



HAL
open science

Loquor, ergo communico-cogito-sum

Hélène Loevenbruck

► **To cite this version:**

Hélène Loevenbruck. Loquor, ergo communico-cogito-sum. Sciences cognitives. Université Grenoble Alpes, 2019. tel-02461867

HAL Id: tel-02461867

<https://theses.hal.science/tel-02461867>

Submitted on 31 Jan 2020

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 destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Université Grenoble Alpes



Mémoire présenté en vue de l'obtention de
L'HABILITATION A DIRIGER DES RECHERCHES
Ecole Doctorale Ingénierie pour la Santé la Cognition et l'Environnement

Loquor, ergo communico-cogito-sum

Hélène Løevenbruck

Laboratoire de Psychologie et NeuroCognition – UMR 5105

Soutenance le 4 juin 2019 devant le jury composé de :

Dr Sylvie Moritz-Gasser

Maître de conférences HDR, Université Montpellier, CHU Montpellier, INSERM U1051

Rapporteuse

Pr Gabriel Bergounioux

Professeur, Université d'Orléans, Laboratoire Ligérien de Linguistique, Université d'Orléans

Rapporteur

Pr Christian Kell

Professeur, Goethe Universität, Brain Imaging Center, Francfort, Allemagne

Rapporteur

Dr François-Xavier Alario

Directeur de recherche CNRS, Laboratoire de Psychologie Cognitive, UMR 7290, Aix-Marseille Université

Examineur

Pr Monica Baciú

Professeure, Université Grenoble Alpes, Laboratoire de Psychologie et NeuroCognition, UMR 5105

Examinatrice

Pr Pascal Perrier

Professeur, Grenoble INP, GIPSA-lab, Département Parole et Cognition, UMR 5216

Président du jury

REMERCIEMENTS

En bientôt vingt-huit ans la liste est tellement longue des personnes à remercier que je ne sais pas par où commencer et je ne voudrais offenser personne et puis les remerciements c'est pathétique et trop intime pour une pudique mais je me lance et je commence par **James Joyce** et son grand monologue intérieur comme ça au moins c'est fait on me prend pour une frimeuse j'assume vu le titre de cette hache des aires de toute façon c'est mort pour la subtilité bon le mieux c'est de commencer par le grand viking qui en 1991 a accepté de m'encadrer en DEA **Pascal Perrier** parce que c'est lui le chercheur-pédagogue qui avec Brassens partage plus que les belles moustaches lui qui m'a appris mes gammes / les secrets des formants et spectrogrammes / vlam ça y est je pars en slam / et puis qui a dirigé ma thèse sans hic / à la recherche des cibles vocaliques / [i a i i A i i è i i È i] ça c'était du slam / puis qui a continué de collaborer avec flamme / sur ces questions de contrôle prédictif / avec lesquelles on s'arrache les tifs / mais qui nous valent quelques articles significatifs / et si j'ai pu simuler l'articulation de [i a i] / c'est grâce à **Rafael Laboissière** merci merci / qui m'expliqua patiemment / la rétropropagation du gradient / tous ses scripts et ses bidules / et son modèle de mandibule puis je ne peux pas penser à Pascal et Rafael sans évoquer leur compère **Jean-Luc Schwartz** le grand chercheur qui fait des gammes / et sait aussi de sa voix aérienne tracer des calligrammes / lui le chercheur / chanteur / enchanteur / je reste dans la famille et dit merci aussi à **Brigitte Colas** / qui nous aidait à délivrer nos mémoires de DEA / et pendant que je pense à ce DEA / je remercie **Pierre Escudier** / qui m'offrit le sésame onirique / l'entrée au fameux nouveau DEA de sciences cognitives de Grenoble / à moi pauvre élève-ingénieure en électronique / l'ouverture aux disciplines les plus nobles / la parole le langage les neurosciences la psychologie la philosophie l'IA la mémoire / le tout sans amalgame / mais avec des gammes / car il portait un deuxième oriflamme / celui de la musique symphonique universitaire / et qui dit musique dit aussi **Pierre Badin** pianiste anti-nucléaire / qui a été co-pilote du vaisseau ICP / aussi bien dans les cieux de paix / que dans les guéguerres / qui dit co-pilote / dit pilote / et donc merci à **Gérard Bailly** ciel / si les [i a i] ont été dicibles / et si la voix du silence s'est faite audible / c'est grâce à tous ses scripts et logiciels / avec lesquels on a segmenté aligné phonétisé étiqueté synthétisé / puisqu'on est dans les B restons-y merci à **Louis-Jean Boë** passionnant passionné de A à Zut des arêtes des poissons aux cervicales des humains de l'indo-européen au sino-tibétain de la photographie à la généalogie je devrais pas mais je le fais quand même merci à **docteur Abry** qui a été inspirateur et questionneur / dommage pour le mister Brya bousilleur / et merci à **Marie-Agnès Cathiard** qui me fit mieux connaître les glides ou comment / de [i a i] on va à [i ya i] tout simplement / maintenant réfléchons comme dirait mon ami **Laurent Girin** / imitant son maître **Gang Feng** son parrain / devenu pro de la séparation de sources musicales / il nous fait passer le sel dans nos soirées amicales / et s'accorde avec **Thomas Hueber** et ses montagnes libres / je les remercie tous les deux au passage merci à **Michel Contini Jean-Pierre Lai** et **Elisabetta Carpitelli** / pour l'inspiration et la motivation lexico-sémantique de kambelibeli / merci à **Yohan Payan** et **Christophe Savariaux** frères de thèse / maintenant collègues qui savent vous mettre à l'aise / même quand brancher un microphone / vous fait ressembler à une bouffonne / grand merci à **David Ostry** sans trac / pour sa générosité sa bienveillance et son Optotrak / chef du fameux Motor Control Lab de Montréal / mais champion du décrochage de téléphone plutôt bancal / merci à **Michel Piquemal** (et **Marion** et **Matthieu**) **Jordi Robert-Ribes** (et **Ester İa** et **Núria**) **François Gaillard** (et **Marie Lola Titouan** et **Plume**) **Mariette Parbelle-Bessac** (et sa famille) **Nadine Ozkan** (et **Jean-Denis** & **Chloé**) **Bertrand Caillaud** **Nathalie Colineau** **Thierry Guiard-Marigny** **Mohammad Akbar Gordon Ramsay** **Olivier Delemar** **Christophe Vescovi** **Lionel Revéret** **Adrian Neagu** **Paul Munteanu** **Lian Apostol** **Antoine Jean-Yves Plinio Barbosa** **Mamoun Alisali** **Bleicke Holm** **Yann Morlec** Ouille Ouille Ouille j'en zappe sûrement et des meilleurs tous ces thésards de la belle époque celle où nous étions encore de jeunes schnoques merci à l'amie cht'i **Anne Vilain** / qui lit l'API sans les mains / que je croise sur de chouettes trajectoires / dans l'équipe articulatoire / dans le DEA en sciences du langage où l'on s'aidait / puis au gouvernail de l'équipe PCMD / et maintenant en tandem / on co-encadre des doctorantes la crème / et on manifeste pour les EULALIES / en espérant développer des applis / une Vilain en cache un autre... merci à **Coriandre Vilain** autrement dit Persil Venimeux / je le remercie ça c'est sérieux / pour toutes ses astuces et ses idées / sa patience devant mes manip mal fagotées merci à tous les membres de l'équipe technique de l'ICP / inlassables supporters / sans qui la technologie pour moi restait un mystère / **Nino Medves** **Christian Bulfone** **Frédéric Elisei** **Monique Revil** merci à toutes les administratives et techniciennes sans qui non plus je n'aurais su avancer **Nadine Bioud** **Dominique Vuillet** **Yvette Gaudé** **Marie-Ange Varin** merci spécial à **Joëlle Miguet** pour son soutien fidèle son réconfort / ses bons conseils ses paroles d'or merci à **Jean-Marc Chassery** grand modeleur / et gouverneur de **GIPSA-lab** / et tous ses équipiers équipières et leurs astrolabes / merci aux collègues du **PRASC** le Pôle Rhône Alpes de Sciences Cognitives et tout particulièrement à **Bernard Amy** **Nicolas Balacheff** **Pierre Bessière** merci aux collègues du plus récent **Pôle Grenoble Cognition** et **Isabelle Maugis** merci à **NeuroCog** merci aux collègues passés et présents de la revue **In Cognito** qui j'espère ressortira de son incognito merci à la fédération **ELESA** et à la **SFR Santé Société**

merci aux enseignants de l'université Stendhal / qui par un DEA à rebours complétèrent mon éducation doctorale / en m'enseignant les beautés de la linguistique / **Chevrot Feu Rakoto** et toute la clique / merci aussi à mes collègues enseignants du temps où à l'IUT j'étais vacataire / puis à l'ENSERG ATER / mais pas à terre/ et en particulier merci à **Xavier Rousset de Pina** / la rime là est facile mais vous ne l'aurez pas / car pour Xavier j'ai infiniment de respect et d'admiration après le DEA le doc le DEA taïaut / post-doc en Ohio / grand grand merci à **Mary Beckman** qui m'accueillit au Phonetics lab / moi la chercheuse bâtarde qui ne savait pas trop ce qu'était une syllabe / puits de science pédagogue infatigable / lève-hyper-tôt pour ses horaires y' a pas d'vocable / elle m'initia à la phonologie de laboratoire / à la prosodie articulatoire / à la palatographie à l'intonation / merci Mary de n'avoir pas rompu le cordon / et de m'avoir intégrée / dans ce projet / sur l'acquisition phonologique / qui nous a fait entrevoir la logique / de pourquoi les bébés anglais disent « ta » pour « ka » / alors que les bébés japonais disent « ka » pour « ta » merci aussi à **Elisabeth Hume** / la phonologue francophone qui fit boum / car avant elle la phonologie était pour moi un mystère abscons / depuis ses cours je me sens dans ce domaine un peu moins con / généreuse végétarienne bon modèle de maman / elle m'a marquée et pas que scientifiquement / merci à **Pauline Welby** elle aussi Colombusienne végétarienne et francophone / elle est experte des montées intonatives et collectionne / les labondance les appelés araignée les noranges les au lion d'or les naprons les dors ange/ en Provence elle vit son idylle / mais a laissé ici des traces indélébiles / Merci aussi à **Keith Johnson Erhard Hinrichs Arnold Zwicky Rebecca Herman Cathy Rogers Mariapaola d'Imperio Stefanie Jannedy Ben Munson Kiyoko Yoneyama Shravan Vasishth Amanda Miller Jennifer Venditti Jen Jenny Bettina Allison Paul Svetlana Mat Janice Marjorie & Gary** et tous les collègues et amis Colombusiens qui ont réussi à me faire apprécier le midwest américain oui oui grand merci à **Lucie Ménard** la balaise / de m'avoir embarquée moi la maudite française / dans son projet très malin / chez les tout p'tits pi les plus grinds / en mesurant leurs lèvres et leurs mâchoires / pour étudier la focalisation articulatoire / pi de garder le contact cordial / j'espère bien qu'on se reverra à Montréal / merci ensuite au **CNRS** / qui m'accepta quelle liesse / au numéro 34 de sa rue ouatée / par moi depuis si longtemps convoitée / merci pour la confiance à long terme sans équivoque / et pour les dots mastocs / pour qu'on ponde des articles-chocs / et puis sachez qu'à **Michel Denis** / et la section 45 / des mercis j'en dis / plus que 45 / puisqu'ils m'ont valu ces gonzes/ de couler un beau bronze / merci à tous les stagiaires sans tarder / que j'ai eu la chance de (co)encadrer / dont les premiers ont fait les frais de mon inexpérience / et les plus récents subissent ma sénescence / notamment merci à **Youcef Semsoum Dorothée Lefebvre Guillaume Rolland Chafika Rabehi Cyril Dubois Tarek Haouam Séverine Gedzelman Laureline Arnaud Caroline Duval Mohamed Ben Jannet Lucie Scarbel Vanessa Bernier Camille Pujol Leslie Lemarchand Manon Gabriel Célise Haldin Louise Charpentier Elena Keracheva Corinne Aubert Flora Gautheron Alexandra Steinhilber Valérie Schott-Brua Myriam El Helou Maïra Tolentino Amélie Castellane Camille Dubrulle Sonja Banjac Marion Daman Clarisse Puissant** / chacun chacune à sa manière / m'a éclairée m'a épatée son et lumière / merci aussi à **Maëva Garnier** pour les manips où l'on s'amusait à crier / dans du bon bruit / avec **Lucie Bailly** / et puis pour les expés mélodieuses / en parole intérieure et silencieuse / après les étudiants/ les doctorants / dont certains sont déjà docteurs merci à **Marc Sato** pour son « sep sep sep sep » infernal / qui nous a bien occupé le lobe pariétal / et pour nous avoir invités à traiter de Neurolinguistique / dans le grand traité de Neurolinguistique / de Pinto-Sato Sato-Pinto /question slam cet édito / c'était du gâteau / merci à ma chère collègue et amie **Marion Dohen** qui a pointé du doigt / que le pointage vocal se voit / lumineuse doctorante étoile filante / elle est devenue enseignante-chercheuse épatante / de la prosodie et du pointage / elle est passée aux gestes sans verbiage / de la parole chez les adultes sains voyons / elle est passée à la recherche-action / chez les personnes avec des difficulté de communication / et avec **Amélie Rochet-Capellan** / elle a monté un projet fascinant / c'est ComEns / c'est immense / une collaboration entre chercheuses / et personnes porteuses / de trisomie 21 / vaste terrain /et puis elle enseigne / à PHELMA / et parfois fff elle-m'a / confié que comme moi / elle en saigne / du peu de temps qu'il reste / sans conteste / pour communiquer ensemble / et quand bon nous semble / merci à **Julia Monnin** thésarde outre-marine / pour ses cartes postales divines / et surtout son savoir faire parler quelle héroïne / les tout petits enfants francophones / et leurs collègues drehu-phones / si seulement la course folle / ralentissait un peu/ comme on pourrait dans les écoles / les aider mieux / merci à **Viet-Anh Tran** qui savait que la voix du silence est d'or / et qui sut la recueillir comme un trésor / chez tout plein de diseurs / d'ailleurs maintenant il assure chez deezer / merci à **Lucile Rapin** qui à l'écoute des voix des schizophrènes / a montré une persévérance de chêne / et même si elle s'est enamourée pour Montréal / elle continue de collaborer joviale / à nos projets sur la parole intérieure / et ses troubles extérieurs / merci à **Ladislav Nalborczyk** comme ses initiales font mon prénom clic / on était fait pour se comprendre / alors bien sûr sans attendre / il a rejoint InnerSpeech notre ANR / et y a contribué brillamment et sans œillère/ en montrant que les ruminations / c'est pas forcément une malédiction / et qu'avec la relaxation / ça peut devenir récréation merci du coup à **Ernst Koster** de l'Université de Ghent Gent ou Gand / pour sa co-direction son savoir et son esprit vaillant / et à **Céline Baeyens** pour ses lumières ses finesses / sa culture son adresse / et sa façon de dire fortuite / que dans sa cuisine il y a huit fuites / merci à **Elsa Spinelli** spécialiste de l'accès lexical / toujours partante et cordiale / qui compense mon absence d'HDR / que depuis 20 ans je dois passer hier / et merci à **Marcela**

Perrone-Bertolotti pour les co-encadrements / les accompagnements / les cheminements / les encouragements / le dynamisme communicatif / et l'engagement créatif / merci à **Geneviève Meloni** / généreuse et prodigieuse / qui sait mêler orthophonie / et recherche fondamentale / et qui nous a entraînées dans un projet international / où on retrouve hourra Montréal / mais aussi Lyon et Liège et à Terre-Neuve l'université Memorial / merci donc aussi au passage à **Estelle Gillet-Perret** orthophoniste épatante / investie dans les données probantes / et ses collègues du CRTLA notamment **Marie-Ange Nguyen-Morel Laurence Casteran Frédérique Bielle Valérie Marcon Rachel Zoubrinetzky Laureline Van der Poorten** et puis aussi à **Anne Monge Anne Descotes Christine Lequette Guillemette Pouget** avec qui on a organisé ces journées internationales / sur les troubles de l'articulation et la Dyspraxie Verbale / merci à **Andrea MacLeod** / pour le co-encadrement commode / à **Sophie Kern et Mélanie Canault** / à **Yvan Rose** et son Phon Oh / Merci à **Laura Machart** / codeuse LPC avec art / pétillante / et brillante / elle apporte une nouvelle énergie / au projet EULALIES merci à **Olivier Clerc** / vif comme l'éclair / pertinent souriant / bienveillant et vaillant / et il est passé des oiseaux aux bébés / après les clins d'œil et l'EMG / Merci à **Annalisa Paroni** pour m'avoir sollicitée sans intox / et tout appris sur le beatbox / les kicks hi-hat et snare / les p'tits cookies / les pâtes au pesto / Merci à **Nathalie Henrich** pour ce co-encadrement / pour la confiance et les encouragements / pour sa voix / et sa joie après les doctorants merci aux post-doctorants **Francesca Carota** et à **Romain Grandchamp** / pour son talent et sa vision grand champ / merci à tous ceux qui m'ont appris l'imagerie cérébrale / merci mille fois merci tellement / à **Monica Baciu** / belle vive touchante et cérébrale / mon académie ma chevalière / devenue mon amie ma conseillère / Merci à **Christoph Segebarth Laurent Lamalle Irène Troprès Cédric Pichat Emilie Cousin Sylvain Harquel** à **Christian Jutten** bâtisseur de ponts costauds /entre les traiteurs /et les utilisateurs / de tous signaux merci à **Mircea Polosan** pour les collaborations / et les discussions / sur les hallucinations / et les ruminations / merci à **Nicolas Franck** pour son franc accueil à Bron / et pour partager sans écueil sa vision / merci à celles et ceux qui m'ont initiée aux recherches sur le nourrisson / à **David Méary** / à **Eve Duperrix** / à **Anne Hillairet de Boisferon** / à **Marie Sarreméjeanne** / à **Mathilde Fort** / à tous les doctorants et étudiants du Babylab / merci surtout à **Olivier Pascalis** / et non pas Pascal Olivieris / avec qui je partage l'anglo-sénégalophilie vraiment / en plus de quelques ANR et encadrements / sous sa veste à carreaux c'est un grand sage / à qui je dois d'avoir osé prendre un virage / pour un quotidien sans nuage / un nouveau voyage / de nouveaux copinages / merci à **Sylviane Valdois** / qui comme directrice a eu confiance en moi / et m'a accueillie au Laboratoire de Psychologie / et NeuroCognition / où j'ai trouvé nouveau logis / et collaborations / merci pour sa compréhension / et son inspiration / merci aux collègues anciens et nouveaux de l'équipe langage pas encore ici remerciés **Marie-Line Bosse** / qui fut ma clean bosse / **Julien Diard** / qui sort Bayes dare-dare / **Sophie Donnadieu** / pour les oreilles elle a des yeux / **Sonia Kandel** / car elle scande elle / **Benoît Lemaire** / poète comme Homère / modèle de mémoire / il modélise la mémoire / **Boris New** / et LEXIQUE version new / **Ronald Peereman** / et son patrimoine / de manulex / bons pour le cortex / merci à mon bienveillant réconfortant / co-bureautant / **David Alleysson** qui répond quand y'a l' téléphone qui son/ merci à toutes celles et ceux au **LPNC** / qui ont partagé sans grimacer / leur science / leur appétence / leur gaïté / leur doigté / leurs connaissances / leurs compétences / en vision / en perception / en mémoire / en savoir / en apprentissage / en psycholangage / pour leur impact et leur tact / merci en vrac / à **Julien Barra Carole Berger Martine Bouvard Aurélie Campagne Christine Cannard Annik Charnallet Corinne Cian Jérôme Clerc Nathan Faivre Nathalie Fournet Christian Graff Michel Guerraz Pascal Hot Caroline Jolly Karine Mazens Martial Mermillod Christopher Moulin Richard Palluel-Germain Dominic Pérennou Céline Souchay Laurent Torlay** mais ma parole c'est fait exprès tous ces noms ils sont tous là j'espère qu'il n'en manque m'pas-un dans le joli moulin chaud de la campagne pâle et pérenne où elle tord les souches et après sa fournée fait vrai ! un graphe cyan clair d'un canard dans un char n'allez pas croire que ça finit dans un bar amer mi eau mi anisette berger merci à **Guylaine Omnès-Sabardak** / pour la gentillesse la finesse et les nombreux tours dans son sac / à **Claire Leroy** / la reine des gestionnaires pour l'énergie et la joie / merci à **Sanie Claraz** et **Nathalie Dupuy** / merci à **Eric Guinet** et **Thierry Rakotoarisoa** / pour les astuces informatiques maintes et maintes fois / et merci aux anciens collègues LPNCiens / je peux pas les citer tous mais chacun reconnaîtra les siens / merci à toutes les directrices et directeurs de l'école doctorale ingénierie pour la santé la cognition et l'environnement / notamment **Anne Guérin** la dernière au commandement / pour sa patience / et son influence / pour son sens du partage / pour son cœur à l'ouvrage / pour la complicité musicale / et la confiance amicale / merci aux collègues responsables du master de Sciences Cognitives de Grenoble / attentifs éducatifs et cœurs nobles / **Jeanny Hérault** / ce vrai héros / **Stéphane Rousset** / partager et enseigner il sait / **Nathalie Guyader** grâce à elle le master ne manque pas d'air / **Alan Chauvin** / le collègue qui en vaut vingt / merci pour les échanges aux anciens et présent co-organisateurs de la Semaine du Cerveau dont **Bettina Debû Antoine Depaulis Isabelle Le Brun Marie-Jo Moutin Sylvie Bretagnon Laurent Vercueil Vanessa Delfau Géraldine Favre** / pour l'ouverture aux techniques d'activation de la conscience et à l'auto-hypnose / merci à **Jean Becchio** ses transes et sa prose/ grand merci à **Stéphanie Smadja** / qui il y a presque 5 ans déjà / m'a gentiment embarquée / dans son passionnant Monologuer / pour les rencontres les cogitations / les découvertes les discussions many thanks to **Charles Fernyhough** / scientist yet artist also / for the delightful Voices Within and the poetic Box of Birds / for the continuous inspiration / and the generous invitation / thanks also to **Agustin Vicente** and

Peter Langland-Hassan for the incitation to write and for orchestrating these New Voices in Inner Speech merci aux collègues qui m'ont invitée dans leurs projets nationaux ou internationaux / financés ou en attente du gros lot / notamment **Pascal Barone Sophie Dufour Denis Perrin** / merci aux Aixois notamment **Noël Nguyen Michel Piterman Yohann Meynadier Maud Champagne-Lavau et Mariapaola D'Imperio** pour l'alignement tonal articulatoire et l'ironie **Núria Esteve-Gibert** pour la prosodie et les gestes chez les petits / merci aux Strasbourgeois articulatoires / particulièrement **Béatrice Vaxelaire** / et **Rudolph Sock** / pour la parole est spéciale sans choc / merci aux parisiens notamment **Evelyne Bloch-Gallego Pierre-Louis Patoine et Sharon Peperkamp Anne Christophe** / stimulantes neurophilosophes / merci au (ex)lyonnais **Melissa Barkat François Pellegrino Jean-Philippe Lachaux Olivier Bertrand Johanna Lévêque** / merci aux Lillois notamment **Renaud Jardi Charles-Edouard Notredame & Anahita Basirat** / merci en vrac aux collègues prosodistes **Corine Astésano Cécile Fougeron Anne Lacheret** / merci au **Neurocercle** notamment **Claude Marti** et encore Jeanny Hérault pour les passionnantes discussions / merci à **Jean-François Dars et Anne Papillault** pour les photographies éclairées les pointages à gogo les histoires courtes mais à longue portée / merci pour les campements sauvages scientifiques et les impromptus minutieusement réglés pour les oxymores croustillants et les réflexions savantes à **Mickaël Chouquet Balthazar Daninos et Léo Larroche** / merci pour le lancement des Langues de Babylab à la pétillante et enthousiasmante **Christelle Pillet** de Médiarts et pour leur participation à **Aurélie Maisonneuve Myriam Roulet Bertille Puissat** et pour leur soutien **Christophe Monge Simon Barral-Baron** et aussi **Célie Rodriguez Marie-Christine Bordeaux** et les collègues de l'EST à l'UGA

merci à **Sylvie Moritz-Gasser Gabriel Bergounioux Christian Kell François-Xavier Alario Monica Baciú et Pascal Perrier** de me faire l'honneur et le bonheur d'être membres de mon jury

merci plus personnels à **Brassens Higelin Grand Corps Malade Rimbaud et Анна Ахматова** aux musiciens de **Piccola Musica** et à notre incroyable chef **Michel Viguier** violoniste et homme-orchestre / merci aux grands **Mark Drobinsky** et son Testore et **Amanda Favier** et son Goffriller / merci à **Marieke Bouche** / qui joue Bach un brin manouche / merci à **Nathalie Dombre** / pour son piano de lumière sans ombre / et merci aux amis fidèles et leurs petits et grands / qui pour moi comptent tant et tant chacune et chacun / **LN & Serge Youss & Wood Jibe & Nat Beu & Læt Adel & Lamia RV Valoche Fabrice Pat & Poops Catherine Clara Agathe & Christian Fred & Béa Heather & Fabrice Meriel Annie & John Rebecca & Arif The Tudors Xavier & Hélène Pierric Wajih** tous les **Vauquoirin docteur Valérie Cristo & Juliette Guillaume & Aurélie Alain Lionel Annabelle Beth Lisbeth Malene Cawa Marie-Jo & JP Danielle & JP Charlotte & Jean-Noël** qui m'a évité quelques dérapages / dans ces 30 premières pages / merci à **Marie-No et Joëlle** les Mary Poppins qui ont fait que le matin / je quittais sans trop de chagrin / mes deux petits coquins / car je savais qu'ils étaient entre de bonnes mains / merci à **Martine Montagne** qui par l'Idogo nous accompagne / merci à **Maria** / merci à **Yves** et sa famille pour ses encouragements / merci à et **Jean-Marc** et sa famille pour les discussions et pour Vygotski / merci au regretté **Jean-Claude** pour ses lumières et son initiation à Descartes et Saint-Augustin / merci à **Pistoule** chorégraphe inspirée fascinante et sa famille / merci à **Riri** écrivain éclectique clairvoyant et sa famille / si j'étais aussi créative / et aussi imaginative / que mon petit frère et ma petite sœur / la hache des ères aurait été déterrée à l'heure / un merci immense incommensurable à **Kiki et Japy** / pour tout ce qu'ils m'ont transmis / l'amour des langues des sciences et de la pédagogie / merci pour leur soutien sans compromis / merci pour leur joie de vivre et leur énergie / merci à **Nicolas et Jean** / mes deux précieux enfants / précoces babilleurs et pointeurs / le violoncelliste et le pianiste / le bassiste et le batteur / le littéraire et le bricol-artiste / leurs questions leurs réactions / m'ont fait faire des ascensions / je les ai pas élevés c'est grâce à eux que je m'élève / c'est grâce à eux que je me lève / car tous les deux créateurs / ils me font mes plus belles heures / merci infini à l'amour de ma vie mon **Philippe** / s'il n'était si présent je partirais en flip / merci pour ton soutien fort / ton réconfort / ta foi en moi / pour ta patience / ta constance / ton esprit / pour tes épaules / ton humour / tes frichtis / l'ouverture de ce manuscrit / tu l'auras compris / c'est pour toi

alors que je finissais ces dernières lignes / ma tendre et lumineuse tante **Sabine** / s'est éteinte / à elle et sa famille je dis merci / et avec une émotion sans feinte / je pense à celles et ceux de nos collègues et proches qui bien trop tôt sont partis et à leurs familles / en particulier **Christian Benoit Bernard Teston Eric Vatikiotis-Bateson / Colin Evans Magali Laurencin Delphine Brozzoni Frédéric Bouaniche Dalila Ouadah** et notre cher **Laurent Durkalec**

A Philippe, Nicolas et Jean

« Si on s'accroche à vouloir saisir le mieux possible ce qu'on voit, qu'on fasse de la science ou de l'art, la démarche est la même. Un savant qui se spécialise dans n'importe quel domaine, plus il en connaît, plus il a à connaître, et il ne doit jamais y avoir d'espoir non plus d'arriver à une connaissance totale. D'ailleurs, la connaissance totale, ce serait la mort même. L'art et la science c'est tâcher de comprendre. »

Giacometti (1962)

SOMMAIRE

Table des figures	xii
Table des encadrés	xiv
Liste des abréviations	xv
INTRODUCTION	1
CHAPITRE I. ESSAIS DE FORMALISATION THEORIQUE	3
I.1. TROIS FONCTIONS PRIMORDIALES DU LANGAGE	6
I.1.1. LA FONCTION COMMUNICATIVE : <i>LOQUOR, ERGO COMMUNICO</i>	7
I.1.2. LA FONCTION COGNITIVE D'ELABORATION ET D'EXPRESSION DE LA PENSEE : <i>LOQUOR, ERGO COGITO</i>	9
I.1.2.1. Unité entre pensée et langage : le langage comme reflet et expression de la pensée	10
I.1.2.2. Le langage comme outil de construction et de structuration de la pensée	11
I.1.2.3. Le langage comme système dédié prioritairement à la pensée	12
I.1.2.4. La pensée sans langage ?	14
I.1.2.4.1. Conception cognitive forte : le langage comme pilier de la pensée	15
I.1.2.4.2. Conceptions cognitives faibles : le langage comme étai de la cognition	16
I.1.2.4.3. Arguments contre la modularité du langage et de la cognition	20
I.1.2.4.4. Arguments contre le rôle nécessaire du langage dans la pensée : pensée sans langage et modes de pensée visuel, kinesthésique, abstrait	23
I.1.2.4.5. L'indicible et l'innommable	27
I.1.2.4.6. Synthèse : fonctions communicative et cognitive	28
I.1.3. LA FONCTION D'AUTONOËSE : <i>LOQUOR, ERGO SUM</i>	30
I.1.3.1. La métacognition : prise de conscience de sa propre pensée par le sujet parlant	30
I.1.3.2. La construction du tandem soi/autrui	31
I.1.3.2.1. Langage et conscience de soi : de « Je pense donc je suis » à « Je est un autre »	32
I.1.3.2.2. Le soi dialogique : subjectivité et altérité	32
I.1.3.2.3. Le soi narratif	35
I.1.3.2.4. La conscience de soi minimale et le développement de la conscience de soi	36
I.1.3.2.5. Le soi temporellement étendu (le soi dans le temps) : la conscience auto-noétique	38
I.1.3.2.6. La pratique du langage intérieur et le renforcement de la conscience de soi	41
I.1.3.2.7. La vertu thérapeutique du langage et du langage intérieur	46
I.1.4. LA TRIADE COMMUNICATION-PENSEE-AUTONOËSE	48
I.2. UNE CYBERNETIQUE ADAPTEE A CES FONCTIONS	50
I.2.1. LE <i>MONITORING</i> VERBAL DANS LA COMMUNICATION	50
I.2.1.1. La production du langage et le <i>monitoring</i> verbal égocentré	54
I.2.1.1.1. Les différents niveaux de traitement impliqués dans la production du langage	54
I.2.1.1.2. Le <i>monitoring</i> verbal égocentré : suivi des verbalisations auto-produites	55
<i>Monitoring</i> par contrôle prédictif de l'étape d'articulation	59
Nature des objectifs phonétiques : somatosensoriels, auditifs, supramodaux ?	63
Implémentation neuronale du contrôle prédictif de l'étape d'articulation	67
Rôle du cervelet dans l'étape d'articulation : modèles contrôleur et prédicteur	69
<i>Monitoring</i> hiérarchique par contrôle prédictif	75
I.2.1.2. La perception du langage et le <i>monitoring</i> verbal allocentré	86
I.2.1.2.1. Les différents niveaux de traitement impliqués dans la perception du langage	86
I.2.1.2.2. Le <i>monitoring</i> verbal allocentré : suivi des verbalisations d'autrui	91
Atténuation sensorielle de la parole auto-produite et renforcement de la parole d'autrui	99

Importance de la temporisation et de la synchronisation des signaux comparés	100
I.2.2. LE <i>MONITORING</i> VERBAL DANS LA PENSEE	101
I.2.2.1. Trois axes du langage intérieur : condensation, dialogalité et intentionnalité	101
Parole intérieure dialogale ou dialogique ?	106
I.2.2.2. Le <i>monitoring</i> verbal intérieur : suivi du contenu de nos productions intérieures	110
Données empiriques sur le <i>monitoring</i> verbal intérieur	110
Modélisation du <i>monitoring</i> verbal intérieur : modèle Condialint	111
I.2.3. L'ATTRIBUTION DE L'AGENTIVITE DANS L'AUTONOESE	116
Pertinence de l'attribution d'agentivité en parole intérieure	117
Hallucinations auditives verbales : défaillance de l'attribution de l'agentivité en parole intérieure	118
Hallucinations auditives verbales non psychiatriques : les entendeurs de voix	118
Hallucinations auditives verbales : explication dans le cadre du <i>monitoring</i> par contrôle prédictif	119
Le mécanisme d'atténuation sensorielle lors d'actions auto-produites, dont la parole	119
Probable atténuation sensorielle lors de la parole intérieure	120
Rôle de la dialogalité dans les hallucinations auditives verbales	123
Rôle de la non-intentionnalité dans les hallucinations auditives verbales	125
I.2.4. PREMIERES BRIQUES D'UN FORMALISME INTEGRE DE LA CYBERNETIQUE DE LA TRIADE COMMUNICATION-COGNITION-AUTONOESE	130
I.2.4.1. La cybernétique de la fonction de communication	132
I.2.4.2. La cybernétique de la fonction de pensée	134
I.2.4.3. La cybernétique de la fonction d'auto-noèse	134
I.2.4.4. Vers une cybernétique intégrée des trois fonctions du langage	135
CHAPITRE II. TRAVAUX EXPERIMENTAUX	139
II.1. MONITORING VERBAL POUR LA FONCTION DE COMMUNICATION	139
II.1.1. LES CAPACITES DE POINTER ET DE COMPRENDRE LE POINTAGE	139
II.1.1.1. Production du pointage prosodique chez l'adulte	140
II.1.1.2. Perception audiovisuelle du pointage prosodique chez l'adulte	141
II.1.1.3. Corrélats cérébraux de la production de pointage multi-formes	141
II.1.1.4. Corrélats cérébraux de la perception du pointage prosodique	142
II.1.2. MANŒUVRES ET VARIATIONS PROSODIQUES LIEES AU SUIVI D'AUTRUI	143
II.1.2.1. Structuration prosodique des énoncés en français : données acoustiques, articulatoires et perceptives	143
II.1.2.2. Alignement tonal en français : données acoustiques et articulatoires	147
II.1.2.3. Ironie et sarcasme	147
II.1.2.4. Effet Lombard sur la prosodie : données acoustiques et articulatoires	148
II.1.3. SUIVI VISUEL DE SOI ET D'AUTRUI	149
II.2. MONITORING VERBAL POUR LA FONCTION DE PENSEE	150
II.2.1. <i>MONITORING</i> DU LANGAGE INTERIEUR	150
II.2.1.1. Hypothèses théoriques	150
II.2.1.2. Données physiologiques	151
II.2.1.3. Données de neuroimagerie	152
II.2.1.4. Données introspectives	155
II.2.2. LES RUMINATIONS MENTALES : DEFICIT DE CONTROLE DE LA PAROLE INTERIEURE	157
II.3. SUIVI DE L'AGENTIVITE POUR LA FONCTION D'AUTONOESE	159
II.3.1. DEFICIT DU SUIVI DE L'AGENTIVITE : PHYSIOLOGIE DES HALLUCINATIONS AUDITIVES VERBALES	159
II.3.2. DEFICIT DU SUIVI DE L'AGENTIVITE : NEUROIMAGERIE DES HALLUCINATIONS AUDITIVES VERBALES	160
II.4. DEVELOPPEMENT DU LANGAGE : MISE EN PLACE DU <i>MONITORING</i> VERBAL DE SOI ET D'AUTRUI	163
II.4.1. ROLE DU <i>MONITORING</i> VERBAL D'AUTRUI DANS LE DEVELOPPEMENT PHONOLOGIQUE : PROJET PAIDOLOGOS	163

II.4.2. ROLE DU <i>MONITORING</i> VERBAL EGOCENTRE DANS LA PERCEPTION PRECOCE	167
II.4.3. <i>MONITORING</i> VERBAL DE SOI POUR AUTRUI : POINTAGE CHEZ L'ENFANT	169
II.4.3.1. Corrélats acoustico-articulatoires du pointage prosodique chez l'enfant	169
II.4.3.2. Corrélats acoustico-corporels du pointage prosodique chez l'enfant	170
II.4.3.3. Production de pointage et de gestes manuels chez les enfants présentant des déficits langagiers	172
II.4.4. DEVELOPPEMENT PHONOLOGIQUE TYPIQUE ET ATYPIQUE : <i>MONITORING</i> VERBAL DE SOI COMPLEXIFIE	172
II.4.5. PRODUCTION DE PAROLE CHEZ LES ENFANTS IMPLANTES COCHLEAIRES : LE <i>MONITORING</i> VERBAL DE SOI ET D'AUTRUI ALTERE	175
II.4.6. DEVELOPPEMENT DU <i>MONITORING</i> INTER-MODAL DU LANGAGE ET DU VISAGE D'AUTRUI	177
CHAPITRE III. APPLICATIONS THERAPEUTIQUES, EDUCATIVES ET TECHNOLOGIQUES	179
III.1. REMEDIATION ET PREVENTION DES TROUBLES DU LANGAGE CHEZ L'ENFANT	179
III.1.1. APPORT DE LA MODALITE GESTUELLE DANS LES TROUBLES DEVELOPPEMENTAUX DU LANGAGE ORAL	179
III.1.2. INTERVENTION ORTHOPHONIQUE FONDEE SUR UNE ANALYSE DIFFERENTIELLE DES PROFILS LANGAGIERS	180
III.1.3. PREVENTION DES TROUBLES LANGAGIERS : INTERVENTIONS LANGAGE ET MUSIQUE CHEZ LE NOURRISSON	181
III.2. PRISE EN CHARGE DES TROUBLES DE LA PAROLE INTERIEURE : HAV ET RUMINATIONS	181
III.3. TROUBLE DE LA VOIX : SYNTHESE DE PAROLE AUDIBLE A PARTIR DE PAROLE SILENCIEUSE	182
III.4. APPORT D'INFORMATIONS VISUELLES ARTICULATOIRES DANS LA REEDUCATION DE PATIENTS PRESENTANT UNE APHASIE DE BROCA NON-FLUENTE	185
CHAPITRE IV. PROGRAMME DE RECHERCHE	187
IV.1. VERS UN MODELE NEUROCOMPUTATIONNEL INTEGRE DE LA CYBERNETIQUE VERBALE SOUTENANT LA TRIADE COMMUNICATION-COGNITION-AUTONOESE	187
IV.2. MISE A L'EPREUVE DE LA MODELISATION PROPOSEE PAR LE RECUEIL DE DONNEES EXPERIMENTALES	188
IV.2.1. CYBERNETIQUE VERBALE POUR LA FONCTION DE COMMUNICATION	188
IV.2.2. CYBERNETIQUE VERBALE POUR LA FONCTION DE PENSEE	190
IV.2.3. CYBERNETIQUE VERBALE POUR LA FONCTION D'AUTONOESE	192
IV.3. APPLICATIONS THERAPEUTIQUES ET EDUCATIVES ENVISAGEES	192
IV.3.1. REEDUCATION ET PREVENTION DES TROUBLES DU LANGAGE	192
IV.3.2. HALLUCINATIONS AUDITIVES VERBALES ET RUMINATIONS MENTALES : A L'ECOUTE DES ENTENDEURS DE VOIX ET DES RESSASSEURS DE MOTS	193
IV.3.3. SYNTHESE VOCALE A PARTIR DE LA PAROLE SILENCIEUSE	194
IV.3.4. APPORT THERAPEUTIQUE DE LA PRATIQUE DU LANGAGE INTERIEUR	194
CONCLUSION	195
RÉFÉRENCES BIBLIOGRAPHIQUES	197
ANNEXES	219
CV détaillé	220
Sélection de publications	229
Résumé	351
Abstract	352

Table des figures

Figure I-1. Le mur des Je t'aime, à Paris XVIIIème	3
Figure I-2. Les trois fonctions du langage étudiées ici.....	6
Figure I-3. Déterminatif de l'égyptien ancien qui représente les actions de la bouche et qui exprime indifféremment les idées de manger, boire, crier, parler, penser, méditer, connaître, juger. Tiré de Maspero (1875).....	10
Figure I-4. Illustration de la conception modulaire des liens entre langage et pensée telle que défendue par Carruthers(2002).....	18
Figure I-5. Illusion de l'échiquier d'Adelson.....	21
Figure I-6. Schéma classique de la communication linguistique avec la symétrie locuteur-rice / allocutaire des trois niveaux fondamentaux.	51
Figure I-7. Boucles de rétroaction induites par le monitoring verbal de soi et d'autrui.....	53
Figure I-8. Le monitoring des actions par contrôle prédictif, avec modèles internes inverse (contrôleur) et direct (prédicteur) et copie d'efférence.....	60
Figure I-9. Le monitoring de l'étape d'articulation par contrôle prédictif.....	62
Figure I-10. Réseau neuronal impliqué dans l'étape d'articulation par contrôle prédictif, incluant la production de parole intérieure.....	67
Figure I-11. Conjecture (non publiée) sur l'implémentation dans le cervelet des modèles internes inverse (contrôleur) et direct (prédicteur) nécessaires au MCP de la parole, et sur les connexions entre le cervelet et les cortex cérébraux sensoriel et moteur.	74
Figure I-12. Modèle hiérarchique du monitoring prédictif de la production de langage.....	80
Figure I-13. Modèle hiérarchique du monitoring de la perception du langage (circuit rouge), greffé sur le monitoring de la production de la parole (circuit vert).....	98
Figure I-14. Trois dimensions de la parole intérieure : condensation, dialogalité et intentionnalité.....	109
Figure I-15. ConDialInt : Modèle hiérarchique de la production de parole intérieure	113
Figure I-16. Modèle explicatif des hallucinations auditives verbales fondé sur une défaillance de la production de parole intérieure vagabonde, avec la voix d'autrui.	127
Figure I-17. Modèle intégré de l'attribution de l'agentivité lors de la production ou de la perception de parole, incluant les variantes de parole intérieure monologale et dialogale.	129
Figure I-18. Proposition provisoire de formalisme intégré de la cybernétique des trois fonctions du langage : communication-pensée-autonoèse	136
Figure II-1. Recueil de données articulatoires : a. suivi vidéo, b. suivi optoélectrique	140
Figure II-2. Activations observées dans chacun des pointages chez les 15 participants.	141
Figure II-3. Contraste entre perception d'énoncés Focalisés vs. Neutres	142
Figure II-4. Suivi de F ₀ et durées syllabiques pour.....	145
{[Le rat marron] _{SA1} [voulait manger] _{SA2} [le long mulot.] _{SA3}] _{SI}	145
Figure II-5. Recueil de données prosodiques articulatoires.....	146
Figure II-6. Variabilité de l'alignement pour une même syllabe CVC : a. le pic de la montée finale est aligné avec la voyelle ; b. le pic est aligné avec la consonne coda.....	147
Figure II-7. Signaux acoustiques correspondant aux productions littérale (en haut) et sarcastique (en bas) de la phrase « il est en pleine forme » chez une locutrice française.....	148
Figure II-8. Positionnement des capteurs électromyographiques de surface pour mesurer l'activité des muscles labiaux lors de la production de parole intérieure.....	151
Figure II-9. Positions des capteurs électro-oculographiques et électromyographiques de surface pour mesurer les clignements des yeux et l'activation des muscles orofaciaux.	152
Figure II-10. Réseau cérébral activé lors de la production intérieure avec focalisation prosodique (vs énoncé neutre). Tiré de Lævenbruck, Baciù, Segebarth & Abry (2005)	152
Figure II-11. Contrastes entre chaque condition et la ligne de base, tracés sur une image de cerveau standard (issue de BSPMview, Spunt, 2016).	155
Figure II-12. Activation EMG (lèvre supérieure, lèvre inférieure, bras, front) et scores de rumination après induction de ruminations.....	157
Figure II-13. Corrélats EMG des hallucinations auditives verbales chez des patients souffrant de schizophrénie.....	160
Figure II-14. Réseau impliquant les aires spécialisées dans le traitement de la parole et de la voix et montrant une différence entre patients (hyperactivité) et sujets contrôles dans les deux tâches (Perception de la parole et Génération de pensée verbale).....	161
Figure II-15. Réseau impliquant les aires du traitement de la parole (panneau A), montrant une différence entre patients schizophrènes hallucinés et non-hallucinés dans la tâche de perception (panneau B).	162
Figure II-16. à gauche : Langues de Nouvelle-Calédonie ; à droite : Enregistrement d'une enfant à domicile, dans la tâche de répétition de mots illustrés	164

Figure II-17. Scores de production correcte pour les consonnes /t, d, k, g, s, ʃ et tʃ/, en fonction de l'âge, en drehu (rouge) et français (bleu).....	165
Figure II-18. Scores de production correcte pour les consonnes /t, d, k, g, s, ʃ et tʃ/, en fonction de la voyelle, en drehu (à gauche) et français (à droite)	165
Figure II-19. Scores de production correcte pour les consonnes /t, d, k, g, s, ʃ, tʃ/ en fonction de la fréquence de la syllabe CV dans la langue, en français (en haut) pour les tranches d'âge de 2 à 5 ans et drehu (en bas) pour les tranches d'âge de 3 à 5 ans.....	166
Figure II-20. Scores d'association audiovisuelle correcte pour /b, d/ à gauche et /v, z/ à droite, en fonction des capacités de babillage des nourrissons de 6-9 mois	168
Figure II-21. Données articulatoires extraites de données vidéo	169
Figure II-22. Analyse des données vidéo des enfants enregistrés au cours d'un jeu élicitant une focalisation prosodique semi-spontanée	171
Figure II-23. Analyse en radar des scores des enfants à la batterie de tests EULALIES, sur un petit nombre d'enfants tests.	174
Figure II-24. A gauche : Schéma d'un implant cochléaire (IC) ; au centre : espaces vocaliques des enfants normo-entendants ; à droite : espaces vocaliques des enfants porteurs d'IC.	175
Figure II-25. A gauche : les stimuli présentés aux bébés ; à droite : un bébé en plein monitoring visuel d'autrui	177
Figure III-1. Analyse multilinéaire des productions d'un enfant présentant un trouble phonologique	180
Figure III-2. Courbe de F ₀ originale et resynthétisées par les deux méthodes pour l'énoncé « Armstrong tombe et s'envole. ».	183
Figure III-3. Enregistrements NAM et vidéo.	184

Table des encadrés

Encadré I.1. La Fonction Communicative du langage : Loquor, ergo communico	9
Encadré I.2. La Fonction Cognitive du langage : Loquor, ergo cogito	14
Encadré I.3. Le langage comme nécessaire à la pensée conceptuelle.....	16
Encadré I.4. Le langage comme nécessaire à la pensée propositionnelle intégrée ou comme étayage de la cognition et de la métacognition	20
Encadré I.5. Contre la conception du langage comme un module autonome	23
Encadré I.6. Modes de pensée non verbaux.....	27
Encadré I.7. Fonction communico-cognitive : Loquor, ergo communico et cogito	29
Encadré I.8. Langage et métacognition	31
Encadré I.9. Langage et prise de conscience de soi	32
Encadré I.10. Dialogue et prise de conscience de soi et d'autrui	35
Encadré I.11. Le soi narratif.....	36
Encadré I.12. Développement de la conscience de soi chez l'enfant	38
Encadré I.13. Langage et conscience autoéotique	41
Encadré I.14. La spécificité humaine : capacité délibérée de simulation de la parole	46
Encadré I.15. Langage et autoéotie	48
Encadré I.16. La triade systémique qui a mené à la complexité et la diversité du langage	49
Encadré I.17. La nature supramodale des objectifs phonétiques	65
Encadré I.18. Distinction entre objectifs supramodaux et petite voix audible	66
Encadré I.19. Conjecture sur les rôles respectifs des lobes antérieur et postérieur du cervelet dans l'étape d'articulation : contrôleur et prédicteur	72
Encadré I.20. Monitoring localisé à chaque niveau ou monitoring global ?	81
Encadré I.21. Ajustement probabiliste des prédicteurs aux différents niveaux ?	82
Encadré I.22. Alignement temporel des signaux à comparer ?	83
Encadré I.23. $M_S(v_S)$: monitoring verbal égocentré	86
Encadré I.24. Rôle crucial de l'alignement temporel des signaux à comparer.....	100
Encadré I.25. $M_S(v_O)$: monitoring verbal allocentré	101
Encadré I.26. Dimension de condensation	106
Encadré I.27. Dimension de dialogalité	107
Encadré I.28. Dimension d'intentionnalité	108
Encadré I.29. Monitoring verbal intérieur	111
Encadré I.30. Monitoring et agentivité lors de la parole intérieure non délibérée ?	114
Encadré I.31. $M_S(<v_S>)$, $M_S(<v_S'>)$, et $M_S(<v_O>)$: monitorings verbaux intérieurs égocentrés et allocentré	116
Encadré I.32. Agentivité : subjectivité vs. altérité	116
Encadré I.33. Agentivité : comparaison entre signaux désiré et prédit	123
Encadré I.34. Dissociation entre monitoring verbal et attribution de l'agentivité	126
Encadré I.35. Les hallucinations auditives verbales : défaillance spécifique de l'attribution de l'agentivité lors de l'imagination non délibérée d'autrui parlant.....	128
Encadré I.36. Attribution de l'agentivité	128

Liste des abréviations et remarques préliminaires

AMS : aire motrice supplémentaire

C : consonne

CV : syllabe formée d'une consonne suivie d'une voyelle

CCA : cortex cingulaire antérieur

DLPF(C) : (cortex) dorsolatéral préfrontal

ECoG : électrocorticographie

EEG : électroencéphalographie

EMG : électromyographie

F_0 : fréquence fondamentale

F1, F2, F3 : premier, deuxième et troisième formants acoustiques

GFI : gyrus frontal inférieur

GFM : gyrus frontal moyen

GFS : gyrus frontal supérieur

GSM : gyrus supramarginal

GTM : gyrus temporal moyen

GTS : gyrus temporal supérieur

HAV : Hallucination Auditive Verbale

IC : Implant Cochléaire

IRMf : imagerie par résonance magnétique fonctionnelle

JTP : jonction temporo-pariétale

L1 : première langue

L2 : seconde langue

MCP : *monitoring* par contrôle prédictif

M1 : cortex moteur primaire

NCP : noyaux cérébelleux profonds

OI : olive inférieure

PM : (cortex) prémoteur

Pont : noyaux pontiques

S1 : cortex sensoriel primaire

SA : syntagme accentuel

SI : syntagme intonatif

STS : sillon temporal supérieur

TDSP : Troubles du Développement des Sons de Parole (*Speech Sound Disorder*)

Th : thalamus

TMS : Transcranial Magnetic Stimulation (stimulation magnétique transcrânienne)

TSLO : Trouble Spécifique du développement du Langage Oral

V : voyelle

Fonction du langage : terme pris ici au strict sens biologique, c'est-à-dire au sens de la simple description de ce que fait le langage, son rôle, sa contribution, son usage, à quoi il sert, pour quoi (en deux mots) on parle.

Modèle interne : en robotique et contrôle moteur : simulation directe de l'appareil moteur (transforme la commande motrice en action et en état sensoriel) ou son inverse en termes de contrôle (prend en entrée un état sensoriel, fournit en sortie la commande motrice à appliquer)

Modèle interne direct (prédicteur) : simulateur de l'appareil moteur, génère une prédiction sensorielle à partir de la commande motrice qui lui est fournie en entrée.

Modèle interne inverse (contrôleur) : représentation inverse de celle du simulateur de l'appareil moteur, calcule la commande motrice à appliquer pour atteindre l'état sensoriel désiré qui lui est fourni en entrée

Monitoring : contrôle, suivi, régulation

A : attribution de l'agentivité à une action en cours

M : *monitoring* : suivi et éventuellement correction d'une action en cours

v : production verbale extériorisée (ou verbalisation extériorisée)

<v> : production verbale imaginée (ou verbalisation intérieure)

S : *Self*, l'agent de l'action est soi-même

O : *Other*, l'agent de l'action est autrui

M_S(v_S) : *monitoring* verbal égocentré ou suivi par soi-même des verbalisations auto-produites

M_S(v_O) : *monitoring* verbal allocentré ou suivi par soi-même des verbalisations d'autrui

M_S(<v_S>) : *monitoring* verbal intérieur égocentré, de notre parole intérieure avec notre propre voix

M_S(<v_S'>) : *monitoring* verbal intérieur égocentré, de notre parole intérieure avec une voix modifiée

M_S(<v_O>) : *monitoring* verbal intérieur allocentré ou *monitoring* de l'imagination d'autrui verbalisant

Une écriture inclusive a été tentée au long de ce manuscrit. Toutefois, au pluriel, pour faciliter la lecture et alléger l'écriture, j'ai conservé la règle de primauté du masculin formulée au XVIIème siècle, le masculin étant en effet plus court...

INTRODUCTION

Loquor, ergo communico-cogito-sum

ou je parle, donc je communique-je pense-je suis.

Ce titre latin n'est pas de la pédanterie. C'est un hommage à René Descartes, qui, bien qu'il soit décrié par certains chercheurs contemporains en sciences cognitives pour sa conception dualiste du corps et de l'esprit, a mené des réflexions sur le langage et le contrôle moteur qui me paraissent fondamentales et encore inspirantes aujourd'hui.

C'est aussi une façon de goûter cette langue qu'on dit morte, dont la comparaison avec le français et les autres langues romanes, nous renseigne sur les mécanismes évolutifs qui façonnent les langues du monde.

C'est, enfin, la reconnaissance, à la suite de André Jacob (1967/1992, p. 346) et de Gabriel Bergounioux (2004), que, comme le verbe latin *loquor* qui est déponent, « parler » ou « dire », en français, peuvent s'employer dans une diathèse moyenne. C'est, justement, le « *moyen de parler* » qui pose question en science du langage, car comme le remarque Bergounioux (2004, p. 60 et suivantes), si l'être humain est locuteur, il est souvent aussi allocuteur de sa propre parole : « je parle », « je me parle » (plus rare et socialement censuré) ou « je dis », « je me dis » (usuel). André Jacob insiste aussi sur le caractère déponent du langage (Brachet, 2007), qui ne serait « ni une activité pure, ni une passivité pure » (on retrouve peut-être le « ça parle » de Lacan). Cette irruption de la grammaire latine incite donc à prendre en considération les fonctions du langage, ses usages et ainsi à convoquer la philosophie dans la science du langage.

La citation du sculpteur Alberto Giacometti n'est pas non plus anodine. Je tente modestement de résumer ici les travaux de recherche que j'ai menés pendant déjà vingt ans, depuis mon entrée au CNRS en 1998, à l'Institut de la Communication Parlée (devenu Département Parole et cognition de GIPSA-lab) puis, depuis 2012, au Laboratoire de Psychologie et NeuroCognition. Je présente les cheminements, essais, esquisses, qui m'ont conduite petit à petit à sculpter des ébauches théoriques que j'espère faire tenir debout quelques temps, tout en sachant qu'il faudra les reprendre, encore et encore.

Je défends l'hypothèse que le langage est le fondement de l'humanité car il sous-tend la communication, la pensée et l'autoconscience, ou la conscience de soi dans le temps. Ces trois fonctions – communicative, cognitive et métacognitive – s'appuient sur le langage, sont entrelacées de façon systémique et requièrent une cybernétique commune qui déploie le suivi de soi et d'autrui et la régulation des messages verbaux produits et perçus :

- régulation et suivi de notre propre flux verbal et de ceux d'autrui dans la communication,
- contrôle de nos productions verbales intérieures pour l'élaboration de la pensée,
- suivi de soi et d'autrui dans nos constructions narratives autobiographiques pour l'autoconscience.

Mes travaux relèvent, de façon générale, de cette cybernétique verbale. Je cherche à mieux comprendre les mécanismes neurobiologiques qui sous-tendent la régulation de la production et de la réception du langage (dans ses versions extériorisée et intérieure).

Ma démarche est triple.

Tout d'abord, je mène une activité théorique et ai commencé à développer un formalisme neurocomputationnel visant à décrire les pressions communicatives, cognitives et métacognitives qui s'exercent sur le langage et à mieux comprendre les fondements neurophysiologiques de la cybernétique verbale qui en découle. Ces essais de formalisation théorique sont présentés dans le Chapitre I.

En parallèle, je suis une démarche empirique, qui me permet de tester les premières ébauches de ce formalisme, et de leur apporter des corrections ou de formuler de nouvelles hypothèses ou questions. J'ai recueilli ainsi des données comportementales et neurophysiologiques sur la production ou la perception du langage, chez l'adulte, au cours du développement chez l'enfant et dans des populations souffrant de troubles du langage, de l'audition ou de pathologies neurologiques ou psychiatriques (aphasie, schizophrénie). Ces travaux expérimentaux sont détaillés dans le Chapitre II.

Enfin, j'ai commencé à envisager des applications éducatives, thérapeutiques et technologiques de ces propositions théoriques et de ces résultats expérimentaux, dans le domaine de la remédiation des troubles du langage. Ces applications visent à maintenir, rétablir ou réparer la triade fonctionnelle communication-pensée-autoconscience. Elles sont décrites de façon succincte dans le Chapitre III.

Le dernier chapitre présente le programme de recherche que j'envisage pour les années à venir. Je propose de contribuer à mieux caractériser les substrats neurophysiologiques de la cybernétique verbale qui sous-tend la triade systémique communication-cognition-autoconscience. J'envisage notamment de recueillir des données en électrophysiologie temporellement et spatialement plus précises, d'examiner plus en détail les étapes du développement de la parole chez l'enfant, de mener des enquêtes et des expériences plus approfondies sur la parole intérieure et ses liens avec le raisonnement et la mémoire épisodique, en m'intéressant à diverses populations. Je compte enfin m'investir davantage dans les applications thérapeutiques et éducatives de ces travaux, notamment dans les domaines du développement, de la remédiation et de la rééducation du langage.

Chapitre I.

ESSAIS DE FORMALISATION THEORIQUE

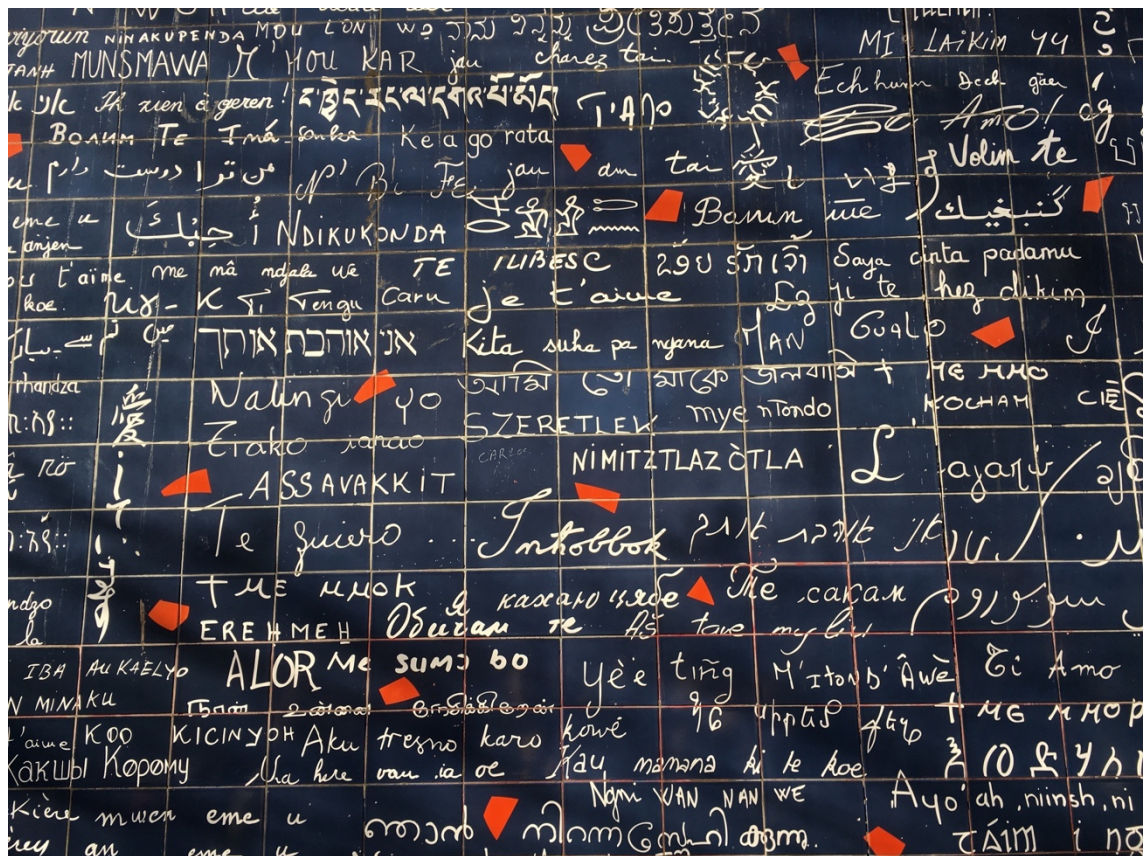


Figure I-1. Le mur des Je t'aime, à Paris XVIIIème

Cette photographie du « mur des Je t'aime », l'œuvre de Frédéric Baron et Claire Kito, inaugurée en 2000, square Jehan Rictus, place des Abbesses à Montmartre, illustre galamment la complexité et la diversité des langues du monde. Je t'aime. Ce sentiment si universel s'exprime de façon extraordinairement multiple dans les quelque sept mille langues parlées aujourd'hui (Anderson, 2010 ; Hagège, 1985 ; Ladefoged, 2003).

Si l'on écoute les formes sonores des langues du monde, environ 600 consonnes, 200 voyelles et de multiples tons ont été inventoriés (Ladefoged, 2003 ; Ladefoged & Maddieson, 1996 ; Maddieson, 2013) dont une grande partie se trouve déclinée et transcrite sur le mur de Paris. Sélectionnons, au hasard des tendresses, sept langues orales : « je t'aime », « I love you » (anglais), « я тебя люблю » (russe), « ich liebe dich » (allemand), « rwi'n dy garu di » (gallois), « eni a hnimi eö » (drehu), « dama la bëgg » (wolof). Rien que dans ce millième des langues du monde, et pour un même sentiment exprimé, on observe une gamme de sons étonnante. On trouve des consonnes fricatives post-alvéolaires comme /ʒ/ dans « je », labiodentales comme /v/ dans « love », palatales comme /ç/ dans « ich » et glottales comme /h/ dans « hnimli », des occlusives bilabiales comme /b/ dans « liebe, люблю, бëgg », alvéolaires comme /t, d/ « t'aime, тебя, дy di, даma », vélaires comme /g/ dans « garu, бëgg », des nasales comme /m, n/ dans « t'aime, rwi'n, eni, hnimi, даma », des latérales /l/ dans « love, liebe, люблю, la », des spirantes alvéolaires comme /ʎ/ dans « rwi'n, garu », labio-vélaires comme /w/ dans « rwi'n » ou palatales comme /j/ dans « я тебя люблю », des géminées comme /gg/ dans « bëgg ». On découvre des voyelles antérieures, postérieures, arrondies, étirées, tendues, relâchées, longues et brèves. Pour dire « je t'aime », les humains ont développé toute une panoplie de configurations articulatoires, d'accolement des cordes vocales, de positions complexes et coordonnées du larynx, du voile du palais, de la langue et des lèvres. Une source d'inspiration pour les psychanalystes...

Si l'on se penche maintenant sur le lexique et l'étymologie, la variété est grande également. C'est la racine indo-européenne *leub^h- (amour) qui a donné « love, люблю et liebe », ce serait *am- (maman / saisir) ou *sem (un, uni) qui aurait donné « aime ». Ce serait la racine proto-indo-européenne *keh²- (désirer) qui aurait donné le gallois « garu ou caru » et c'est probablement une racine nigéro-congolaise signifiant « vouloir » qui aurait donné le wolof « bëgg » (cf. <http://starling.rinet.ru/babel.php?lan=en> ou <https://www.etymonline.com/>).

Sur le plan de la morphologie, on observe aussi une diversité foisonnante. Dans les langues isolantes, le verbe « aimer » est invariable, comme en vietnamien (yêu / anh yêu em : aimer / je t'aime (d'un homme à une femme) et yêu / em yêu anh : aimer / je t'aime (d'une femme à un homme)). Dans les langues agglutinantes, on peut accoler au radical du verbe un grand nombre de suffixes qui expriment la possibilité, la négation, le temps, la personne, le mode, la diathèse (voix passive ou active), l'aspect (progressif, aoriste), comme en turc (sevmek / seni seviyorum : aimer / je t'aime). Dans les langues flexionnelles (ou synthétiques), le radical du verbe peut varier en fonction du temps ou de la personne (comme en français : aller/je vais, nous allons, j'irai) et être complété de désinences, marquant la personne, le temps, l'aspect, le mode, la diathèse. Les pronoms qui accompagnent ce verbe peuvent eux-mêmes porter ou non la marque du cas (nominatif, accusatif) ou du genre, selon les langues.

La diversité se retrouve aussi dans la syntaxe. L'ordre des mots dans la phrase est variable selon les langues : le verbe pouvant se trouver en position initiale, centrale ou finale dans

l'énoncé (oral ou gestuel). Les structures de phrases sont elles-mêmes diverses. Dans les langues ergatives, comme le basque, une construction spécifique est employée pour associer un verbe transitif à son agent. Si la langue a des déclinaisons, l'agent (le sujet) d'un verbe transitif se met à l'ergatif et l'objet (le patient) se met à l'absolutif (en basque : *maite zaitut* : « je vous ai comme mon aimé-e », « maite » signifiant « aimé-e » et le « t » final du verbe « zaitut » indiquant le cas ergatif de « je », la racine du verbe avoir étant « u » et « z » indiquant le cas absolutif de « vous », « a » indiquant le présent, « it » étant un infixé lié à l'emploi de vous). Dans les langues accusatives, comme le japonais, c'est le patient (l'objet) du verbe qui porte une marque (en japonais, *watashi ha anata ga suki desu* : littéralement « moi toi aimer », « watashi » : moi, je, « ha » : particule d'ouverture du sujet, « anata » : toi, « ga » : marque de l'accusatif, « suki desu » : aimer).

Si l'inventaire oral est riche, la variété gestuelle est elle-même impressionnante mais n'est pas une transposition des variations de l'oral. Ainsi, alors que les locuteurs de l'anglais oral américain et britannique se comprennent, la BSL (*British Sign Language*) n'est pas intelligible pour les utilisateurs de l'ASL (*American Sign Language*). Dans la langue des signes française (LSF), *je t'aime* se dit en posant la main à plat sur le torse, en la glissant vers le haut puis en la dirigeant vers l'avant, la paume se tournant vers le ciel, tout en regardant la personne aimée (je te donne mon cœur). En ASL, pour signer « je t'aime », on pointe sur son torse avec l'index d'une main, puis on ferme les poings des deux mains et on les croise au-dessus de son cœur, comme pour prendre quelqu'un dans ses bras, puis on désigne avec l'index la personne aimée. On peut aussi signer la suite des lettres I.L.Y : en fermant le point puis en relevant l'auriculaire pour faire « I », puis l'index et le pouce à sa suite, pour créer L et le Y. En langue des signes japonaise, on pointe avec l'index vers son propre nez, puis de la même main, on désigne la personne aimée en la pointant du doigt, puis on ferme le poing et on frotte le dos de cette main fermée avec l'autre main à plat, avec des gestes circulaires.

Enfin sur le plan de la transcription écrite, l'œuvre illustre quelques-unes des écritures que nous utilisons de nos jours, notamment les alphabets latin, cyrillique, grec, arménien, géorgien, hébraïque, arabe, perso-arabe, tibétain, cinghalais, coréen (hangul), les alphasyllabaires éthiopien, devanāgarī (hindi), bengali, tigrigna, tamoul et khmer, les caractères chinois, les kanji, hiragana et katakana japonais, les signes de la LSF, les caractères braille.

Pourquoi tant de formes sonores, gestuelles et écrites pour un même sentiment ?

Sur le plan cérébral, la complexité est encore à l'œuvre. Alors qu'un système moteur vocal primaire (PVMN, *Primary Vocal Motor Network*), impliquant un réseau de structures sous-corticales peut suffire à émettre des vocalisations chez les primates non-humains (cf. Hage & Nieder, 2016), la production de la parole et des signes chez l'être humain engage un réseau de régions sous-corticales, corticales et cérébelleuses d'une grande complexité.

Pourquoi une machinerie cérébrale aussi sophistiquée pour émettre des sons et des gestes ?

Je défends ci-dessous l'hypothèse que la diversité sonore et gestuelle des langues du monde, ainsi que la complexité de l'architecture neuronale sous-tendant le langage, sont issues de pressions évolutives, qui permettent aujourd'hui au langage de remplir trois fonctions primordiales : la communication, la pensée et l'autonoèse.

I.1. TROIS FONCTIONS PRIMORDIALES DU LANGAGE

« *Le langage est l'instrument par lequel l'homme façonne sa pensée, ses sentiments, ses émotions, ses efforts, sa volonté et ses actes, l'instrument grâce auquel il influence et est influencé, l'ultime et le plus profond fondement de la société humaine. Mais, il est aussi le dernier, l'indispensable recours de l'homme, son refuge aux heures solitaires où l'esprit lutte avec l'existence, et où le conflit se résout dans le monologue du poète et la méditation du penseur* » (Hjelmslev, 1943/1971, c'est moi qui souligne).

Dans ces quelques lignes, Hjelmslev évoque élégamment les trois fonctions du langage auxquelles je m'intéresse ici, illustrées sur la Figure I-2 : la fonction sociale de communication avec autrui, la fonction cognitive d'élaboration et d'expression de la pensée, et la fonction métacognitive d'autoconscience, i.e. de construction de soi, la possibilité de se concevoir et de s'attribuer une identité, au cours du temps.

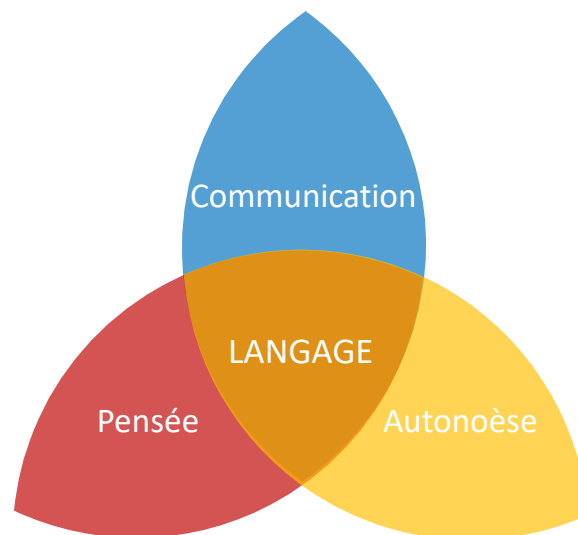


Figure I-2. Les trois fonctions du langage étudiées ici

Notons d'abord que le terme de fonction est pris ici au strict sens biologique, c'est-à-dire au sens de la simple description de ce que fait le langage, son rôle, sa contribution, et non pas au sens téléologique de « but du langage », de sa finalité intentionnelle, ni au sens évolutionniste de « fonction primaire », qui serait la raison pour laquelle le langage non seulement existerait actuellement mais aussi pour laquelle il aurait perduré au cours de l'évolution, la raison pour laquelle il aurait été sélectionné (pour des débats sur les notions de fonction et d'explication fonctionnelle en philosophie des sciences, voir par exemple Nagel, 1961 ; Cummins, 1975 ; Millikan, 1989).

Notons également que les six fonctions du langage identifiées par Jakobson (1963) peuvent être considérées comme des sous-fonctions de la fonction communicative. La fonction *expressive* (ou *émotive*) permet l'expression des pensées ou des sentiments des locuteurs, elle vise à communiquer à autrui l'attitude du sujet à l'égard de ce dont il parle (« Bof ! »). La fonction *conative* est utilisée par l'émetteur du message pour induire chez le récepteur un certain comportement (« Dépêche-toi ! »). Elle a donc elle aussi trait à la communication inter-locuteurs. La fonction *phatique* est utilisée pour établir, maintenir ou interrompre le contact physique et psychologique avec le récepteur (« Allô ? »). Elle

permet donc de s'assurer que la communication sera effective. La fonction *métalinguistique* consiste à utiliser le langage pour expliquer le langage, i.e. de s'assurer que le code utilisé dans la communication est bien partagé par les interlocuteurs (« Qu'entendez-vous par fonction ? »). La fonction *référentielle* ou *dénotative* décrit une réalité objective, fournit une information sur une situation, oriente la communication vers ce dont l'émetteur parle (« Jakobson était un linguiste russo-américain »). Enfin la fonction *poétique* permet de faire du message un objet esthétique, dans le domaine poétique ou littéraire, mais aussi dans les communications quotidiennes (« Jakobson ? va lui ouvrir ! » ou « Saussure, c'est le pied ! »). Ces six fonctions sont toutes reliées à la communication, elles sont donc pertinentes ici, mais j'en ajouterai d'autres, ayant trait aux fonctions de pensée et d'autoïèse.

I.1.1. La fonction communicative : *loquor, ergo communico*

La première fonction évidente du langage, que le sens commun reconnaît immédiatement (comme le note Searle, 1972), est une fonction sociale de communication, le langage nous permettant de transmettre à autrui nos pensées, nos émotions et nos sentiments et de recevoir ceux de nos congénères. Comme le constate Asoulin (2016), l'hypothèse que la fonction primordiale du langage soit communicative est défendue par la plupart des linguistes, philosophes, psychologues et spécialistes de sciences cognitives contemporains. S'y rallient, par exemple, Martinet (1960, 1989), Fromkin, Rodman & Hyams (1983), Deacon (1997), Jackendoff (2002, 2011), Millikan (2005) ou Scott-Philips (2015). Cette position était déjà avancée par Locke en 1690 :

« Lorsqu'un homme parle à un autre, c'est afin de pouvoir être entendu ; le but du langage est que ces sons ou marques puissent faire connaître les idées de celui qui parle à ceux qui l'écoutent. » (Locke, 1690, c'est moi qui souligne).

Cette fonction de communication interhumaine est parfois distinguée de la fonction expressive du langage, qui serait personnelle, et qui permet de manifester les affects, les sentiments, les passions, les idées (Lucrèce, livre V, cité par Ballet, 1886, p.2-3 ; Rousseau, 1781 ; Mounin, 1968). Je préfère ne pas retenir cette distinction, car la fonction expressive est également communicative, dans la mesure où elle est manifestation. Je distingue cependant la fonction expressive de la fonction d'élaboration et d'expression de la pensée, détaillée dans le paragraphe I.1.2, qui sous-entend un processus de construction conceptuelle, non inclus dans la notion de communication. La fonction de communication interpersonnelle du langage est éminemment sociale, ainsi que l'explique Saussure (1916/1995) :

« [la langue] n'existe qu'en vertu d'une sorte de contrat passé entre les membres de la communauté » (p. 31).

Elle est, selon Saussure, le lien social qui relie les individus d'une communauté. Dans cette même lignée, Jakobson (1963) explicite et schématise la fonction de communication :

« (...) tout acte de parole met en jeu un message et quatre éléments qui lui sont liés : l'émetteur, le receveur, le thème (topic) du message, et le code utilisé. (...) Je pense que la réalité fondamentale à laquelle le linguiste a affaire, c'est l'interlocution — l'échange de messages entre émetteur et receveur, destinataire et destinataire, encodeur et décodeur. » (Jakobson, 1963, p. 28-29 et p.32).

D'après le préhistorien Leroi-Gourhan (1964), la fonction de communication du langage se serait développée selon trois phases successives : une phase de communication pratique, une phase de communication indirecte (le récit), et une dernière phase où le langage permet d'exprimer les sentiments et ses idées abstraites :

« Le langage du Néanderthalien ne devait pas différer beaucoup du langage tel qu'il est connu chez les hommes actuels. Essentiellement lié à l'expression du concret, il devait assurer la communication au cours des actes, fonction primordiale où le langage est étroitement lié au comportement technique ; il devait aussi assurer la transmission différée des symboles de l'action sous forme de récits. Cette seconde fonction a dû émerger progressivement chez les Archanthropiens, mais il est difficile d'en faire la démonstration. Enfin, au cours du développement des Paléanthropiens, apparaît une troisième fonction, celle dans laquelle le langage dépasse le concret et le reflet du concret pour exprimer des sentiments imprécis dont on sait à coup sûr qu'ils entrent pour une part dans la religiosité. » (A. Leroi-Gourhan, 1964).

Cette fonction sociale de communication est jugée essentielle par certains auteurs. Selon Martinet (1960), par exemple, la langue est « *un instrument de communication doublement articulé et de manifestation vocale* ». La fonction du langage étant de communiquer, le langage serait donc structuré selon des contraintes de pertinence : les énoncés seraient par conséquent déterminés par la production du sens et le transport de l'information. C'est aussi l'hypothèse de Searle (1972), selon qui toute théorie du langage doit tenir compte de sa fonction, car cette fonction communicative détermine la structure du langage. On ne peut pas étudier la structure du langage indépendamment de cette fonction de communication :

« The purpose of language is communication in much the same sense that the purpose of the heart is to pump blood. In both cases it is possible to study the structure independently of function but pointless and perverse to do so, since structure and function so obviously interact. » (...) « Any attempt to account for the meaning of sentences must take into account their role in communication, in the performance of speech acts, because an essential part of the meaning of any sentence is its potential for being used to perform a speech act. »¹

Millikan (2005) considère même que la fonction de communication du langage est une fonction fondamentale, primaire, originelle, au sens évolutionniste :

« I will argue (...) that a primary function of the human language faculty is to support linguistic conventions, and that these have an essentially communicative function »²

Ainsi, selon Millikan, le langage est par essence destiné à la communication interpersonnelle, il a évolué pour permettre aux individus de communiquer entre eux.

¹ « La fonction du langage est la communication tout comme la fonction du cœur est de pomper le sang. Dans les deux cas, il est possible d'étudier la structure indépendamment de la fonction, mais il est vain et absurde de le faire, puisque la structure et la fonction interagissent de façon si évidente. (...) Toute tentative de rendre compte du sens des phrases se doit de tenir compte de leur rôle dans la communication, dans l'accomplissement d'actes de langage, car une part essentielle du sens de toute phrase réside dans son potentiel à être utilisée pour accomplir un acte de langage. »

² « Je défendrai l'hypothèse qu'une fonction primaire de la faculté humaine de langage est de soutenir les conventions linguistiques, et que celles-ci ont une fonction communicative par essence. »

Faire l'hypothèse que la fonction essentielle du langage est communicative revient à postuler que le langage est un simple moyen de communiquer la pensée et qu'il ne joue pas un rôle central dans la cognition, autrement dit que la pensée est pré-existante au langage. Russell (1897/1989) pousse cette hypothèse plus loin et considère même que le langage ordinaire entrave la pensée, qu'il est un obstacle à l'exercice philosophique car les mots usuels du langage sont trop ambigus et trompeurs pour représenter adéquatement le monde physique. Bergson (1934) fait également l'hypothèse que la fonction primitive du langage « *est d'établir une communication en vue d'une coopération. Le langage transmet des ordres ou des avertissements. Il prescrit ou il décrit.* ». Cette fonction primitive est utilitaire et liée aux nécessités de l'action. Cependant Bergson (1889) note qu'elle est inapte à décrire la réalité car, pour les sentiments notamment, les mêmes mots désignent des états différents, propres à la personnalité de chacun. Ainsi explique-t-il que « nous échouons à traduire entièrement ce que notre âme ressent : la pensée demeure incommensurable avec le langage. ». L'hypothèse que la pensée est indépendante du langage et que celui-ci est un simple medium (parfois inadapté et réducteur) de communication correspond, selon Carruthers (2002) et Pelletier (2004), au modèle standard en sciences cognitives, dans lequel l'esprit humain est constitué de modules autonomes et isolables (et innés). Dans ce courant de pensée, que Carruthers nomme la « Conception Communicative » du langage, le langage est considéré comme le module spécialisé dans la production et l'interprétation d'énoncés, parmi d'autres modules qui permettent le raisonnement et la pensée. Le langage n'est considéré que comme un module d'entrée-sortie, spécialisé dans l'interprétation et la production d'énoncés. Il est un simple conduit par lequel transitent les pensées d'un individu à l'autre, vers l'intérieur ou vers l'extérieur de l'esprit. La pensée elle-même se produirait en dehors du langage, dans un autre mode de représentation.

Encadré I.1. La Fonction Communicative du langage : *Loquor, ergo communico*

En somme, selon certains théoriciens, le langage joue une fonction essentiellement communicative : il nous permet d'échanger avec autrui des pensées, des émotions, des sentiments et il a évolué pour cette fonction ; le langage ne jouerait pas un rôle central dans la cognition, il serait un moyen (d'ailleurs imparfait) de communiquer la pensée, qui serait elle-même pré-existante au langage. Cette hypothèse peut ainsi se résumer en la formule suivante : « je parle, donc je communique ».

L'hypothèse selon laquelle le langage est *avant tout* un système de communication interpersonnelle n'est cependant pas partagée par tous les linguistes ou philosophes. Comme nous allons l'aborder, pour certains, la fonction essentielle et première du langage n'est pas la communication avec autrui, mais plutôt celle de construction et d'expression de la pensée, qui ne peut s'envisager comme indépendante du langage.

I.1.2. La fonction cognitive d'élaboration et d'expression de la pensée : *loquor, ergo cogito*

Si une première fonction évidente du langage paraît être une fonction sociale de communication, le rôle primordial du langage dans l'élaboration de la pensée a été perçu dès les penseurs égyptiens. Maspero (1875) remarque que, dans la tradition des savants égyptiens de l'Ancien Empire (III^{ème} millénaire av. J.C.), l'action de penser est

intrinsèquement liée à celle de parler. C'est ainsi qu'il cite le hiéroglyphe de la Figure I-3, qui est un déterminatif représentant les actions de la bouche et exprimant indifféremment les notions de manger, boire, crier, parler, penser, méditer, connaître, juger (Maspero, 1875, p. 223 ; cf. aussi Gardiner, 1927, code A2 « homme assis portant main à la bouche »). Pour les égyptiens de l'Ancien Empire, penser est donc parler.



Figure I-3. Déterminatif de l'égyptien ancien qui représente les actions de la bouche et qui exprime indifféremment les idées de manger, boire, crier, parler, penser, méditer, connaître, juger. Tiré de Maspero (1875).

Pour les philosophes grecs, le *logos* (λόγος), qui signifie au départ le discours (parlé ou écrit), la parole, le langage, le verbe, désigne aussi par extension, la pensée, le concept, la raison. Le concept de *logos*, s'il a recouvert des acceptions différentes, de Héraclite à Aristote en passant par Platon, a toujours impliqué une unité voire une identité entre le langage et la pensée (cf. Sokolov, 1972, chapitre 1). Ainsi dans *Théétète*, Platon explicite la position de Socrate selon laquelle penser est une « *discussion que l'âme elle-même poursuit tout du long avec elle-même à propos des choses qu'il lui arrive d'examiner. (...) Car voici ce que me semble faire l'âme quand elle pense : rien d'autre que dialoguer, s'interrogeant elle-même et répondant, affirmant et niant. Et quand, ayant tranché, que ce soit avec une certaine lenteur ou en piquant droit au but, elle parle d'une seule voix, sans être partagée, nous posons que c'est là son opinion. De sorte que moi, avoir des opinions, j'appelle cela parler, et que l'opinion, je l'appelle un langage, prononcé, non pas bien sûr à l'intention d'autrui ni par la voix, mais en silence à soi-même.* » (Platon, *Théétète*, 189e-190a). En somme, chez les philosophes grecs anciens, penser équivaut à dialoguer avec soi-même.

I.1.2.1. Unité entre pensée et langage : le langage comme reflet et expression de la pensée

La philosophie moderne reprend cette notion d'unité entre pensée et langage. Pour Descartes, par exemple, le langage n'a d'autre fonction que d'exprimer la pensée : il dérive de la pensée, proprement humaine.

« Ce langage est en effet le seul signe certain d'une pensée latente dans le corps ; tous les hommes en usent, même ceux qui sont stupides ou privés d'esprit, ceux auxquels manquent la langue et les organes de la voix, mais aucune bête ne peut en user ; c'est pourquoi il est permis de prendre le langage pour la vraie différence entre les hommes et les bêtes. » (Descartes, Lettre à Morus du 5 février 1649).

Descartes exclut du langage « *les cris de joie ou de tristesse, et semblables* » et « *aussi tout ce qui peut être enseigné par artifice aux animaux* » (Descartes, 1646). Pour lui, le langage est la manifestation de la pensée élaborée humaine et ne doit pas être assimilé à la simple communication (qui peut exister entre animaux, tel que le défend Montaigne).

Ce lien fort entre langage et pensée élaborée se retrouve dans « la Logique de Port-Royal » (1662) :

« L'esprit a coutume d[e] lier si étroitement [les sons aux idées] que l'idée de la chose excite l'idée du son, et l'idée du son celle de la chose. On peut dire en général sur ce sujet, que les mots sont des sons distincts et articulés, dont les hommes ont fait des signes pour marquer ce qui se passe dans leur esprit. Et comme ce qui s'y passe se réduit à concevoir, juger, raisonner et ordonner, [...], les mots servent à marquer toutes ces opérations ». (Arnauld & Nicole, 1662/1878, p. 107-108).

Selon les jansénistes Arnauld et Nicole, les mots sont les signes de la pensée et le langage la représente. « L'art de penser » se décline en 4 opérations (« *concevoir, juger, raisonner et ordonner* ») qui sont reflétées dans l'expression verbale. Le langage est ainsi le miroir de la pensée. Dans cette conception, le langage apparaît donc d'abord comme un outil servant à l'expression de la pensée avant d'être un instrument de communication. On retrouve une affirmation similaire chez Leibniz :

« (...) je suis persuadé que les langues sont le meilleur miroir de l'esprit humain et qu'une analyse exacte de la signification des mots ferait mieux connaître que toute autre chose les opérations de l'entendement. » (Leibniz, 1765/1921, 3.7.6).

Une unité entre pensée et langage a donc été perçue depuis les penseurs égyptiens de l'Ancien Empire jusqu'aux philosophes européens du XVIIIème siècle. Mais, comme je le détaille ci-après, Leibniz lui-même a questionné plus avant les liens entre pensée et langage.

I.1.2.2. Le langage comme outil de construction et de structuration de la pensée

Comme le défend Parmentier (2014), pour Leibniz « *les mots sont des " caractères " destinés aussi bien, sinon plus, à l'élaboration des pensées qu'à leur communication.* ». Leibniz perçoit le lien entre pensée et langage, entrevu déjà chez Descartes ainsi qu'Arnauld & Nicole, mais il le décrit de façon plus précise. Selon lui, la pensée ne précède pas le langage, comme le défendait Descartes, le langage n'est pas l'expression d'idées préalablement conçues, mais il construit la pensée :

« Je crois qu'en effet sans le désir de nous faire entendre nous n'aurions jamais formé de langage ; mais étant formé, il sert encore à l'homme à raisonner à part soi, tant par le moyen que les mots lui donnent de se souvenir des pensées abstraites que par l'utilité qu'on trouve en raisonnant à se servir de caractères et de pensées sourdes ; car il faudrait trop de temps s'il fallait tout expliquer et toujours substituer les définitions à la place des termes. » (Leibniz, 1765, liv. III, p. 223).

Près de deux siècles plus tard, au fondement de la linguistique contemporaine, on retrouve cette conception du lien entre langage et élaboration de la pensée chez Saussure, même si celui-ci donne à la fonction de communication un rôle fondamental (cf. paragraphe I.1.1 ci-dessus) : « *ce n'est pas la pensée qui crée le signe, mais le signe qui guide primordialement la pensée* » (Saussure, posthume 2002, 46). Merleau-Ponty (1945) affirme même que la pensée n'existerait pas sans le langage. C'est, selon lui, la parole qui permet de construire, d'élaborer la pensée. « *La pensée n'est rien 'd'intérieur', elle n'existe pas hors du monde et hors des mots* », elle coïncide avec l'acte d'expression, avec la gesticulation verbale, même intérieure. Contrairement à Bergson qui affirme que la pensée existerait pour soi avant l'expression, Merleau-Ponty considère la parole

comme un geste, un processus sensori-moteur, qui constitue et qui actualise le sens. Pensée et expression seraient ainsi élaborées simultanément.

Ryle (1972) explicite de même la façon dont l'acte de se parler peut permettre d'élaborer sa pensée, de mieux saisir les choses :

*« Thinking, then, can be saying-things-tentatively-to-oneself with the specific heuristic intention of trying, by saying them, to open one's own eyes, to consolidate one's own grasp, or to get oneself out of a rut, etc. »*³

Le rôle du langage, et en particulier de la parole intérieure, dans la construction de la pensée a été envisagé très tôt chez les psychologues soviétiques, comme Vygotski (1934/1985) ou Sokolov (1972), avec la perspective philosophique matérialiste que « *le langage est l'immédiate réalité de la pensée* » (Marx & Engels, 1846) et que « *les idées n'ont pas d'existence en dehors du langage* » (Marx & Engels, 1857). Sokolov (1972) affirme ainsi que la pensée humaine est, par essence, une pensée verbale, la parole étant non seulement un moyen d'expression de la pensée mais aussi et de façon primordiale, le moyen de sa formation et de son développement, le médium de l'analyse, de la synthèse, de l'abstraction et de la généralisation.

I.1.2.3. Le langage comme système dédié prioritairement à la pensée

Cette idée que le langage est un instrument permettant la pensée est défendue par Chomsky (1969), qui se place dans la lignée de Descartes, et intègre les conceptions leibniziennes. Il traduit ainsi les positions de Descartes en termes contemporains :

*« We have seen that the Cartesian view, as expressed by Descartes and Cordemoy as well as by such professed anti-Cartésiens as Bougeant, is that in its normal use, human language is free from stimulus control and does not serve a merely communicative function, but is rather an instrument for the free expression of thought and for appropriate response to new situations »*⁴ (Chomsky, 1969, p. 13).

Selon Chomsky, Descartes a remarquablement décrit le caractère « spécial » du langage humain. Ce qui distingue le langage humain de la communication des animaux et de tous les autres systèmes de communication (les automates), c'est la créativité, la faculté de générer des énoncés nouveaux, non encore formulés par autrui, dans des contextes nouveaux. Chomsky interprète la linguistique cartésienne en termes leibniziens, proposant que la fonction d'expression de la pensée que joue le langage n'est pas simplement la traduction de la pensée pour autrui (qui serait donc une opération de communication), mais aussi l'opération de structuration de la pensée, pour soi-même. Ainsi, Chomsky développe et prolonge la conception cartésienne du langage en affirmant que le langage a évolué non pas pour permettre aux individus de communiquer entre eux mais, avant tout, pour permettre la pensée élaborée. Selon Chomsky (1977),

³ « Penser, donc, cela peut être se-dire-les-choses-provisoirement avec l'intention heuristique spécifique d'essayer, en les disant, de s'ouvrir les yeux à soi-même, de consolider sa propre compréhension ou de se sortir de l'ornière, etc. »

⁴ « Nous avons vu comment dans la conception cartésienne, telle qu'elle a été exprimée par Descartes et Cordemoy ainsi que par des anti-cartésiens revendiqués tels que Bougeant, le langage, dans son usage ordinaire, est affranchi de tout contrôle par le stimulus, sa fonction n'est pas simplement communicative, il est plutôt un instrument servant à l'expression libre de la pensée, et capable de fournir une réponse appropriée à des situations nouvelles »

contrairement aux propositions de Searle (1972), le langage sert principalement à penser et secondairement à communiquer. Chomsky (2002) écrit même :

« *Language is not properly regarded as a system of communication. It is a system for expressing thought, something quite different . . . language use is largely to oneself: "inner speech" for adults, monologue for children.* » (Chomsky 2002, pp. 76 - 77).

Plus loin, de façon peu prudente, comme le fait remarquer à juste titre Wiley (2014), il ajoute :

« *Inner speech is most of speech. Almost all the use of language is to oneself.* » (Chomsky, 2002, p. 148).

Et cette idée de primauté du langage intérieur sur la parole externalisée est de nouveau avancée dans un article sur le programme biolinguistique : « *Statistically speaking, for whatever that is worth, the overwhelming use of language is internal—for thought. It takes an enormous act of will to keep from talking to oneself in every waking moment—and asleep as well, often a considerable annoyance.* » Berwick & Chomsky (2011, p.25-26).

Ces propos sont assez peu convaincants (cf. I.1.2.4.4), mais sans adhérer à l'idée extrême que le langage est quasi-exclusivement employé pour la pensée, l'hypothèse plus nuancée qu'il ait évolué principalement pour la pensée reste attrayante. Cette conception du langage comme un système représentationnel de la pensée est d'ailleurs partagée par Bickerton (1990) ou Pinker & Bloom (1990) par exemple, selon qui l'organisation du langage reflète non pas les besoins ou les caractéristiques de la communication interpersonnelle mais les propriétés structurelles des concepts (voir Newmeyer, 2016).

Cette primauté de la fonction d'expression des idées sur celle de communication est explicitée également par Reboul (2015, 2017), dans une perspective évolutionniste. Selon cette auteure, la plupart des théories contemporaines sur l'évolution du langage (notamment la théorie codique de la communication de Millikan, 2005 et la théorie de la communication ostensive de Scott-Philips, 2015) reposent sur une hypothèse majeure, suivant laquelle le langage est un système de communication dans le sens fort, c'est-à-dire ayant évolué pour satisfaire des besoins de communication sociale et ayant une visée coopérative altruiste. Toutefois, si le langage est, en effet, utilisé dans la communication humaine interpersonnelle, cela ne signifie pas qu'il a évolué *pour* la communication. Une autre hypothèse, défendue par Reboul, est que le langage a évolué pour d'autres raisons et qu'il a ensuite été exapté pour la communication. Il serait donc un système de communication dans le sens faible. En effet, la description du langage comme un système de communication dans le sens fort va à l'encontre des caractéristiques structurelles spécifiques du langage humain. Selon Reboul (2015, 2017), le langage est excentrique parmi les systèmes de communication animaux, parce que c'est le seul à incorporer l'infinité discrète (la production d'une infinité de signaux linguistiques différents à partir d'un nombre fini d'unités discrètes), la sémantacité (la capacité des signaux linguistiques à transmettre des messages porteurs de sens) et le découplage (qui permet aux locuteurs de parler d'objets absents ou non-existants). Ces trois caractéristiques structurelles donnent lieu à la créativité linguistique (la faculté de produire un nombre infini de phrases dans des contextes différents). Cette créativité ne se retrouve pas dans la communication animale et ne semble pas une condition nécessaire à la communication. Comment expliquer, dans une hypothèse de communication au sens fort, que le langage

humain présente cette propriété de créativité ? De plus, dans l'hypothèse du langage comme un système de communication au sens fort, le message correspond à ce qui est communiqué, i.e., à la signification du·de la locuteur·rice. Or, de par la propriété de découplage du langage notamment, la signification du·de la locuteur·rice ne se réduit généralement pas à la signification de la phrase, elle n'est pas récupérable par simple composition sémantique. Elle dépend du contexte et de l'intention du·de la locuteur·rice. Le découplage permet également la duperie et le mensonge, qui ne sont pas expliqués par la pression sélective en vue d'une communication altruiste réussie. Selon Reboul, ce ne sont donc pas les besoins de la communication qui ont façonné le langage, mais la pensée humaine, proprement créative et sophistiquée. Dans la lignée de Fodor & Pylyshyn (2015), Reboul remarque que pensée et langage partagent la même organisation structurelle : tout comme les phrases organisent structurellement les mots d'une façon créative, les pensées sont elles-mêmes composées structurellement de concepts, d'une façon créative. Ainsi, et comme le conçoit Chomsky (2014), Reboul suppose que le langage n'a pas évolué pour permettre la communication mais pour structurer les concepts et rendre possible une pensée sophistiquée. Le langage, cet instrument de la pensée, aurait *ensuite* été externalisé pour la communication avec autrui et sa version externalisée aurait hérité des propriétés inhérentes à une pensée élaborée.

Encadré I.2. La Fonction Cognitive du langage : *Loquor, ergo cogito*

En résumé, si selon certains théoriciens, le langage joue une fonction essentiellement communicative (cf. Encadré I.1), d'autres argumentent que sa structure complexe et sa propriété de créativité sont des indicateurs qu'il a en fait évolué pour développer la pensée et que sa fonction est prioritairement cognitive. En d'autres termes, si les tenants de l'hypothèse de la fonction communicative du langage affirment « je parle donc je communique », les défenseurs de la fonction cognitive répondent : « je parle donc je pense ».

I.1.2.4. La pensée sans langage ?

Ainsi, des penseurs égyptiens du III^{ème} millénaire av. JC à certains linguistes contemporains, en passant par les philosophes grecs anciens et certains philosophes modernes, s'est forgée l'hypothèse que le langage aurait avant tout contribué à développer la capacité humaine de penser. Il serait donc façonné par un impératif non pas communicatif mais cognitif et aurait pour fonction primordiale l'élaboration, la construction et l'expression de la pensée.

Le langage a donc une fonction cognitive attestée et son implication dans la pensée semble incontestable. Le rôle du langage et en particulier du langage intérieur, dans les processus cognitifs complexes a été inventorié en détail récemment, notamment dans notre propre revue de la littérature (Perrone-Bertolotti *et al.*, 2014) ou dans le panorama très détaillé de Alderson-Day & Fernyhough (2015). De nombreux travaux récents suggèrent ainsi que le langage joue un rôle dans de nombreuses fonctions cognitives, telles que la catégorisation (Lupyan, 2009 ; Lupyan & Mirman, 2013 ; Langland-Hassan *et al.*, 2017), la mémorisation (Baddeley, 1992 ; Paivio, 1990 ; Pavlenko, 2014), l'orientation et le raisonnement spatial (Loewenstein & Genter, 2005), l'arithmétique mentale (Sokolov, 1972). Il joue également un rôle dans les fonctions exécutives et notamment dans la flexibilité cognitive (cf. le changement de tâches ou *task switching*, Emerson &

Miyake, 2003 ; Laurent *et al.*, 2016), le contrôle cognitif, la résolution de problème complexe (Sokolov, 1972 ; Baldo *et al.*, 2005 ; 2015) et la cognition sociale ou la théorie de l'esprit (Newton & de Villiers, 2007).

La pensée serait donc dépendante du langage. Toutefois, les travaux en aphasiologie de la fin du XIX^{ème} siècle ont questionné la substantialité de ce lien entre pensée et langage. Comme l'explique Bergounioux (2001a), l'observation que dans l'aphasie, la pensée puisse persister, alors que le langage articulé a péri (Broca, 1861, cité par Hécaen & Dubois, 1969) a permis d'entrevoir une dissociation entre pensée verbale et expression orale (ou exophasie). Cette observation a également relancé ce que Bergounioux nomme « l'endophasiologie » — l'étude de la parole du dedans — à la fois par les médecins et neurologues (Baillarger, Ballet, Broca, Charcot), les philosophes (Taine), les linguistes (Emile Egger), les psychologues (Ribot, Binet, Ombredane, Saint-Paul, Victor Egger), qui fait l'objet de la section II.2.1. Selon Bergounioux, il a alors été avancé, notamment par Charcot (1889) ou Ombredane (1951), qu'il existerait des idées sans mots (une pensée sans langage articulé) et, symétriquement, des mots sans idées, ce que Leibniz nommait le « psittacisme », l'emploi de mots vides de sens, à la manière des productions orales des perroquets.

I.1.2.4.1. Conception cognitive forte : le langage comme pilier de la pensée

Ce questionnement sur les liens entre langage et pensée a été repris récemment dans le domaine des sciences cognitives. Carruthers (2002) explique que diverses versions plus ou moins extrêmes de ce qu'il nomme « la conception cognitive » du langage coexistent en sciences cognitives, liées à des perspectives différentes sur les liens entre langage et pensée. Contrairement à la conception purement communicative du langage (cf. I.1.1), la conception cognitive sous-tend que le langage joue un rôle crucial dans la cognition humaine. Mais selon Carruthers, cette conception recouvre en fait des acceptions variées, de fortes à faibles. Selon une des acceptions fortes, le langage est *nécessaire* à la pensée humaine, toute pensée conceptuelle humaine (à la différence des proto-pensées) *requiert* le langage (Davidson, 1975 ; Wittgenstein, 1953).

Une autre acception forte, dans la lignée des travaux sur la relativité linguistique de Whorf (1956), est que le langage est le médium, le support, de toute pensée conceptuelle (Dennett, 1991 ; Bickerton, 1990). L'apparition du langage aurait rendu possible la pensée conceptuelle humaine en démultipliant les capacités computationnelles de l'esprit des premiers hominidés. La pensée conceptuelle dépendrait ainsi du langage, de façon constitutive, voire consubstantielle.

Selon Carruthers, ces acceptions fortes sont peu tenables, car elles sont contredites par les données actuelles en ontogenèse et phylogenèse. Des capacités cognitives et des connaissances complexes du monde ont en effet été observées chez les nourrissons, à des stades pré-linguistiques (*e.g.* Hickmann, 2001 ; Lécuyer *et al.*, 1996). De même, les travaux sur les premiers hominidés (*Homo erectus* et les formes archaïques d'*Homo sapiens*), dont on suppose qu'ils n'avaient pas développé le langage, suggèrent qu'ils étaient capables de planification et d'interactions sociales complexes (Mithen, 1996). Carruthers soutient donc que les versions fortes de la conception cognitive ne sont pas adaptées.

Gauker (2002, 2005, 2011) réfute l'argument qu'il puisse exister une pensée conceptuelle chez des êtres non-linguistiques. Selon lui, les comportements de raisonnements et de

résolution de problèmes observés chez ces êtres ne sont pas une forme de pensée conceptuelle. Pour Gauker (2002) la pensée conceptuelle basique représente un particulier comme appartenant à un type général (ou plusieurs particuliers comme étant liés). Les autres sortes de pensées conceptuelles sont liées par inférence à des pensées conceptuelles basiques. Selon Gauker, les créatures privées de langage ne sont pas capables de penser à un individu particulier et à le représenter comme appartenant à un type général (une catégorie conceptuelle). Les comportements de catégorisation prétendument observés chez les animaux seraient en fait, selon Gauker (2002, 2005), le produit d'une pensée imagée (*imagistic thinking*), permettant aux animaux de percevoir une image x comme plus proche d'une image y que d'une image z sans nécessairement penser que x et y appartiennent à un type commun (un arbre par exemple). De même les comportements prétendument stratégiques observés chez certains animaux, avec des plans se déroulant sur plusieurs jours, ne seraient que la reconnaissance de similarité entre l'événement en train de se dérouler et un événement mémorisé. En résumé, Gauker considère que le type de pensée observée chez certains animaux et chez les nourrissons humains au stage prélinguistique n'est pas conceptuel mais correspond à des jugements de similarité, ne nécessitant pas de posséder des concepts. L'acquisition de concepts est, selon Gauker, liée à l'acquisition de mots. La pensée conceptuelle est donc dépendante du langage car les concepts sont construits par le langage. Gauker oppose ainsi des traitements mentaux de type pensée imagée, reposant sur des jugements de similarité et qui peuvent se faire sans langage à des traitements reposant sur des concepts et donc des mots. En d'autres termes, il n'y aurait pas, selon Gauker, pensée conceptuelle sans mot mais il pourrait exister une forme de pensée imagée, non conceptuelle, sans langage.

Encadré 1.3. Le langage comme nécessaire à la pensée conceptuelle

Ainsi, les tenants de conceptions cognitives fortes du langage (e.g. Bickerton, Davidson, Dennett, Gauker, Wittgenstein) affirment que le langage est le support, de toute pensée conceptuelle, l'acquisition de concepts étant liée à l'acquisition de mots. Les formes de raisonnement et de résolution de problèmes observées chez les animaux ou les nourrissons au stade pré-linguistique ne seraient que des jugements de similarité, non computationnels et non conceptuels.

1.1.2.4.2. Conceptions cognitives faibles : le langage comme étai de la cognition

Parmi les versions faibles de la conception cognitive décrites par Carruthers (2002), la moins controversée est que le langage serait un canal à travers lequel les concepts et les croyances sont acquis (Hampton, 2002). Ainsi, le langage ne serait pas le véhicule représentationnel de toute pensée, mais serait requis pour acquérir certains concepts (comme les concepts scientifiques d'électron, de neutrino ou d'ADN, qui seraient inaccessibles à une personne dépourvue de langage (cf. aussi Hampton, 2002).

Une autre version décrite par Carruthers, plus forte et matière à débats, est que l'acquisition du langage ne servirait pas uniquement à garnir l'esprit de concepts et de croyances, mais sculpterait, façonnerait les processus cognitifs (Bowerman & Levinson, 2001). Cette hypothèse semble étayée par des données comparatives sur l'influence de la langue ambiante sur, par exemple, les processus de catégorisation chez les enfants. Il a été montré que la prépondérance des verbes en coréen vs. des noms en anglais peut

être mise en correspondance avec les performances en résolution de problèmes vs. en catégorisation d'objets par les enfants de ces langues (Gopnik, Choi & Baumberger, 1996). Toutefois, selon Carruthers (2002), le fait que l'acquisition d'une langue plutôt qu'une autre conduise les enfants à porter leur attention sur certains éléments plutôt que d'autres ou même à raisonner différemment, n'indique pas que le langage soit systématiquement impliqué dans la pensée, ni que le langage continue à avoir un rôle dans la cognition adulte.

Selon Carruthers (2002), une troisième version faible de la conception cognitive, inscrite dans la lignée des travaux de Vygotski (1934/1985), considère le langage comme un échafaudage, plus précisément un étayage, pour le développement de la cognition. Un certain nombre de travaux montre en effet que les enfants verbalisent plus lorsqu'ils doivent résoudre des problèmes plus complexes et que les enfants qui verbalisent le plus ont les meilleures performances dans la résolution de problème (Diaz & Berk, 1992).

Dans cette perspective, Clark (1998) propose que le langage, et notamment le langage intérieur, soit un *outil* cognitif, utilisé non pas simplement pour la communication, mais aussi pour augmenter les capacités cognitives humaines. Il défend une conception *supra-communicative* du langage, dans laquelle le langage est considéré comme une ressource complémentaire aux compétences cognitives primitives (ou fondamentales), présentes chez certains animaux non humains. Selon Clark, le langage compléterait les capacités primitives de six façons. La première est *l'accroissement de la mémoire* : l'écriture permet la rétention des informations, des plus simples (liste de courses ou de tâches) aux plus complexes (manuel de statistique...). La seconde est la *simplification de l'environnement*, comme l'utilisation d'étiquettes linguistiques symboliques (« centre ville »), nous permettant de nous repérer dans nos environnements complexes. L'étiquetage linguistique permet aussi d'organiser le monde en classes d'objets (« les chiens ») et même en concepts, d'abord appréhendés par des descriptions linguistiques puis labélisés (comme la charité, l'extorsion, les trous noirs). La troisième est la *coordination et la réduction de la délibération en ligne*. Les échanges linguistiques que nous avons avec autrui ou les formulations que nous nous adressons à nous-mêmes nous permettent de coordonner nos activités, de planifier nos actions, de les enchaîner de façon téléologique, et d'éviter de recourir à des délibérations instantanées, en ligne, pour chaque nouvelle action à produire. La quatrième est le *dépassement de l'apprentissage par parcours individuel (Taming Path-Dependent Learning)*. Tout apprentissage, toute formation, est un parcours qui nécessite des étapes. L'utilisation du langage permet de préserver les idées et de les faire migrer d'un individu à l'autre, permettant de raccourcir certaines étapes. Une idée ayant émergé chez un individu après un parcours intellectuel spécifique peut ensuite, par migration, se développer et s'épanouir plus amplement chez un autre individu diversement expérimenté et équipé. Le langage permet ainsi de transcender l'apprentissage individuel (par étapes) transformant ainsi la cognition humaine en un processus collectif. La cinquième façon est la *guidage de l'attention et l'allocation de ressources*. Les instructions que nous nous formulons linguistiquement nous permettent de nous focaliser, de nous contrôler et de nous corriger. La sixième et dernière façon qu'a le langage de compléter nos ressources cognitives primitives est par *la manipulation et la représentation de données*. L'utilisation du langage et plus particulièrement du langage écrit, permet de prendre des notes, de tracer des schémas, des plans, de manipuler des idées et de laisser l'environnement extérieur interagir avec ces traces, ré-organiser nos idées, les séquencer différemment, de les décomposer ou les regrouper, étendant ainsi

nos arguments ou nos concepts internes initiaux. Pour résumer, le rôle du langage selon Clark (1998) n'est pas simplement de stocker et communiquer des idées, mais plutôt de rendre accessible des concepts, des stratégies et des parcours d'apprentissages qui ne seraient pas accessibles à des cerveaux individuels. Pour Clark, le langage serait un étayage de la cognition humaine, augmentant la portée, la complexité et la variété de nos capacités cognitives. Clark (2002) précise que sa conception du langage n'est pas aussi « faible » que le suppose Carruthers (2002). Selon sa conception, le langage ne sert pas uniquement à fixer des contenus mentaux spécifiques pour les réutiliser ultérieurement et permettre d'étendre la réflexion. Il rend surtout possible la pensée de la pensée, autrement dit la « dynamique cognitive de second ordre » ou la *métacognition*. Le langage re-décrit nos propres pensées dans un format qui les rend disponibles pour toutes sortes d'opérations et manipulations. Ainsi, en véhiculant les pensées sous formes de phrases, résistantes au contexte et trans-modales, le langage permet de les stabiliser, de les rendre scrutables, examinables à loisir et permet ainsi de penser l'acte de penser. C'est, selon Clark, grâce à la parole intérieure, à la production mentale verbale, que des pensées peuvent cibler d'autres pensées. Par conséquent, le fait de pouvoir penser notre propre pensée pourrait nous permettre de nouvelles capacités, telles que l'auto-évaluation, l'auto-critique et l'auto-correction, qui seraient véritablement spécifiques à l'être humain et qui permettraient de repousser les limites de la cognition humaine (je reviendrai sur ce point dans la section I.1.3).

Carruthers (2002) interprète cependant Clark (1998) simplement comme une version faible peu controversée de la conception cognitive du langage. Il affirme que sa propre conception, illustrée sur la Figure I-4, est intermédiaire entre les versions fortes, intenable et les versions faibles, telles que celles de Diaz & Berk (1992) ou Clark (1998). Il considère ainsi le langage comme le medium de la pensée non-spécifique à un domaine, le lieu de l'intégration inter-modulaire.

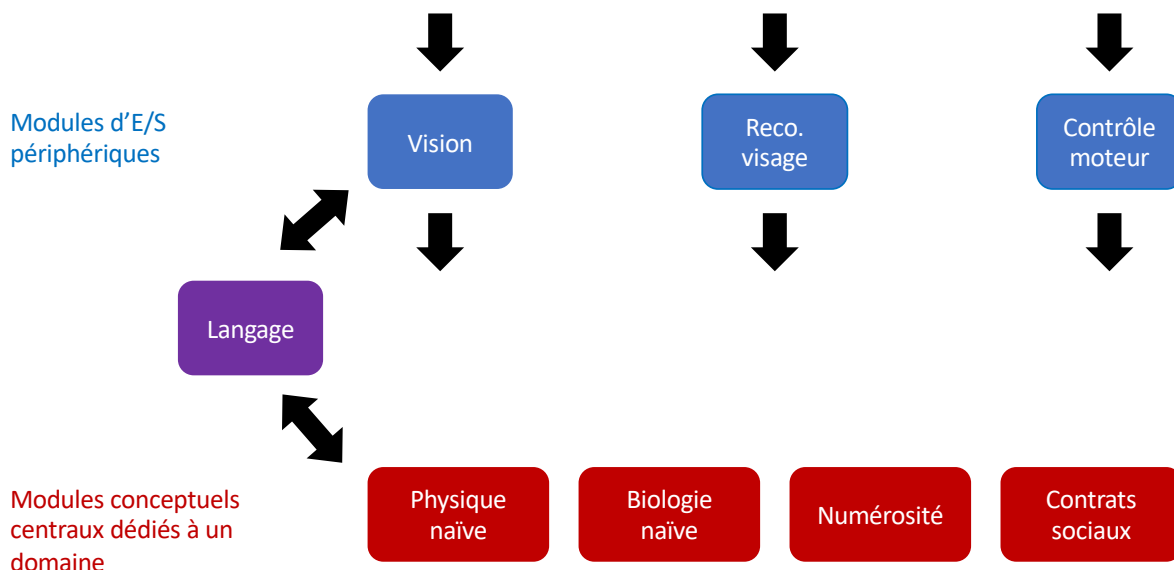


Figure I-4. Illustration de la conception modulaire des liens entre langage et pensée telle que défendue par Carruthers(2002)

Ainsi, Carruthers (2002) révisé la conception modulaire de la cognition développée par Fodor (1983) en étendant la notion de module à tous les processus mentaux, qu'ils soient périphériques ou centraux. Il fait ainsi l'hypothèse que l'esprit humain est constitué de

modules autonomes isolés, qu'il considère comme innés, de deux types : des modules périphériques d'entrée-sortie – tels que la vision, la reconnaissance de visage, le contrôle moteur et le langage – ainsi que des modules centraux de traitement (*central-process modules*), de types conceptuels, qui reçoivent des entrées d'ordre conceptuel et fournissent des sorties elles-mêmes conceptuelles – tels que la psychologie naïve (Baron-Cohen, 1995), la physique naïve (Leslie, 1994) la biologie naïve (Atran, 1990), la numérosité (Dehaene, 1997), la géométrie naïve (Hermer & Spelke, 1994) et les contrats sociaux (Cosmides & Tooby, 1992). Selon Carruthers, les modules conceptuels sont spécifiques, dédiés à un domaine, mais des transferts d'information se font entre modules conceptuels et les entrées-sorties des modules conceptuels sont donc traitées dans un format non-dédié, inter-modulaire. Ce serait, selon Carruthers, le langage qui permettrait d'associer et combiner les informations entre les modules. La pensée propositionnelle inter-modulaire et non-spécifique à un domaine dépendrait du langage. Le langage serait ainsi impliqué de façon constitutive dans la pensée. Il serait le véhicule non-modulaire et non dédié de la pensée conceptuelle, en intégrant les sorties des processus conceptuels centraux modulaires. Car contrairement aux autres modules périphériques (tels que la vision), le langage peut communiquer en entrée-sortie avec les modules conceptuels centraux (selon Carruthers, la vision fournit des informations aux modules conceptuels mais n'en reçoit pas). Les représentations internes du module de langage, les formes logiques structurées (FL, ou *Logical Form*, Chomsky, 1995), constitueraient le format de représentation permettant d'intégrer les informations issues de multiples modules conceptuels. Le langage serait donc, selon Carruthers, le vecteur de la communication inter-modulaire, il unifierait les sorties des différents modules conceptuels, permettant une pensée propositionnelle intégrée, non restreinte à un domaine. La pensée propositionnelle intermodulaire, non-spécifique à un domaine, serait donc exprimée, enchâssée dans le langage.

Carruthers (2002) ne nie donc pas que certaines formes de pensée ou de traitements cognitifs puissent se faire sans le langage, c'est le cas selon lui des traitements conceptuels modulaires dédiés, spécifiques à un domaine. Ainsi une forme de pensée conceptuelle pourrait exister sans le langage et pourrait, selon Carruthers, s'observer chez les mammifères non humains ou chez les nourrissons humains au stade prélinguistique :

« I shall assume that all mammals, at least, are capable of thought – in the sense that they engage in computations which deliver structured (propositional) belief-like states and desire-like states (Dickinson 1994; Dickinson & Balleine 2000). I shall also assume that these computations are largely carried out within modular systems of one sort or another (Gallistel 1990) (...). Furthermore, I shall assume that mammals possess some sort of simple non-domain-specific practical reasoning system, which can take beliefs and desires as input, and figure out what to do. » (Carruthers, 2002, p. 666).

Encadré I.4. Le langage comme nécessaire à la pensée propositionnelle intégrée ou comme étayage de la cognition et de la métacognition

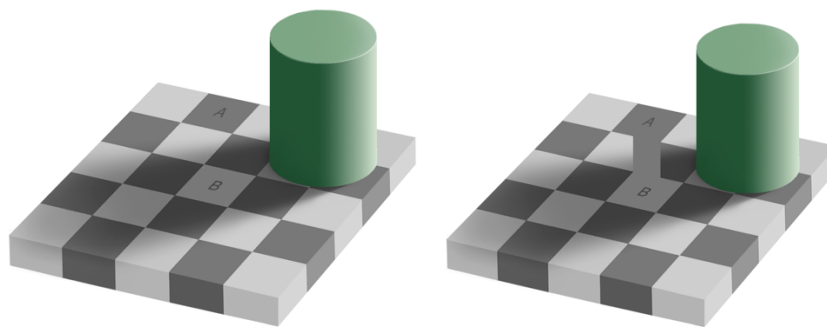
Ainsi, dans la conception cognitive intermédiaire de Carruthers, le langage serait donc un module périphérique d'entrée-sortie parmi d'autres, permettant aux modules conceptuels centraux, dédiés à des domaines spécifiques, d'échanger des informations entre eux. Le langage serait le vecteur de cet échange d'information entre modules conceptuels spécialisés. Il unifierait les sorties des différents modules conceptuels, permettant une pensée propositionnelle intégrée, non restreinte à un domaine. Le langage serait donc nécessaire à la pensée propositionnelle intermodulaire, non-spécifique à un domaine, mais des formes de pensée conceptuelle restreinte à un domaine pourrait exister sans langage, donc chez certains animaux ou chez les nourrissons humains au stade prélinguistique.

Dans la conception cognitive plus faible (moins extrême) de Clark, le langage est considéré comme un étayage de la cognition humaine, augmentant la portée, la complexité et la variété des ressources cognitives humaines et permettant des capacités métacognitives, telles que l'auto-évaluation, l'auto-critique et l'auto-correction, qui seraient véritablement spécifiques à l'être humain.

I.1.2.4.3. Arguments contre la modularité du langage et de la cognition

Postuler l'existence de modules autonomes et innés, dont le langage ferait partie, me paraît être cependant une hypothèse inutilement forte (c'est aussi ce que remarquent Clark ou Bickerton, 2002), sans compter que la notion de module elle-même est sujette à interprétations variées (cf. Coltheart, 1999 ; Fodor, 1983 ; Pinker, 1997). On peut, en fait, reconsidérer les arguments de Carruthers sur les liens entre pensée et langage en émettant l'hypothèse que les différents traitements mentaux (périphériques et conceptuels) soient non autonomes, d'une part, et qu'ils soient acquis ou en tout cas façonnés par l'expérience, d'autre part. La conception modulaire des traitements mentaux est difficilement compatible avec les données actuelles de neuroimagerie qui montrent que chaque traitement mental requiert des réseaux d'aires cérébrales interconnectées et fonctionnant en interaction avec d'autres processus, de façon perméable et absolument non autonome (cf. par exemple sur le traitement des visages, Vuilleumier *et al.* 2001). Les données sur les performances des patients cérébro-lésés ou sur la plasticité cérébrale vont elles aussi à l'encontre des conceptions strictement modulaires de la cognition (e.g. Duffau *et al.*, 2014). De plus la conception modulaire de Carruthers conférant au langage un rôle d'intégrateur inter-modulaire, me paraît se heurter à une difficulté sur les interactions bas-niveau qui semblent bien exister entre les processus mentaux. En effet, si chaque module conceptuel (par exemple le module de la géométrie et le module de physique naïve) fait un traitement sur une scène en cours, ces traitements peuvent se compléter et donc s'influencer mutuellement, ce qui est contradictoire avec leur autonomie. Dans l'exemple de l'échiquier d'Alderson de la Figure I-5, si des modules autonomes sont à l'œuvre, alors le module de géométrie qui permet d'apprécier les relations spatiales entre les objets présents dans la scène, indique (dans son mentalais dédié) qu'*un cylindre est placé sur l'échiquier, avec une source de lumière derrière lui*. Le module de physique naïve, qui permet d'établir les propriétés physiques des objets, fournit l'information (exprimée dans son mentalais propre) que *le cylindre est vert et les cases a et b de l'échiquier sont toutes les deux gris foncé*. La pensée issue des

traitements autonomes de ces deux modules dédiés, et intégrée par le langage serait donc, selon le modèle de Carruthers (2002) : *un cylindre vert est placé sur l'échiquier, avec une source de lumière derrière lui, et les cases a et b de l'échiquier sont toutes les deux gris foncé*. Ce type de modèle n'explique donc pas l'illusion visuelle qui devrait résulter en la pensée intégrée : *un cylindre vert est placé sur l'échiquier, avec une source de lumière derrière lui, et la case a n'étant pas dans l'ombre du cylindre est intrinsèquement plus foncée que la case b*. Pour expliquer cette illusion, il faut que les deux traitements (géométrique et physique) soient non autonomes et que l'information de localisation de la source lumineuse (issue du traitement géométrique spatial) soit intégrée lors du traitement des propriétés des objets de la scène. Ou alors il faut une organisation ultra-complexe entre le langage et les différents modules – le langage traduisant au fur et à mesure les résultats des sous-étapes de traitements de chaque module et les fournissant aux autres modules, créant ainsi de nouvelles informations à traiter – ce qui ralentirait les traitements de façon non plausible. Il n'est pas exclu que ce soit le langage qui permette la communication (en temps réel) entre les deux types de traitement, mais il me semble exclu qu'une organisation modulaire, dédiée, spécifique, soit à l'œuvre.



*Figure I-5. Illusion de l'échiquier d'Adelson
Les cases A et B ont la même teinte de gris mais A est perçue comme plus foncée que B.
Illustrations tirées de https://en.wikipedia.org/wiki/Checker_shadow_illusion*

Il existe de nombreux autres exemples d'influences et d'interactions entre les différents systèmes de traitements cognitifs, tels que la vision, l'audition, la motricité, le langage (cf. Dale & Spivey, 2002), ce qui suggère que les processus mentaux ne fonctionnent pas de façon autonome. De plus, l'innéité de ces traitements mentaux reste à démontrer, de nombreux travaux montrant que l'acquisition de capacités cognitives, y compris le langage, sont façonnés voire intégralement construits par l'usage et l'interaction avec l'environnement (e.g. Tomasello, 2003).

Un autre argument contre la modularité des traitements cognitifs, et en particulier l'indépendance des processus mnésiques et langagiers, est fourni par les études sur le bilinguisme ou le multilinguisme.

Une vision modulaire simpliste de la mémoire suggère que les souvenirs sont stockés de façon non linguistique. Lors de la récupération d'un souvenir, celui-ci est alors encodé par le module du langage et prend une forme linguistique, fonction de la situation de récupération. Dans le cas du bilinguisme, le modèle hiérarchique « Revised Hierarchy Model » fait l'hypothèse que le système de stockage (un réservoir, « store » en anglais) conceptuel non-linguistique est relié à un lexique A pour la première langue (L1) et un lexique B pour la deuxième langue (L2), ces deux lexiques étant eux-mêmes reliés.

Comme le notent Larsen *et al.* (2002), dans ce type de modèle modulaire, le langage applique des étiquettes linguistiques, des noms, aux images mentales stockées de façon purement conceptuelle. Si un tel modèle peut fonctionner pour évoquer des souvenirs d'objets isolés, il est peu adapté aux souvenirs autobiographiques complexes, contenant des enchaînements d'actions, des sensations variables, des obstacles extérieurs, des contextes multiples, des buts et des moyens variés. Il semble plus approprié de concevoir que le langage joue un rôle dans la récupération de souvenirs autobiographiques complexes et une conception modulaire selon laquelle le stockage d'informations serait indépendant du langage semble inadaptée.

Les travaux de Pavlenko (1999) sur les représentations conceptuelles chez les bilingues apportent un éclairage intéressant sur cette question du lien entre mémoire épisodique et langage. S'inscrivant dans le modèle de la représentation mentale du lexique de Paradis (1997), Pavlenko (1999) distingue explicitement les niveaux lexical, sémantique et conceptuel. Le niveau lexical correspond à la forme du mot, incluant ses propriétés phonologiques et morphosyntaxiques. Le niveau sémantique correspond aux informations sur le mot, disponibles de façon explicite et reliant le mot à d'autres mots, aux idiomes ou expressions conventionnelles. Le niveau conceptuel contient des informations multimodales, non linguistiques, incluant de l'imagerie visuelle, des programmes moteurs, des représentations auditives, tactiles et somatosensorielles. Selon Pavlenko (et Paradis), le niveau de la signification des mots (la sémantique) est distinct de celui des concepts. Elle cite les travaux de Lecours & Joanette (1980) sur l'aphasie globale qui montrent qu'on peut en effet accéder à un concept (un récipient que l'on peut utiliser pour boire du thé) en ayant perdu la capacité de produire ou comprendre le mot associé (« la tasse »). Selon Pavlenko, il faut donc distinguer niveaux sémantique et conceptuel, mais les représentations conceptuelles sont façonnées par les mots et sont donc culturellement et contextuellement variables. Ainsi chez les bilingues qui acquièrent une deuxième langue tardivement, les représentations conceptuelles initiales, formées dans le contexte linguistique et culturel de la L1, sont confrontées à l'influence de la L2 et de la nouvelle culture ambiante. Il peut en résulter différentes interactions. Les conceptions initiales peuvent co-exister avec les nouvelles représentations conceptuelles formées dans le nouvel environnement linguistique et culturel ; elles peuvent aussi être modifiées, en amplitude et en portée, par l'environnement ; elles peuvent enfin converger vers des représentations entièrement nouvelles, différant à la fois de la culture de la L1 et de la L2. Dans le cas des souvenirs autobiographiques, Pavlenko fait l'hypothèse que les représentations conceptuelles de la L1 et de la L2 co-existent. Cette hypothèse est étayée par les travaux de Larsen *et al.* (2002) chez des polonais ayant émigré au Danemark à l'âge adulte et y vivant depuis plus de 30 ans. Ces personnes rapportent que la récupération mentale de souvenirs datant de leur passé en Pologne se fait en utilisant le polonais, alors que pour les événements ayant eu lieu dans leur nouveau pays, la récupération mentale se fait en danois. Larsen *et al.* en concluent que la récupération de souvenirs autobiographiques implique des représentations conceptuelles, qui ont été formées dans une culture et un bain linguistique donné. Elles sont associées à des représentations sémantiques propres à la langue ambiante lors de l'événement et à des formes lexicales dans cette même langue. C'est ainsi que les personnes bilingues polonais-danois peuvent avoir la sensation de se souvenir d'événements en polonais ou en danois, en fonction de la situation. Ainsi, chez les individus qui émigrent à l'âge adulte, il semble co-exister deux ensembles de représentations conceptuelles en mémoire : un pour la période de leur vie dans le pays

d'origine, associé aux systèmes sémantique et lexical de la L1, et un pour la nouvelle période dans leur pays d'adoption, associé aux systèmes sémantique et lexical de la L2 (voir aussi Marian & Neisser, 2000).

Encadré I.5. Contre la conception du langage comme un module autonome

En résumé ce type d'études suggère que le langage interagit avec les différents traitements cognitifs, comme il a été expliqué plus haut, et ceci dès la formation de représentation conceptuelle. Le langage est donc bien plus qu'un module de communication, il est engagé aussi dans la conceptualisation.

Par conséquent, je ne suis pas convaincue par la conception modulaire de l'esprit telle que l'entend Carruthers, mais je pense qu'une conception cognitive supra-communicative (à la Clark, 1998) ou extra-communicative (à la Lupyan, 2009) du langage est juste : le langage sert bien sûr à communiquer, mais il peut également entrer en jeu lors de la conceptualisation et du raisonnement.

I.1.2.4.4. Arguments contre le rôle nécessaire du langage dans la pensée : pensée sans langage et modes de pensée visuel, kinesthésique, abstrait

Ceci étant posé, et que le langage ait donc des fonctions communicative ET cognitive étant admis, il reste encore débattu qu'il soit le seul vecteur de la pensée propositionnelle non-spécifique à un domaine, comme le suggère Carruthers (2002), ou qu'il soit nécessairement impliqué dans la pensée conceptuelle, comme l'affirme Gauker (2002, 2011). L'introspection, le sondage, la réflexion philosophique conduisent à l'idée qu'il puisse exister une forme de pensée conceptuelle (et non restreinte à un domaine) non verbale.

Les études de sondage descriptif de l'expérience subjective menées pendant plusieurs années par Hurlburt (Hurlburt, 1990 ; Hurlburt, 2011), à partir d'entretiens itératifs guidés et de sondage immédiat, ont conduit celui-ci à nuancer l'hypothèse de la prépondérance du langage dans la pensée et à introduire la notion de « *unsymbolized thinking* » (pensée non-symbolisée).

En particulier, Hurlburt a remis en cause cette affirmation que le langage intérieur est ubiquitaire et que nous passons la majeure partie de notre temps d'éveil à penser verbalement. Comme nous l'avons rappelé dans Perrone-Bertolotti *et al.* (2016), cette affirmation de l'omniprésence du langage intérieur, défendue encore récemment par Chomsky (cf. I.1.2.3), date du XIX^{ème} siècle. Après les premières descriptions qualitatives du langage intérieur par Bain (1855), Charcot (1880), Binet (1886), et Stricker (1885), le médecin-psychologue Georges Saint-Paul fut, dès 1892, le premier à mener un large questionnaire d'introspection, auprès de 200 personnes (« Des professeurs éminents, des romanciers illustres, des littérateurs distingués », p.5). Bien qu'il en ait tiré des descriptions minutieuses de différentes formes de langage intérieur⁵, celles-ci restent

⁵ « Presque tous les individus emploient en certaines circonstances, spontanément, sans chercher à le faire, le procédé d'articulation mentale. Chez les uns il est un symptôme de préoccupation ou d'attention, chez d'autres il accompagne une joie vive, une grande douleur, certains états émotifs causés par la peur, l'admiration, la prière..., chez beaucoup il prend naissance sous l'influence de l'introspection ; quand leur « moi » cherche à concevoir comment il pense, il parle sa pensée. Il exprime en tout cas la tendance à enfermer l'idée en un mot, partant à l'isoler, à la rendre plus nette. ».

toutefois critiquables, car fondées sur l'introspection *a posteriori*, et non pas sur l'expérience immédiate du langage intérieur. De plus, les personnes questionnées étaient invitées à se reconnaître parmi différents profils proposés, plutôt que de décrire spontanément leur expérience. Plus tard, la méthode dite « *Experience Sampling* » (i.e., échantillonnage ou sondage de l'expérience subjective) a permis des descriptions plus précises (Larson & Csikszentmihalyi, 1983). Elle consiste à faire décrire l'expérience interne vécue dans les microsecondes qui ont précédé un signal sonore et les observations rapportées sont ainsi plus proches de l'expérience vécue. Cette méthode de rapport immédiat a conduit certains auteurs à considérer le langage intérieur comme quasi-continu, omniprésent, et même difficile à empêcher (Klinger & Cox, 1987). Toutefois, ce type de sondage en direct présente encore un biais expérimental, puisque les personnes interrogées répondent en utilisant le langage (oral ou écrit) et ont ainsi tendance à surévaluer la présence de langage dans leur expérience intérieure. Hurlburt a donc développé une méthode rigoureuse, nommée « *Descriptive Experience Sampling* » (DES, sondage descriptif de l'expérience subjective). Cette méthode réduit le biais de langage en évitant d'utiliser un questionnaire avec des questions prédéfinies et en menant des entretiens guidés itératifs à la suite des périodes de sondage. Dans la méthode DES, rien n'est spécifié à l'avance. Les participants portent sur eux un dispositif qui émet un signal sonore de façon irrégulière, plusieurs fois par jour. À chaque retentissement de la sonnerie, ils prennent note de leur expérience intérieure juste avant la sonnerie. Après avoir porté ce dispositif pendant un jour entier, ils rencontrent l'expérimentateur et passent en revue leurs notes et sont guidés pour décrire avec le plus de précision et de justesse possible leurs expériences intérieures. Cette séquence de sondages par sonnerie aléatoire suivis d'un entretien guidé est répétée plusieurs fois, ce qui conduit en général à des comptes-rendus de plus en plus experts. En utilisant cette méthode de façon rigoureuse, avec des centaines de participants, sur plus d'une trentaine d'années, Hurlburt et ses collègues ont fréquemment rencontré des personnes qui ne notaient aucun épisode de langage intérieur. Ils concluent de leurs travaux que la fréquence d'occurrence du langage intérieur est très différente d'un individu à l'autre et varie de 0% à 100% des instants échantillonnés (Hurlburt, 2011 ; Hurlburt *et al.*, 2013 ; Hurlburt & Heavey, 2015). La moyenne d'occurrence se situe à 26 % (Heavey & Hurlburt, 2008). Ainsi, nous pouvons considérer qu'en moyenne un quart de nos périodes d'éveil est constitué de langage intérieur (cf. aussi Morin *et al.*, 2011).

Hurlburt montre que le reste de notre expérience intérieure est constitué de quatre autres éléments : des images visuelles, des sensations ou sentiments (*feelings*), des expériences sensorielles et ce qu'il nomme « *unsymbolized thinking* », et qui fait référence à une pensée abstraite, explicite, qui n'inclut pas l'expérience de mot, ni d'image ni de tout autre symbole (Hurlburt, 2011). Selon Hurlburt, ce type de pensée serait un phénomène distinct de la pensée verbale, mais n'en serait pas un précurseur, et serait spécifique, clair et précis. Cette pensée non-symbolisée ne serait donc pas spécifique à un domaine, pourtant, elle se ferait sans la médiation du langage. Cette notion de pensée non verbale est à rapprocher de la théorie du double codage (*Dual coding theory*) de Paivio (1990) selon laquelle des stimuli peuvent être codés et manipulés mentalement de façon visuelle et/ou verbale. Elle rappelle également les intuitions d'Albert Einstein sur son propre mode de pensée, telles qu'il les a confiées au mathématicien français Jacques Hadamard :

« *Words or language, as they are written or spoken, do not seem to play any role in my mechanism of thought. The psychical entities which seem to serve as elements in thought are certain signs and more or less clear images which can be « voluntarily » reproduced and combined (...) this combinatory play seems to be the essential feature in productive thought – before there is any connection with logical construction in words or other kinds of signs which can be communicated to others. The above-mentioned elements are, in my case, of visual and some muscular type. Conventional words or other signs have to be sought for laboriously only in a second stage, when the mentioned associative play is sufficiently established and can be reproduced at will.* » (Hadamard, 1945, Appendix II).

John-Steiner (1985/1997) a rassemblé les témoignages d'une cinquantaine de personnalités qu'elle considère comme des « penseurs expérimentés » et a montré que la pensée peut prendre différents modes : elle peut être verbale bien sûr, mais aussi visuelle ou kinesthésique (sensorielle et motrice). Pinker (1994) cite également de nombreux témoignages d'artistes ou de scientifiques (Faraday, Maxwell, Tesla, Kekulé, etc.) qui affirment créer ou penser sans utiliser le langage, mais à partir d'images mentales, de représentations visuelles. Pour Pinker (1994) ces témoignages sont un argument pour la notion de « mentalais » et l'hypothèse selon laquelle la pensée se fait dans un langage universel, proche des langues naturelles, composé de symboles représentant des concepts, et incluant des arrangements de symboles véhiculant des relations comme « qui fait quoi à qui ». L'hypothèse de l'existence d'un mentalais universel est difficile à concilier avec certains travaux des courants de la cognition incarnée et située ou de la relativité linguistique (e.g. Barsalou *et al.*, 2003 ; Boroditsky, 2011 ; Wilson, 2002) qui suggèrent que la pensée humaine s'inscrit dans un cadre culturel, linguistique, environnemental, social, et est loin d'être universelle. Mais sans adhérer à la notion de mentalais universel, il semble bien que les témoignages mentionnés ci-dessus indiquent qu'une forme de pensée conceptuelle peut se dérouler sans langage, et même sans mot.

D'ailleurs la conception que la pensée serait essentiellement verbale est probablement typiquement occidentale. Dans d'autres cultures, le lien entre pensée et langage est plus détendu, voire évité. Kim (2002) explique par exemple que si, dans la tradition philosophique et religieuse (judéo-chrétienne et musulmane) occidentale, le langage est sacré et intimement lié à la pensée, ce lien n'existe pas dans la culture asiatique orientale. Selon Armstrong (1993), les trois principales religions monothéistes placent le langage au cœur du divin. Dans l'Ancien Testament, on lit que la parole de Dieu est créatrice, c'est par sa parole que Dieu a créé le monde : *“Dieu dit : Que la lumière soit ! Et la lumière fut.”* (Gen 1,3). Dans le Nouveau Testament, dans le prologue de l'évangile selon Saint Jean, on lit : *« Au commencement était le Verbe, et le Verbe était en Dieu, et le Verbe était Dieu »*. Le Coran est considéré par les musulmans comme la parole sacrée de Dieu, révélée au prophète Mahomet. Il est considéré comme incréé, c'est-à-dire existant sans avoir été créé, divin, parfait. Il est écrit en arabe dit « coranique », proche de l'arabe classique ancien et même s'il a été traduit dans plus de 40 langues, les pratiquants l'apprennent et le récitent en arabe coranique. Selon Kim (2002), dans les traditions philosophiques et religieuses asiatiques orientales, le langage n'a pas cette place centrale et ce lien présumé avec la pensée. La vanité des paroles est souvent rappelée et les états de silence et d'introspection sont considérés comme bénéfiques pour atteindre des niveaux plus

élevés de pensée⁶. Dans les enseignements bouddhistes et taoïstes, un état de silence, de calme pur, est recherché, par la méditation, en pratiquant la visualisation interne, la concentration, et la régulation de la respiration, pour atteindre la « Vérité Suprême » (Rinpoche, 1987 ; Robinet, 1995). Kim (2002) montre que ces conceptions culturelles occidentales et orientales différentes du lien entre langage et pensée correspondent à un rôle différent du langage dans la performance cognitive. Elle a comparé la capacité de résolution de problèmes (à partir des matrices progressives de Raven avancées), accompagnée de verbalisation à voix haute ou non, chez des participants américains d'origine européenne et asiatique. Ses résultats indiquent que la performance des participants d'origine asiatique baisse lorsqu'on leur demande de verbaliser à voix haute les étapes de leur raisonnement alors que la performance des participants d'origine européenne n'est pas affectée. Elle montre également que la suppression articulatoire (le fait de demander aux participants de réciter l'alphabet pendant la tâche cognitive) diminue le score des participants d'origine européenne, mais pas celui des participants d'origine asiatique. Elle conclut que les participants d'origine asiatique utilisent moins le langage intérieur, et utilisent un mode de pensée moins verbal.

Par conséquent, contrairement à l'hypothèse de Gauker (2011), pour qui toute pensée conceptuelle est verbale – le concept ne pouvant se former sans le mot –, il se pourrait bien que chez certains individus, des concepts émergent et soient manipulés mentalement sans la médiation par le langage.

Le cas de Chelsea, rapporté par Glusker (1987), Dronkers (1987) et Dronkers *et al.* (1998), et de nouveau discuté par Baldo *et al.* (2015), est éclairant. Chelsea est une femme issue d'une famille rurale en Californie. Elle est née avec une surdité profonde et a été élevée par ses parents, à la maison, avec ses frères et sœurs. Sa surdité n'a été prise en charge qu'à l'âge de 32 ans, âge auquel elle a été appareillée auditivement et a pu commencer à parler. Selon ses parents, la famille n'utilisait pas la langue des signes avec Chelsea et elle s'exprimait en pointant, ou avec d'autres gestes ou des mimes. Les chercheurs considèrent donc que Chelsea n'a pas été exposée au langage pendant 32 ans. Pourtant, selon sa famille, pendant cette période, elle a pu aider sa famille dans les tâches ménagères, cuisiner, prendre soin de ses plus jeunes frères et sœurs, etc. D'autres cas semblables ont été rapportés, dont Ildefonso, né sourd dans une famille rurale du Mexique et qui n'a découvert la langue des signes qu'à l'âge de 27 ans, à son arrivée aux Etats-Unis, grâce à Susan Schaller, enseignante dans une école pour sourds adultes. Selon Schaller (1991/2012), Ildefonso a commencé par interpréter les signes comme des instructions (« ouvre le livre » pour le signe « livre ») et ce n'est qu'après plusieurs jours qu'il a eu, comme Helen Keller, une révélation en découvrant la fonction symbolique des signes, en comprenant que chaque objet avait un nom. Pourtant, avant de découvrir la langue des signes, il était capable de s'occuper du bétail, d'aider à semer, récolter.

Il semble difficile de faire l'hypothèse que le type de tâches effectuées par Chelsea et Ildefonso puisse s'exécuter sans aucun concept mental ni catégorie et ne reposer que sur la seule capacité de jugement de similarité, définie par Gauker (2005). Il apparaît que des raisonnements et des activités élaborées peuvent s'effectuer sans le langage. Il se peut donc qu'une forme de pensée abstraite remplace (et précède parfois) la pensée verbale (typiquement chez les individus avec un déficit ou un trouble de la modalité verbale, ce

⁶ Notons que certains courants philosophiques ou religieux occidentaux encouragent également la pratique du silence intérieur (cf. la prière apophasique, le silence cartusien).

qui était probablement le cas d'Albert Einstein et ce qui était le cas de Chelsea). Cette forme de pensée abstraite, non exprimée par le langage est à rapprocher de la notion de *verbum interius* (le verbe intérieur) ou de *verbum in corde* (le verbe engendré dans le cœur) déjà introduite par Saint Augustin au début du Vème siècle. On peut ainsi lire :

« vous comprenez bien que la parole ou l'idée était en mon esprit avant de choisir un terme, une voix pour arriver jusqu'à vous. Tous comprennent aussi, je pense, que ce qui se fait en moi se produit également dans tous ceux qui parlent. Je sais donc ce que je veux dire, je le possède dans mon esprit, je cherche des termes pour l'exprimer ; avant que ces termes soient prononcés par ma voix, je possède assurément la parole, la pensée en moi-même. Ainsi la parole est en moi antérieure à la voix ; elle existe d'abord, la voix ne vient qu'ensuite. En toi au contraire, c'est l'oreille qui est frappée d'abord du son de ma voix pour porter ma pensée, ma parole à ton esprit. » (Augustin, Sermon 288, 4).

Augustin distinguait le *verbum in corde*, le verbe à l'intérieur de celui qui parle (en son cœur) de la voix exprimée, ou *locutio*, qui survient ensuite. Le *verbe intérieur*, défini par Augustin dans *De Trinitate*, n'appartient à aucune langue, il est universel, il est antérieur à toute forme linguistique (Cary, 2011 ; Panaccio, 2014). Augustin l'explique lui-même :

« Voyez une parole conçue dans l'intelligence, elle cherche à en sortir, elle veut qu'on la profère; on examine à qui on va la porter. Rencontre-t-on un Grec ? On cherche une expression grecque pour la lui faire comprendre. Un Latin ? C'est un terme latin. Un Carthaginois ? C'est une expression punique. Supprime ces différents interlocuteurs, et la parole intérieure n'est ni grecque, ni latine, ni (431) punique, ni d'aucune autre langue. » (Augustin, Sermon 288, 3).

Ces lignes de l'évêque d'Hippone résonnent, une quinzaine de siècles plus tard, avec les conceptions contemporaines des liens non exclusifs entre pensée et langage.

Encadré I.6. Modes de pensée non verbaux

En résumé, il se pourrait bien que dans certains cas et chez certains individus, des concepts émergent et puissent être manipulés mentalement sans la médiation par le langage. Il semble qu'il existe des modes de pensée visuel, kinesthésique et peut-être abstrait qui ne requièrent pas le langage. En somme, si le langage peut parfois être impliqué dans la pensée et peut la faciliter, l'étayer, il n'en est pas le seul vecteur, il n'est pas systématiquement constitutif.

I.1.2.4.5. L'indicible et l'innommable

Un dernier argument sur la non contingence entre langage et pensée réside dans l'indicible, magnifiquement décrit dans Manon Lescaut de l'Abbé Prévost :

« Je demeurai, après cette lecture, dans un état qui me serait difficile à décrire car j'ignore encore aujourd'hui par quelle espèce de sentiments je fus alors agité. Ce fut une de ces situations uniques auxquelles on n'a rien éprouvé qui soit semblable. On ne saurait les expliquer aux autres, parce qu'ils n'en ont pas l'idée ; et l'on a peine à se les bien démêler à soi-même, parce qu'étant seules de leur espèce, cela ne se lie à rien dans la mémoire, et ne peut même être rapproché d'aucun sentiment connu. » Prévost (1731/1839, p. 100).

Il existe des sentiments et des émotions que l'on éprouve mais que l'on peine à exprimer avec les mots de notre langue ou que l'on a du mal à traduire d'une langue à l'autre. Dans son ouvrage encyclopédique des émotions humaines, Watt Smith (2015) recense 156 états d'âmes qui sont associés à un vocable dans une des langues du monde. On y découvre que les Baining de Nouvelle-Guinée ont le mot « awumbuk » pour ce sentiment de nostalgie et d'inertie qui nous prend lorsqu'un invité bien-aimé s'en va, ou que les Inuit nomment « iktsuarpok » le sentiment d'attente insupportable qui nous submerge lorsque nous attendons une réponse à un message téléphonique, un bus qui n'arrive pas, des invités en retard, ou encore qu'en polonais le sentiment de chagrin multi-forme, qui va de la mélancolie à la rage en passant par la plainte ou le regret intense, se nomme « žal ». Ces émotions peuvent sembler familières, pourtant comment les désigner en français ?

L'indicible peut naître aussi de situations qu'on préfère taire. Benjamin Orenstein, rescapé de la Shoah, a gardé le silence pendant 48 ans sur l'horreur des camps de concentration. Devant la montée du négationnisme, dans les années 1990, il a résolu de s'exprimer et donne chaque année de nombreuses conférences, pour faire de ses auditeurs des « témoins de témoin ». Mais il se tait sur certains de ses souvenirs, de ses souffrances trop extrêmes : « certaines choses, je ne pourrai jamais les raconter » (Orenstein, 2018). Les mots peuvent faire peur, être impossibles à énoncer. Mais, plus que l'indicible, il y a ce qui ne peut s'exprimer par le langage. Il y a l'innommable – évoqué dans « les carnets de Marc Masson », dans « J'irai tuer pour vous » le dernier roman de mon frère Henri, inspiré de l'histoire poignante de son ami, agent clandestin français :

« Mais jamais je ne trouverai les mots pour décrire ce que ces animaux – non, ces humains – me font endurer pendant les plus horribles journées de mon passage sur la Terre des hommes. C'est que les mots, sans doute, n'existent pas pour dire l'inhumanité de mes bourreaux, ces fils de rien, ces âmes vides, et le seul fait d'y penser me déchire encore les entrailles. On ne décrit pas la barbarie. Il faudrait bien du cœur pour y parvenir. Aucune langue ne possède le vocable à même de traduire cela, aucun idiome ne porte en lui assez de puissance évocatrice pour faire toucher du doigt à celui qui écoute la laideur réelle des ténèbres. » (Henri Lævenbruck, 2018, p. 538).

La littérature, l'ethnologie, l'histoire nous rappellent qu'il est des sentiments et des pensées que le langage ne parvient pas à nommer : un argument de plus pour la non-unité entre pensée et langage.

I.1.2.4.6. Synthèse : fonctions communicative et cognitive

Je préfère donc m'en tenir à une position selon laquelle le langage facilite la pensée, et particulièrement la métacognition (cf. Clark, 1998 ; Alderson-Day & Fernyhough, 2015), mais n'en est pas le seul vecteur, même dans le cas de la pensée propositionnelle non-spécifique à un domaine (contra Carruthers, 2002) ou dans le cas de la pensée conceptuelle (contra Gauker, 2011). Autrement dit, le langage peut faciliter et soutenir la pensée, mais il n'en est probablement pas systématiquement constitutif. L'argument suggéré par Reboul (2017) selon lequel le langage n'a pas simplement évolué pour les besoins de la communication, puisqu'il existe des systèmes de communication qui ne présentent pas la complexité du langage humain (voir les cas de communication animale) peut se retourner contre la position selon laquelle il aurait évolué pour les besoins de la cognition : car il existe des systèmes de pensée qui se passent des outils linguistiques.

Tout comme il est possible de communiquer sans langage, il est possible de penser sans langage. Compte tenu de ces éléments, je reste agnostique sur la primauté de la fonction cognitive du langage sur sa fonction communicative. Contrairement à Chomsky (2002, cf. I.1.2.2 ci-dessus), je ne pense pas que l'utilisation la plus importante que nous fassions du langage soit pour nous-mêmes. Que ce soit une pression sociale et communicative qui ait en premier lieu façonné le langage ou que ce soit le besoin de développer une pensée élaborée et complexe, ou que ce soit les deux à la fois est difficile à trancher. Car il n'existe pas de langage sans groupe social et si le langage, dans sa version d'évolution actuelle, permet en effet de structurer les concepts et développer une pensée sophistiquée, pourquoi aurait-on eu besoin d'une pensée complexe, si ce n'est pour construire et entretenir des relations sociales elles-mêmes de plus en plus élaborées, nécessitant des formes de communication elles-mêmes plus évoluées.

C'est aussi la position de Hagège (1985), qui propose une théorie de l'homme (l'humain) « dialogal », dans laquelle la cognition joue un rôle majeur :

« Chacun peut voir le langage servir à la communication. [...] Mais il convient de se garder des vues réductrices. Interaction dialogale ne signifie pas simple transfert d'information. D'abord, le discours, en quoi les langues prennent corps, institue un échange qui commande une hiérarchisation de l'information selon l'importance, et va bien au-delà d'une pure transmission de messages. Ensuite, si ces derniers communiquent, c'est qu'ils ont quelque chose à communiquer, qui n'est pas le produit d'un simple prélèvement sur le monde et l'événement. Les langues sont des modèles, façonnés par la vie sociale, d'articulation du pensable, grâce auxquelles se déploie une réflexion capable d'ordonner le monde. L'expérience est perçue en bloc, mais elle est hiérarchisée linéairement dans l'étendue du discours. Cette opération est à la fois dialectiquement, la trace de la pensée et cela même qui la nourrit. Méthodes d'analyse, les langues sont du même coup des facteurs essentiels de construction de la personnalité, aussi bien pour l'individu depuis sa naissance que pour l'espèce au cours de son histoire. [...] le langage est à la fois, en contexte de groupe, méthode de pensée et produit de la pensée au sens le plus général. Né, peut-être pour des fins pratiques et des significations communes, il a perfectionné l'espèce en même temps qu'elle le perfectionnait. » (Hagège, 1985, p.260-261, c'est moi qui souligne).

Encadré I.7. Fonction communico-cognitive : *Loquor, ergo communico et cogito*

Ainsi, la pression communicative exercée sur le langage est issue de pressions sociales et culturelles. Et en réponse aux besoins d'une communication de plus en plus efficace et socialement adaptée, la pensée doit se structurer. La pression cognitive exercée sur le langage serait donc en fin de compte elle-même issue de pressions sociales, culturelles et communicatives, et réciproquement, puisqu'une pensée élaborée entraîne des besoins d'échanges complexifiés. Par conséquent, le langage dérive, selon moi, de contraintes à la fois communicatives (sociales) et cognitives, les deux étant intimement liées. Les besoins d'échanges, de communication ont créé des besoins de langage et de pensée complexes. La pensée complexe a elle-même créé des besoins de langage évolué et a contribué à façonner une communication plus élaborée.

Ce que l'on peut résumer en : « je parle, donc je communique et je pense ».

I.1.3. La fonction d'autoconscience : *loquor, ergo sum*

Le langage humain a donc une fonction communicative visant les échanges avec autrui, ainsi qu'une fonction cognitive contribuant à construire et élaborer la pensée propre du locuteur. Il est ainsi capital dans le développement des interactions avec autrui autant qu'avec soi-même. Mais plus que cela, le langage contribue également à la fonction métacognitive d'autoconscience, en précisant, voire établissant, la frontière entre soi⁷ et autrui, participant à la conscience de soi au cours du temps ou conscience autoconsciente, que l'on peut définir comme une illusion nécessaire à la cohésion et la coopération sociale, à la communication humaine (Hume, 1739 ; Frith, 2010).

I.1.3.1. La métacognition : prise de conscience de sa propre pensée par le sujet parlant

Tout d'abord, avant de nous intéresser à la conscience de soi, notons que le langage est considéré par de nombreux philosophes et chercheurs comme permettant au sujet de prendre conscience de son activité de pensée. Hegel (1817) écrit ainsi que pour que nous soyons conscients de nos pensées réelles et déterminées, il faut que nous leur donnions une forme objective, à travers l'utilisation du langage. Il reconnaît l'existence d'une pensée non verbale « *ineffable* » mais celle-ci serait une « *pensée obscure, la pensée à l'état de fermentation, et qui ne deviendrait claire que lorsqu'elle trouverait le mot. Ainsi le mot donne à la pensée son existence la plus haute et la plus vraie.* ». C'est alors que la pensée, selon Hegel, devient claire et que l'esprit peut la saisir comme une pensée. Autrement dit, c'est par l'expression verbale que le sujet prend conscience de sa pensée.

Plus récemment Frankish (1998, 2010) a proposé que le langage (la parole intérieure) joue un rôle causal dans nos décisions d'accepter, de rejeter ou d'agir sur les propositions exprimées mentalement. C'est l'instanciation, l'activation de phrases en mode intérieur qui nous permet d'émettre des décisions sur notre pensée propositionnelle. Frankish se place dans le cadre des théories du double traitement ou du double système de pensée, développées en sciences cognitives par Evans & Over (1996), Evans (2003), Evans & Frankish (2009) ainsi que Stanovich & West (2000), ainsi que, de façon parallèle, dans le domaine des recherches sur la prise de décision par Kahneman, Slovic & Tversky (1982) ou Kahneman (2011). Selon ces théories, deux systèmes de pensée (ou de processus cognitifs) sont mis en œuvre dans la pensée et le raisonnement. Le système 1, ou système implicite, est celui de l'intuition, constitué de réactions instinctives, automatiques et rapides sur lesquelles reposerait la plus grande partie de nos décisions. Il est décrit comme une forme de cognition universelle, innée et commune aux animaux non-humains et humains. Il serait non délibéré, non conscient et correspondrait à une pensée non propositionnelle incluant des traitements en parallèle, dédiés chacun à un domaine. Le système 2, ou explicite, est celui de la pensée analytique, lente, avec des traitements en série. Il serait délibéré et dirigé consciemment. Il aurait évolué tardivement et serait propre aux humains. Il correspondrait à la pensée propositionnelle. Selon Frankish, la parole intérieure interviendrait dans le système 2. Le raisonnement conscient serait donc dépendant du langage et serait une internalisation des échanges verbaux qui interviennent en situation sociale réelle de résolution de problème, par exemple.

⁷ En français, dans la tradition de Pascal (« Qu'est-ce que le moi ? », Pensées, 1670), on utilise souvent le terme de « moi » pour désigner le concept qui est référé par « self » en anglais. Je préfère utiliser le terme de « soi » qui permet une description plus objective, à la 3^{ème} personne. Et j'emploie le morphème « auto- » pour « self- ».

Carruthers (2016) reconnaît la pertinence de la conception duale de la pensée, impliquant processus automatiques et non verbaux d'une part et conscients et verbaux d'autre part, mais défend qu'il n'est pas nécessaire de séparer système 1 et système 2, qui peuvent être finalement conçus comme s'appuyant sur les mêmes ressources cognitives. Si Carruthers (2002) donne un rôle au langage à la fois pour la pensée consciente et non-consciente (cf. I.1.2.4), il précise que les pensées non conscientes n'incluent que le lexique et la syntaxe et sont dépourvues de traits phonologiques (i.e. ce sont les formes logiques de Chomsky, 1995). Elles seraient donc moins complètes, moins spécifiées dans le détail ce qui expliquerait qu'elles soient moins accessibles à la conscience. Plus récemment, Carruthers (2009) a explicité le rôle de la verbalisation dans la prise de conscience de ses pensées par le sujet, soit la métacognition. Carruthers (2009) voit la métacognition ancrée dans une capacité pré-existante, la capacité d'attribuer des états mentaux à autrui et de les comprendre (capacité de *mindreading* que l'on peut traduire par « identification d'états mentaux »). La métacognition serait la capacité de *mindreading* appliquée à soi-même plutôt qu'à autrui. Cette capacité accède aux signaux perceptifs, puisque pour comprendre les actions d'autrui, il faut avoir accès aux représentations perceptives de ses actions. Carruthers reprend le cadre théorique de Baars sur l'espace de travail conscient (Baars, 1988, cf. aussi Dehaene & Naccache, 2001). Dans ce cadre, différents processeurs dédiés fonctionnent en parallèle, de façon non consciente. Lorsqu'à un instant donné une coalition de processeurs dominants envoie les résultats de leur traitement dans l'espace de travail global, un message est diffusé à l'ensemble des processeurs, créant l'expérience consciente. Ainsi, selon Carruthers, le système de *mindreading* (un des processeurs dédiés) peut recevoir en entrée toute sorte d'états sensoriels diffusés de façon globale aux différents systèmes de traitements cognitifs. Parmi ces états sensoriels figurent les signaux verbaux (parole intérieure), particulièrement accessibles car audibles. Ainsi, selon Carruthers, nous pourrions être conscients que nous pensons, en interprétant les représentations audibles que nous générons lors de la production de parole intérieure.

Encadré I.8. Langage et métacognition

En résumé, Hegel a formulé l'hypothèse que c'est par l'expression verbale que le sujet prend conscience de sa pensée, que c'est le mot qui permet à la proto-pensée (non verbale) de devenir claire et saisissable. Cette hypothèse a été développée plus avant par Evans, Frankish ou Kahneman, entre autres, qui ont introduit la théorie du double système de pensée, avec un système 1 intuitif, implicite, automatique, rapide, non délibéré et non verbal et un système 2, explicite, analytique, lent délibéré, propositionnel, verbal et propre aux humains. Le raisonnement conscient (et donc la métacognition) mettrait en jeu le système 2 et serait dépendant du langage. Selon Carruthers, systèmes 1 et 2 s'appuient sur les mêmes ressources cognitives et la métacognition émerge du langage intérieur. C'est en interprétant les signaux audibles générés lors de la verbalisation intérieure que nous prenons conscience que nous pensons.

I.1.3.2. La construction du tandem soi/autrui

Si le langage contribue à nous faire prendre conscience que nous pensons, une fois que l'on sait que l'on est en train de penser, comment sait-on que c'est soi qui pense ?

I.1.3.2.1. Langage et conscience de soi : de « Je pense donc je suis » à « Je est un autre »

Descartes, dans le *Discours de la Méthode* (1637) a perçu le lien entre la pensée (et donc, selon lui, le langage, cf. paragraphe I.1.2.1 ci-dessus) et l'identification (la reconnaissance) du sujet pensant par lui-même : « Je pense donc je suis », je pense donc je sais qu'il y a une personne en train de penser et cette personne qui est en train de penser, c'est moi. Pascal fait également le lien entre pensée et conscience de soi lorsqu'il écrit « *le moi consiste dans ma pensée* » (*Pensées*, Laf. 135, Sel. 167). Mais il n'en déduit pas qu'il y ait une substance du soi. L'évidence cartésienne de l'identité de la substance pensante comme étant le soi stable, est même questionnée par Pascal dans les *Pensées* (1670). Selon Pascal, le soi est inconstant, il change selon les moments de la vie, il est contingent aux situations, il est donc insaisissable. En partant de ce même constat que la pensée et le corps ne cessent de changer, Locke (1690) soutient que ce n'est pas l'identité de la substance qui fait l'identité de la personne, mais la mémoire personnelle, la conscience de nos actes et pensées passés. La notion cartésienne de soi comme substance a également été remise en question par Hume (1739), qui considère que l'impression d'un soi unifié et stable est une fiction, car les impressions, les perceptions qui se succèdent les unes aux autres sont changeantes, sont dans un flux et un mouvement perpétuel. Le soi est une hypothèse qui ne correspond à rien dans l'expérience que nous faisons de notre vie intérieure. Selon Hume, il n'y a pas de sujet substantiel. C'est cette découverte que le soi est une illusion, une chimère, que fait Rimbaud à l'âge de 17 ans lors qu'il écrit à Paul Demeny.

« *Car Je est un autre. Si le cuivre s'éveille clairon, il n'y a rien de sa faute. Cela m'est évident : j'assiste à l'éclosion de ma pensée : je la regarde, je l'écoute : je lance un coup d'archet : la symphonie fait son remuement dans les profondeurs, ou vient d'un bond sur la scène.* » (Rimbaud, 1871).

C'est le concept de soi qui nous amène à revendiquer ce qui se trame à l'intérieur de nous comme étant sous notre contrôle propre. Mais, selon le poète éclairé et « voyant », nous n'y sommes pour rien...

Encadré I.9. Langage et prise de conscience de soi

En résumé, le langage permet l'identification (la reconnaissance) du sujet pensant par lui-même, même si le soi est instable et inconstant et si cette reconnaissance est une illusion, une construction.

I.1.3.2.2. Le soi dialogique : subjectivité et altérité

Les réflexions de la fin du XIX^{ème} siècle en philosophie, notamment dans le cadre du courant pragmatiste, puis au XX^{ème} siècle en linguistique et en sciences cognitives ont permis d'avancer de nouveaux éléments sur cette question de l'identité et de la constance du sujet pensant. Ce serait justement le langage (et en particulier le langage intérieur) qui créerait la permanence de l'identité du soi et qui permettrait l'illusion d'un soi intégré et stable dans le temps. Selon Wiley (2006), les philosophes du courant pragmatiste américain (Peirce, James, Mead et Dewey, notamment) ont introduit des idées innovantes sur les liens entre langage intérieur et conception du soi, notamment celle que le soi est un dialogue interne. Selon Peirce, en particulier, c'est par le dialogue interne entre le soi immédiat et un soi en construction, avec qui le soi immédiat dialogue,

comme il le ferait à voix haute avec autrui, que s'échafaude notre identité. Peirce explique ainsi :

« *a person is not absolutely an individual. His thoughts are what he is "saying to himself," that is, saying to that other self that is just coming into life in the flow of time. When one reasons, it is the critical self that one is trying to persuade, and all thought whatsoever is a sign, and is mostly of the nature of language.* »⁸ (Peirce, 1934, p.421).

Ce passage est particulièrement éclairant car il contient deux idées majeures de Peirce : que le soi est dialogique et que la pensée est de nature langagière. Le soi s'élabore, se détermine, se construit, se contrôle, se corrige, à travers le dialogue intérieur. Comme l'explique Petrilli (2013), Peirce envisage même que le soi et le signe soit indissociable, le langage construit le soi et le soi enrichit le langage, de façon réciproque :

« *The man-sign acquires information, and comes to mean more than he did before. But so do words. Does not electricity mean more now than it did in the days of Franklin? Man makes the word, and the word means nothing which the man does not make it mean, and that only to some man. But since man can think only by means of words or other external symbols, these might turn round and say: "You mean nothing which we have not taught you, and then only so far as you address some word as the interpretant of your thought." In fact, therefore, men and words reciprocally educate each other; each increase of a man's information involves and is involved by, a corresponding increase of a word's information.* »⁹ (Peirce, 1868, p. 156).

Et dans le paragraphe suivant : « *the word or sign which man uses is the man himself. (...) Thus my language is the sum total of myself; for the man is the thought.* »¹⁰

Selon Petrilli (2013), la reconnaissance par Peirce de la nature dialogique du soi s'accompagne de l'idée que les notions de soi et d'autrui sont intimement liées. Pour qu'il y ait dialogue, il faut qu'il y ait un autre. C'est parce que le soi a pris conscience d'autrui qu'il peut prendre conscience de sa propre existence en dialoguant intérieurement avec son *alter-ego*. Nous prenons conscience de notre soi à travers l'expérience, à travers nos interactions avec le monde, et en particulier avec les autres, via le langage. Ainsi, le fondateur du pragmatisme américain a reconnu le rôle du langage dans la conscience de soi et d'autrui, ouvrant une perspective sur la troisième fonction du langage à laquelle je m'intéresse ici, la fonction d'auto-néose (voir aussi sur ce sujet les travaux des autres tenants du pragmatisme, notamment James, Mead et Dewey, cf. Wiley, 2006). Cette

⁸ « une personne n'est pas absolument un individu. Ses pensées sont ce qu'il est 'en train de se dire à lui-même', c'est-à-dire ce qu'il dit à cet autre soi qui vient tout juste à la vie dans le flux du temps. Lorsque l'on raisonne, c'est le soi critique que l'on essaie de persuader, et toute pensée de quelque ordre que ce soit, est un signe, et est de nature principalement langagière ».

⁹ « L'homme-signe acquiert de l'information et en vient à signifier davantage qu'avant. Mais les mots font de même. Le mot électricité ne signifie-t-il pas davantage qu'à l'époque de Franklin ? C'est l'homme qui fait le mot et le mot n'a de signification que celle que l'homme lui a donnée, et encore cette signification ne vaut-elle que pour un certain homme. Mais puisque l'homme ne peut penser que par le moyen de mots ou d'autres symboles extérieurs, ceux-ci pourraient bien se retourner et dire : 'Tu ne veux rien dire d'autre que ce que nous t'avons enseigné, et seulement dans la mesure où tu t'adresses à un mot comme interprétant de ta pensée.' Effectivement, donc, les hommes et les mots s'éduquent réciproquement ; chaque accroissement d'information de l'homme implique et est impliqué par un accroissement correspondant de l'information du mot. » (ma traduction, la partie en italique est à vérifier car le texte présente une ambiguïté que je ne suis pas sûre d'avoir résolue)

¹⁰ « Le mot ou le signe que l'homme utilise, c'est l'homme lui-même ». « Ma langue, c'est la totalité de moi-même car l'homme est la pensée. »

conception du rôle du langage dans la construction du soi est proche de celle de Vygotski (1934/1985) selon laquelle la parole intérieure est dialogale et dialogique¹¹, c'est une internalisation de notre propre parole à voix haute mais également de nos échanges sociaux eux-mêmes, i.e. de nos dialogues et conversations. C'est, selon Vygotski, par ces dialogues internes que nous pouvons nous considérer et considérer nos actions selon différentes perspectives : la nôtre et celle d'autrui.

Plus tard, Merleau-Ponty (dans *Causeries*, 1948) a explicité encore davantage comment se forge la conscience de soi, à travers le langage et l'interaction avec autrui :

« Nous ne vivons pas d'abord dans la conscience de nous-même – ni même d'ailleurs dans la conscience des choses – mais dans l'expérience d'autrui. (...) Le contact de nous-même avec nous même se fait toujours à travers une culture, au moins à travers un langage que nous avons reçu du dehors et qui nous oriente dans la connaissance de nous-même. »

Avec les outils de la linguistique, Benveniste (1966) a donné un éclairage renouvelé à la question de la subjectivité dans le langage : *« C'est dans et par le langage que l'homme se constitue comme sujet ; parce que le langage seul fonde en réalité, dans sa réalité qui est celle de l'être, le concept d' "ego". [...] Est "ego" qui dit "ego". [...] La conscience de soi n'est possible que si elle s'éprouve par contraste. Je n'emploie je qu'en m'adressant à quelqu'un, qui sera dans mon allocution un tu. »*

« Le fondement de la subjectivité » dit encore Benveniste *« est dans l'exercice de la langue »*. J'y ajoute celui de l'altérité. C'est, selon Benveniste, par le dialogue et le monologue qui en dérive, que se constituent nos identités, notre sens de la subjectivité et de l'altérité, qui ne s'opposent pas mais au contraire se co-construisent. C'est ainsi que, par le langage, nous acquérons cette capacité de reconnaître à autrui des pensées, des sentiments, des émotions, des attitudes, des intentions, un passé, un présent, un avenir, tous distincts des nôtres. C'est par le langage que cette capacité de découvrir autrui nous forge nous-mêmes, dans un processus de polarité que Benveniste lui-même identifie comme singulier. Car cette (bi-)polarité représente une opposition réversible, une complémentarité indissociable que l'on peut rapprocher de la notion de « non-dualité » exprimée dans certains courants philosophiques ou spirituels (Loy, 2012). Ainsi, explique Benveniste, le langage, notamment intérieur, donne les outils permettant de mieux se représenter soi et autrui : par les pronoms personnels « je » et « tu » et les déictiques (démonstratifs, adverbes, adjectifs) qui organisent les relations spatiales et temporelles avec soi comme origine : « ceci, ici, maintenant » (cf. aussi Wolff, 2009). Wilkinson & Fernyhough (2017) défendent aussi l'hypothèse que le dialogue intérieur contribue à la constitution de connaissances sur soi et à la prise de conscience de soi. Le langage façonne ainsi la distinction entre soi et autrui et le sentiment d'un soi intégré.

Ainsi, de Hume à Benveniste puis Fernyhough, s'est petit à petit forgée la notion de « soi dialogique, langagier et culturel », l'idée selon laquelle le soi est une construction (Hume), qu'il est en devenir permanent et se façonne au cours de nos interactions avec le monde, avec autrui, avec nous-mêmes et par le langage (Peirce, Merleau-Ponty, Benveniste, Fernyhough, entre autres).

¹¹ Les termes de « dialogal » et « dialogique » font débat. Je me réfère à Bres (2005) pour la distinction entre les deux. Il me semble que ce qu'entendait Vygotski était à la fois « dialogal », au sens de conversation, d'alternance entre plusieurs tours de parole produits par des locuteurs distincts et « dialogique », au sens d'interaction de l'énoncé en cours avec d'autres discours (du locuteur lui-même ou d'un allocutaire). Cf. I.2.2.1 pour plus d'explications.

Encadré I.10. Dialogue et prise de conscience de soi et d'autrui

En résumé, c'est par le dialogue interne entre soi immédiat et soi en construction que s'échafaude notre identité : le soi peut être vu comme un alter-ego (Peirce, James, Mead, Dewey, Vygotsky). Pour Merleau-Ponty, la conscience de soi se forge même à travers le langage et le dialogue avec autrui. Benveniste note justement que le langage donne les outils pour se représenter soi et se distinguer d'autrui : *je, tu, ceci, ici, maintenant*. Le dialogue contribue ainsi à la stabilité de l'identité du soi.

I.1.3.2.3. Le soi narratif

Cette conception que le soi repose sur le langage et l'environnement culturel a été importée dans le domaine des sciences cognitives par Dennett (1988, 1991) qui a introduit la notion de « soi narratif ». L'hypothèse de Dennett est que le soi est le résultat d'une construction narrative (passant donc par le langage) nécessaire à la vie de l'individu. Le soi narratif comporte une dimension temporelle, incluant les souvenirs du passé et les intentions futures. Nous utilisons donc le langage pour créer des récits et à travers ces récits nous créons notre identité, notre soi. Dennett conçoit le soi comme un centre de gravité narratif abstrait, point d'intersection des différents récits sur le sujet, construits par le sujet lui-même et par les autres.

Ricœur (1983-85, 1990) aboutit à des conclusions similaires sur la construction du soi narratif, sauf qu'il envisage le soi narratif non pas comme un point de convergence abstrait des différentes narrations, mais comme un soi plus concret d'une part, et distribué, décentré, d'autre part. Ricœur a introduit une distinction entre *identité* de la personne et *ipséité*. La « mêmété », ou l'identité comprise au sens d'un « même » (*idem*), suppose une permanence dans le temps, s'oppose au changeant, au variable. L'ipséité, l'identité au sens d'un soi-même (*ipse*), ne présuppose pas l'immuabilité et intègre la pluralité temporelle et narrative. Wolff (1997) a formulé les définitions suivantes :

« Par "identité", nous entendons ce qui fait qu'un individu demeure le même numériquement tant qu'il est » et « par "ipséité", nous entendons ce qui fait qu'un individu est lui-même et non un autre, c'est-à-dire ce en quoi il est unique et différent de tous les autres de son espèce » (Wolff, 1997, p. 72).

Selon Gallagher (2000), la notion distribuée de Ricœur représente le soi comme la somme de tous ces narratifs, incluant toutes les contradictions, les tergiversations, les équivoques, les difficultés et tous les messages cachés qui s'expriment dans la vie personnelle. Un soi distribué permet ainsi de mieux rendre compte des conflits intérieurs, de l'indécision morale et de l'auto-illusion qu'un soi abstrait ponctuel. Ainsi, selon Dennett et Ricœur, le langage, et en particulier la parole intérieure, contribue à la construction du soi et à l'ipséité par la création et l'intégration de récits intériorisés, auto-crésés ou créés par autrui.

Le témoignage de Eva Hoffman, dans son ouvrage autobiographique « Lost in translation » illustre de façon poignante le rôle du langage dans la construction narrative du soi. D'origine polonaise, sa famille émigre au Canada, alors qu'elle a 13 ans. Elle se met à apprendre l'anglais et connaît alors une phase de désintégration de sa parole intérieure, son polonais n'étant pas adapté aux nouvelles situations vécues et son anglais n'étant pas encore suffisamment fluide pour relater ses nouvelles expériences :

« *The worst losses come at night. As I lie down in a strange bed in a strange house ... I wait for that spontaneous flow of inner language which used to be my nighttime talk with myself ... Nothing comes. Polish, in a short time, has atrophied, shriveled from sheer uselessness. Its words don't apply to my new experiences; they're not coeval with any of the objects, or faces, or the very air I breathe in the daytime. In English, words have not penetrated to those layers of my psyche from which a private conversation could proceed. This interval before sleep used to be the time when my mind became both receptive and alert, when images and words rose up to consciousness, reiterating what had happened during the day, adding the day's experiences to those already stored here, spinning out the thread of my personal story. Now, this picture-and-word show is gone; the thread has been snapped. I have no interior language, and without it, interior images – those images through which we assimilate the external world, through which we take it in, love it, make it our own – become blurred too.* » (Hoffman, 1989, p. 107–108).

Elle entrevoit ainsi clairement le rôle crucial du langage intérieur dans l'élaboration de l'histoire personnelle, dans la construction du soi sous forme de récits verbalisés :

« *I understood how much our inner existence, our sense of self, depends on having a living speech within us. To lose an internal language is to subside into an inarticulate darkness in which we become alien to ourselves; to lose the ability to describe the world is to render that world a bit less vivid, a bit less lucid.* » (Hoffman, 1989, p. 48).

Le langage est donc essentiel à l'élaboration de récits sur notre histoire personnelle, constitutifs de notre conscience de soi.

Encadré I.11. Le soi narratif

En résumé, le soi résulte d'une construction narrative, de récits sur nos souvenirs du passé et nos intentions futures. D'après Dennett, le soi narratif est le centre de gravité abstrait des différents récits sur le sujet. Pour Ricœur, le soi narratif est concret, distribué, changeant, non immuable, en d'autres termes, il y a ipséité (soi-même) mais pas identité (mêmeté) de la personne. Le sentiment de désintégration du soi parfois observé chez les immigrés allophones en phase d'apprentissage de la langue ambiante peut s'expliquer par le rôle fondamental de la pratique du langage dans la construction du soi.

I.1.3.2.4. La conscience de soi minimale et le développement de la conscience de soi

Bermúdez (2001) remarque que la construction de la conscience de soi par le langage conduit à un paradoxe. Car pour être capable d'auto-réflexion verbale, l'individu doit pouvoir utiliser le pronom à la 1^{ère} personne (« Je »). Or utiliser ce pronom signifie que l'on a une conscience de soi. Ce qui revient à dire que pour construire une conscience de soi, il faut une conscience de soi. Pour résoudre ce paradoxe, Bermúdez (2001) suggère qu'il existe une distinction entre la conscience de soi élaborée, incluant la capacité de conduire des réflexions sur soi-même, qui s'appuierait sur le langage, et une conscience de soi minimale, primitive, qui, elle, ne nécessiterait pas le langage. Il a introduit la notion de « conscience de soi non-conceptuelle », qui reposerait sur des expériences perceptives pures, sans médiation verbale. Ainsi, par exemple, la proprioception de la position de son corps par un individu ne recourt pas au langage, mais elle contribue à la conscience du corps et à se situer soi-même dans l'environnement. Bermúdez (2001) s'appuie sur des travaux expérimentaux sur les nourrissons pour démontrer que cette

conscience de soi non-conceptuelle existe dès la naissance, ce que Neisser (1988) a appelé le « soi écologique ». Ainsi, il a été suggéré que la contingence visuo-proprioceptive pourrait permettre au nourrisson, dès 5 mois, d'associer le mouvement ressenti de son corps avec l'observation de son corps en mouvement (Bahrick & Watson, 1985). Cette contingence permettrait la construction du soi écologique, l'expérience d'un soi différencié, situé dans un lieu, avec un corps ayant des frontières délimitées. Il semble, par ailleurs, que le nourrisson puisse, dès la naissance, distinguer les enregistrements de ses propres cris de ceux d'autres nourrissons (Martin & Clark, 1982) et soit capable d'une proprioception permettant l'imitation d'expressions faciales (Metzoff & Moore, 1977).

Dans ce contexte, les travaux récents de Blanke sur la conscience de soi corporelle ou somatique sont éclairants. Blanke (2012) explique que ce type de conscience repose sur trois aspects : l'auto-identification avec le corps (l'expérience de posséder un corps), l'auto-localisation (expérimenter où l'on se situe dans l'espace) et la perspective à la 1^{ère} personne (expérimenter d'où l'on perçoit le monde). Il montre que les processus en jeu dans les différents aspects de la conscience de soi corporelle reposent sur des signaux intéroceptifs et homéostatiques. Ce type de conscience pourrait être à la base d'une conscience de soi narrative plus élaborée, mettant en jeu, elle, le langage. Cette vision résonne avec la proposition de Bermúdez (2001) pour résoudre le paradoxe du développement de la conscience de soi.

Les études sur le développement de la maîtrise des pronoms « je » et « tu » par les jeunes enfants semblent aller dans le sens de cette hypothèse. Ainsi, Bates (1990) repère trois étapes dans l'acquisition de la référence pronominale chez les enfants. La première étape, qui concerne les nourrissons de 9 à 18 mois, est une étape de référence non-contrastive, au cours de laquelle les enfants montrent une capacité de communication référentielle en triangle impliquant l'enfant comme sujet, l'adulte comme partenaire de la communication et un objet, comme objet de référence partagée. Cette étape commence par des gestes de monstration, l'enfant tendant l'objet en direction de l'adulte et attendant un acquiescement. La phase suivante consiste en des gestes de don, l'enfant donnant un ou plusieurs objets à l'adulte. Et le stade développemental clé est celui du pointage communicatif, au cours duquel l'index étendu, le bras tendu, l'enfant indique à l'adulte l'objet ou l'événement intéressant. C'est une étape fondamentale d'individuation de l'objet par le geste déictique. Le geste de pointage individualise sa référence. Au cours de cette étape de référence non-contrastive, le langage de l'enfant se développe, mais les pronoms restent très rarement employés. Selon Bates, la seconde étape, celle de la référence contrastive, démarre autour de 19-20 mois pour s'achever autour de 28 mois. C'est à ce stade que les enfants commencent à employer les pronoms de façon contrastive, en opposant notamment « je » à « tu ». C'est également à cette étape développementale qu'ils utilisent leur propre prénom, ce qui semble indiquer l'émergence consciente du concept de soi chez l'enfant. Et, de façon intéressante, cette étape de contraste linguistique rudimentaire entre soi et autrui, coïncide avec des indicateurs non langagiers de l'émergence d'un concept de soi et d'autrui explicite et conscient, une amorce de cognition sociale (Kagan, 1981 ; Lewis & Brooks-Gunn, 1979 ; Bretherton & Beeghly, 1982). Enfin, la troisième étape, de référence paradigmatique, commence à 28 mois et se poursuit tout au long de la vie. C'est alors que les différentes versions du soi social commencent à être envisagées et que l'enfant apprend à naviguer parmi les différentes attentes que son environnement a de son moi « multidimensionnel » (cf. aussi Frith, 2010 : le concept de soi se développe en partie par

ce que les autres disent et pensent de soi). Ainsi, selon Bates (1990), le langage et les compétences sociales (dont l'élaboration de la théorie de l'esprit) interagissent au cours du développement : le langage fournit à l'enfant des clés pour s'adapter à son environnement social et, réciproquement, ses expériences sociales enrichissent ses compétences linguistiques.

Encadré I.12. Développement de la conscience de soi chez l'enfant

Une conscience de soi minimale, constituée d'expériences perceptives pures, sans médiation verbale, semble exister chez les nourrissons pré-verbaux et serait fondée sur la contingence audio-visuo-perceptive et la perception des frontières du corps. La conscience de soi élaborée serait échafaudée sur la conscience de soi minimale et mettrait en jeu le langage, les outils lexicaux et syntaxiques. On observe ainsi des étapes du développement de la conscience de soi élaborée, avec l'apparition du pointage et de la capacité de monstration entre 9 et 18 mois, suivie de l'émergence d'un concept de soi et d'autrui, avec l'utilisation de l'opposition entre les pronoms « je » et « tu » entre 19 et 28 mois puis de la capacité d'envisager différentes versions du soi social, accompagnées de compétences linguistiques enrichies.

I.1.3.2.5. Le soi temporellement étendu (le soi dans le temps) : la conscience autoéotique

On l'a vu ci-dessus, ce qui distingue les êtres humains des autres animaux n'est pas la capacité de communiquer (cf. I.1.2.2). Ce n'est pas non plus la capacité de mener des raisonnements simples (cf. I.1.2.4) ou d'apprendre (Weiskrantz, 1985). Selon Tulving (2002), ce n'est pas non plus simplement la conscience de soi. Une conscience de soi minimale existe chez certains animaux, notamment les chimpanzés et les orangs-outans qui sont capables de se reconnaître dans un miroir (Gallup, 1977). Selon de nombreux chercheurs contemporains, la spécificité humaine résiderait dans la capacité de voyager mentalement dans le temps et de disposer ainsi d'un système de mémoire épisodique (permettant le rappel d'épisodes personnellement vécus avec la récupération d'informations liées à cette expérience personnelle, superbement illustré par Proust avec le « goût du morceau de madeleine trempé dans le tilleul ») lié à une conscience de soi étendue dans le temps (e.g. Tulving, 2002 ; Nelson & Fivush, 2000 ; Suddendorf & Corballis, 1997). Ce lien crucial entre mémoire et conscience de soi a été envisagé depuis Platon (avec la doctrine de la réminiscence) et a été explicité par Saint Augustin (397-401) qui montre que c'est la mémoire qui rend possible le sentiment de soi, avec l'évocation de souvenirs passés et l'imagination de situations futures :

« C'est au-dedans que j'accomplis ces actes, dans la cour immense du palais de ma mémoire. Oui, là, le ciel et la terre et la mer sont à ma disposition, avec toutes les sensations que j'ai pu trouver en eux, hormis celles que j'ai oubliées. Là, je me rencontre aussi moi-même; je me souviens de moi, de ce que j'ai fait, quand et où je l'ai fait et quelle impression j'ai ressentie quand je le faisais. Là se trouve tout ce dont j'ai fait l'expérience ou que j'ai cru, et dont je me souviens. De la même abondante réserve, je tire également par ressemblance avec les choses dont j'ai fait l'expérience ou auxquelles, d'après cette expérience, j'ai cru, je tire d'autres et d'autres images; je les relie moi-même à la trame du passé et, de là, je tisse même celle de l'avenir, actes, événements, espérances; et je pense et repense tout cela comme si c'était du présent. «Je ferai ceci ou cela, me dis-je en moi-même, là, dans les replis immenses de mon esprit, pleins d'images de choses si

nombreuses et si grandes et il s'ensuit ceci ou cela. « Oh! s'il arrivait ceci ou cela ! ». Que Dieu écarte ceci ou cela » Voilà ce que je dis en moi-même, et, dès que je le dis, se présentent les images de tout ce que je dis, jaillissant de ce même trésor de la mémoire; je ne dirais absolument rien de tout cela, si elles m'avaient manqué. » (Confessions, livre X, chapitre VIII, Traduction de E. Tréhorel et G. Bouissou).

Cette capacité de voyage mental évoquée par Saint Augustin serait donc, d'après les recherches contemporaines sur la cognition animale, une spécificité humaine. La conscience de soi minimale des grands singes ne serait qu'une conscience de soi au présent, alors que la conscience de soi humaine existe dans un temps subjectif et permet la mémoire épisodique. Certains animaux non humains (mammifères et oiseaux) possèdent une mémoire sémantique : ils peuvent mémoriser des connaissances sur le monde et auraient ainsi une conscience noétique (une conscience de l'information sur le monde, un accès à des connaissances). Mais la mémoire épisodique implique la capacité de répondre aux quatre questions : « quoi, quand, où et qui » sur un même événement encodé en mémoire (Tulving, 1972, complété par Schwartz & Evans, 2001). Elle implique aussi la capacité de voyager mentalement dans le temps, de revivre mentalement un événement passé ou d'anticiper un événement futur (Tulving, 1983). Schwartz & Evans (2001) montrent que certains grands singes sont capables de récupérer les informations répondant aux questions « quoi » et « qui » (qui a donné la nourriture, quelle nourriture est donnée) et que certains oiseaux (comme les geais) peuvent mémoriser « quoi », « où » et « quand » (quelle nourriture a été enfouie, où elle est cachée et quand elle a été cachée). Schwartz & Evans (2001) suggèrent donc qu'une forme de mémoire proche de la mémoire épisodique humaine existe chez certains animaux, même si la capacité de récupérer des informations relatives aux quatre questions, de façon intégrée, sur un événement du passé ne semble avoir été démontrée chez aucun animal. De plus, la capacité de voyager dans le temps (dans le passé comme dans le futur correspond, selon Schwartz & Evans (2001), à deux états subjectifs : la sensation de véracité ou de conviction (la conviction que l'événement que l'on est en train de se remémorer s'est passé ainsi) et la sensation de passéité (savoir que l'événement que l'on est en train de se remémorer n'a pas lieu dans le présent mais s'est déroulé dans le passé). Plutôt que la sensation de passéité, il faut peut-être utiliser le terme de sensation d'ailleurs temporo-spatial, qui inclut ainsi la sensation que l'événement remémoré a eu lieu dans le passé ou aura lieu dans le futur, dans un lieu peut-être différent. Seuls les humains possèderaient la conscience auto-noétique associée à la mémoire épisodique, c'est-à-dire la capacité de connaître et de réfléchir sur le contenu, la véracité et la nature de leurs expériences personnelles, de les revivre, d'en anticiper de nouvelles. Selon Droege (2013), la mémoire épisodique, qui relie les expériences passées, présentes et futures, ou plus précisément qui intègre l'expérience passée (ou future) dans le présent, permet une conscience de soi étendue temporellement. C'est ce qui permet aux humains de suivre, de mémoriser et d'anticiper les états mentaux des autres et de soi-même, au fil du temps.

Plus que la mémoire épisodique, ce serait la mémoire autobiographique qui jouerait un rôle majeur dans la construction et le maintien d'une conscience de soi (Wilson & Ross, 2003). Elle est définie par Conway & Rubin (1993) comme la mémoire des événements de la vie personnelle, c'est un système dynamique permettant de stocker des souvenirs personnels et des connaissances sur soi. Elle a donc une composante épisodique (les événements de la vie personnelle, situés dans le temps et l'espace) impliquant une

conscience auto-noétique, ainsi qu'une composante sémantique (les informations accumulées sur soi tout au long de la vie, sans contexte spatio-temporel), impliquant une conscience noétique. Elle est liée à nos objectifs de vie, à nos émotions et aux sens que nous donnons aux événements. Conway (2005) montre que le lien entre mémoire autobiographique et conscience de soi est bidirectionnel. En effet, la mémoire autobiographique participe à la construction d'un soi intégré, mais le soi influence la récupération d'informations en mémoire (Greenwald, 1980). Rathbone, Moulin & Conway (2008) expliquent que la distribution temporelle des souvenirs autobiographiques n'est pas régulière tout au long de la vie. Il existe un pic de réminiscence, pour les souvenirs encodés entre l'âge de 10 et 30 ans (ou 15-25 ans, Conway & Rubin, 1993). Cette période de la vie, de l'adolescence à l'âge adulte, serait celle de l'émergence du soi stable et l'encodage des souvenirs serait donc privilégié sur cette période, constitutive du soi. Plus précisément, Rathbone *et al.* montrent qu'il existe en fait plusieurs pics de réminiscence, centrés sur les différents aspects du soi et les périodes où ces aspects se sont construits (« je suis une scientifique », « je suis une compagne », « je suis une mère »).

La mémoire autobiographique et la conscience de soi étendue temporellement sont donc intimement liées et semblent différencier l'humain des autres animaux. C'est à travers la mémoire autobiographique que le soi est constitué, défini, qu'il peut être régulé. Les troubles de la mémoire autobiographique peuvent entraîner des troubles de la conscience de soi. Dans le cas de l'amnésie, la sensation de cohérence du soi peut disparaître (Conway *et al.*, 2004). Une mémoire autobiographique défaillante peut être associée à des troubles de l'identité personnelle dans le cas de maladies neurologiques ou psychiatriques, comme la schizophrénie (Conway, 2005).

La mémoire autobiographique s'appuie sur des constructions narratives qui permettent d'organiser les événements de façon cohérente dans le temps (Conway, 2005). Les rôles du langage, des interactions sociales et de la culture dans la mémoire autobiographique semblent ainsi indéniables. Toutefois la question de la nécessité du langage dans le développement de la mémoire autobiographique chez l'enfant reste débattue. Certains chercheurs considèrent que le langage est central car il permet de transformer le simple rappel d'expériences en leur donnant une perspective et une interprétation liées à l'histoire personnelle (Bruner, 1987 ; Nelson, 2005 ; Fivush, 2011). D'autres travaux nuancent cette position. Il est généralement admis que les adultes ont peu de souvenirs datant d'avant l'âge de 7 ans, encore moins précédant leurs 5 ans et quasiment aucun datant de la période pré-verbale, i.e. avant 2 ans, (Bauer, 2007 pour une revue). On parle d'amnésie infantile pour ce déficit de mémoire autobiographique entre la naissance et deux ans. Toutefois, il semble que les enfants avant l'âge de 7 ans et même avant 2 ans aient des souvenirs liés à des expériences spécifiques, mais qu'ils ne puissent y accéder verbalement. Les souvenirs qui peuvent être remémorés verbalement seraient liés au développement du langage, mais des souvenirs semblent pouvoir être formés avant l'acquisition du langage. Ces souvenirs ne pourraient être accédés que visuellement (reconnaissance de photographies) et resteraient inaccessibles verbalement (Simcock & Hayne, 2003 ; Morrison & Conway, 2010). Dans une revue de la littérature sur la mémoire épisodique et autobiographique chez l'enfant, Picard, Eustache & Piolino (2009) expliquent que dans les deux premières années de vie, les souvenirs autobiographiques sont quasiment absents. Les enfants sont simplement capables d'évoquer le passé sous forme de bribes de souvenirs, d'actions, de comportements, de sensations. Ces traces

mnésiques ne sont pas des souvenirs autobiographiques car ils ne sont pas associés à un événement unique dans le temps. Ils ne sont pas non plus verbalisables ultérieurement. Ce serait le développement du langage qui permettrait, plus tard, la structuration des connaissances personnelles et l'établissement de souvenirs autobiographiques détaillés et organisés, contribuant ainsi à la conscience de soi temporellement étendue.

Le cas de Helen Keller devenue sourde et aveugle à l'âge de 19 mois, illustre de façon poignante le rôle fondamental du langage dans le développement de la conscience de soi. Helen Keller (1908) écrit qu'avant son apprentissage du langage à 7 ans grâce à son éducatrice, qui lui apprit à communiquer et à utiliser la langue des signes et le braille, elle n'avait pas la conscience d'exister :

« Before my teacher came to me, I did not know that I am. I lived in a world that was a no—world. I cannot hope to describe adequately that unconscious, yet conscious time of nothingness. I did not know that I knew aught, or that I lived or acted or desired. I had neither will nor intellect(...) My inner life, then, was a blank without past, present, or future, without hope or anticipation, without wonder or joy or faith. »¹² Keller (1908, p. 114).

Ainsi il semble bien que le langage ait une fonction d'auto-noèse, c'est-à-dire qu'il permette non seulement la conscience de soi, mais une conscience de soi inscrite dans le temps. Ce rôle important du langage dans l'auto-noèse a une conséquence forte chez les personnes bilingues ou multilingues. Pavlenko (2014) défend l'idée que chez les bilingues la conscience de soi peut être multiple. Elle rapporte que, dans un questionnaire adressé à 1039 personnes bi- ou multi-lingues, à la question posée « avez-vous parfois l'impression d'être une personne différente lorsque vous utilisez vos différentes langues ? », 65% des personnes interrogées à travers le monde ont répondu positivement (Pavlenko, 2006). S'appuyant sur cette étude et d'autres travaux menés par d'autres chercheurs, elle estime que pour certains bi- ou multi-lingues, il existe différentes versions du soi, en fonction de la langue, et que ces multiples versions fragilisent la cohérence du sentiment de soi.

Encadré I.13. Langage et conscience auto-noétique

En résumé, la conscience de soi dans le temps est fondée sur la mémoire autobiographique, qui s'appuie sur des constructions narratives, qui organisent temporellement les souvenirs. Le langage contribue à l'illusion d'un soi intégré et stable dans le temps en préservant nos pensées, en enrichissant notre soi narratif, au cours du temps. Il semble donc intervenir dès l'émergence de la conscience auto-noétique et contribue à la consolider, à la façonner tout au long de la vie.

I.1.3.2.6. La pratique du langage intérieur et le renforcement de la conscience de soi

Si l'une des spécificités humaines semble bien la capacité de conscience auto-noétique, renforcée et alimentée par le langage, un deuxième mécanisme est constitutif de la

¹² « Avant que mon éducatrice vienne à moi, je ne savais pas que je suis. Je vivais dans un monde qui était un non-monde. Je ne peux espérer décrire adéquatement cette période de néant inconscient pourtant conscient. Je ne savais pas que je savais quoi que ce soit, ni que je vivais ou que j'agissais ou désirais. Je n'avais ni volonté ni intelligence. (...). Ma vie intérieure d'alors était un vide sans passé, présent ni futur, sans espoir ni attente, sans émerveillement, ni joie, ni foi. »

cognition humaine, celui de l'inhibition et de la simulation. C'est parce que la production de langage à voix haute peut être inhibée et simulée intérieurement que les êtres humains peuvent passer une grande partie de leur temps à construire et développer mentalement leur ego, sans embarrasser la terre entière.

Comme décrit dans la section I.1.2.2, Chomsky soutient que, contrairement à l'hypothèse courante que la fonction première du langage est la communication, le langage a évolué d'abord pour permettre la pensée élaborée et qu'il est même mal adapté à la communication. Avec Berwick, défendant le programme « biolinguistique », il soutient même que le langage a commencé par être pratiqué de façon intérieure avant d'être externalisé et utilisé à voix haute pour communiquer :

« language evolved as an instrument of internal thought, with externalization a secondary process » (Berwick & Chomsky, 2011, p. 32)

Cette hypothèse paraît difficilement conciliable avec les remarques ci-dessus (cf. I.1.2.4) sur la possibilité de mener une pensée conceptuelle sans langage, même intérieur (chez certains animaux, chez les nourrissons, chez certains artistes et scientifiques de haut niveau, chez certains sourds non verbaux). Si la pensée est possible sans langage, alors la pression évolutive pour le langage ne peut être uniquement liée à un besoin d'augmenter les capacités computationnelles mentales. Il semble plus raisonnable de proposer l'inverse, c'est-à-dire que l'utilisation du langage s'est d'abord faite de façon externalisée, sous la pression sociale de communication accompagnée probablement de la pression cognitive. La communication avec autrui, la prise en compte des points de vue d'autrui, ainsi que des habitudes langagières d'autrui enrichit notre propre répertoire langagier et par là même notre propre pensée. Comme le développe Bakhtine, le langage est, par essence, dialogique¹³ et hétéroglossique :

« Our speech, that is, all our utterances (including creative works), is filled with others' words, varying degrees of otherness or varying degrees of "our-own-ness", varying degrees of awareness and detachment. These words of others carry with them their own expression, their own evaluative tone, which we assimilate, rework, and re-accentuate. » (Bakhtin, 1986, p. 89).

Au cours du développement du langage, le nourrisson, puis l'enfant, tente de reproduire les énoncés produits par son entourage, ou s'en inspire. Nous pouvons refaire cette expérience de l'intégration des énoncés d'autrui lorsque nous voyageons et devons parler une langue étrangère : il nous arrive alors répéter mentalement des constructions langagières entendues autour de nous. Nous les intégrons alors à notre répertoire de sons et énoncés préfabriqués dans cette langue. C'est ce que fait remarquer très adroitement Eva Hoffman (déjà citée plus haut, cf. I.1.3.2.3) dans son roman autobiographique, qui livre l'expérience d'une adolescente émigrée cherchant à maîtriser sa seconde langue avec détermination.

« Since I lack a voice of my own, the voices of others invade me as if I were a silent ventriloquist. They ricochet within me, carrying on conversations, lending me their modulations, intonations, rhythms. I do not yet possess them; they possess me. But some of them satisfy a need; some of them stick to my ribs. ... Eventually, the voices enter me; by assuming them, I gradually make them mine. » (Hoffman, p. 219–220).

¹³ Cf. note 11 ci-dessus

Ainsi, ce sont, selon moi, les pressions conjuguées de la complexification de la communication et du développement de la pensée, associée à une troisième pression, celle du renforcement du sentiment de soi, qui ont conduit au langage humain actuel. Et les deux dernières (pensée et construction de soi) ont favorisé sa version intérieure.

C'est parce que le langage permet d'étayer la pensée, mais aussi de focaliser notre attention sur notre propre pensée et de placer le soi comme objet de réflexion, qu'il s'est internalisé. On peut donc proposer, à l'inverse de Berwick & Chomsky (2011), que l'internalisation est venue secondairement à la parole à voix haute. C'est d'ailleurs ce que l'on observe lors du développement du langage chez l'enfant. Comme le note Vygotski (1934/1985), la parole intérieure paraît héritée de la parole à voix haute, via un processus graduel d'internalisation, au cours duquel le discours privé de l'enfant, d'abord produit à voix haute (ce que Piaget a nommé le « langage égocentrique », cf. ses descriptions détaillées dans Piaget, 1923), devient petit à petit intérieur, à partir de l'âge de 5 à 7 ans (cf. aussi Conrad, 1971 ; Cooley, 1922 ; Fernyhough, 2016). Même si des résultats récents invitent à nuancer cette hypothèse (Mani & Plunkett, 2010 ; Ngon & Peperkamp, 2016) et montrent que des prémices de langage intérieur existent bien dès l'âge de 20 mois, il reste que la pratique du discours privé à voix haute est peu inhibée avant l'âge de 5 ans.

Et c'est, selon Jackendoff (1996), parce que nous sommes capables d'internaliser le langage, de le pratiquer mentalement, que nous pouvons focaliser notre attention sur nos propres pensées, les modifier, les mémoriser. Le langage intérieur permet de complexifier la pensée, mais il permet aussi d'en prendre conscience (la méta-cognition) et de la préserver dans le temps (la conscience auto-noétique). Jackendoff (2011) distingue la pensée de la parole intérieure et explique que sans la pratique du langage intérieur, nous n'aurions quasiment pas conscience de nos pensées. Selon Jackendoff, la pensée, sous forme de « structure conceptuelle formelle », est presque entièrement non-consciente et ce que l'on ressent consciemment comme notre monologue intérieur est en fait la structure phonologique associée à la pensée :

« conceptual structure, that is, the formal structure of the thought conveyed by a sentence, is almost completely unconscious, and that what we EXPERIENCE as our inner monologue is actually the phonological structure linked to the thought. We are aware of our thinking because we hear the associated sounds in our head. This analysis immediately explains the fact that we seem to be 'thinking in a language'. » (Jackendoff, 2011, p. 613).

Selon Jackendoff, la conscience n'est donc pas liée à l'acte de pensée initial, mais à sa transformation en forme phonologique, audible mentalement. La production de langage intérieur nous permet donc de prendre conscience de nos pensées, de nous focaliser sur elles. Nous pouvons prendre conscience de nos pensées via le langage intérieur, mais ceci ne signifie pas que nos pensées soient sous forme de mots ou de représentations linguistiques ou phonologiques. Quand nous entendons notre voix intérieure, nous nous écoutons nous parler à nous-mêmes. Et dans cette conversation avec nous-même nous utilisons les mêmes moyens linguistiques que ceux que nous utilisons avec autrui pour exprimer notre pensée. La parole intérieure est une traduction en mots, pour nous-mêmes, du contenu de nos pensées, nous faisant prendre conscience de notre pensée, nous la rendant accessible, nous permettant d'y porter notre attention. Ceci rappelle la lettre de Rimbaud à Demeny, déjà citée plus haut (cf. I.1.3.2.1 ci-dessus) :

« j'assiste à l'éclosion de ma pensée : je la regarde, je l'écoute. » (Rimbaud, 1871).

Ce point de vue sur l'implication du langage intérieur dans la prise de conscience de nos pensées ou activités mentales est partagé par de nombreux chercheurs. On le retrouve chez Vygotski (1934/1985), Clark (1998), Clowes (2007) ou Martínez-Manrique & Vicente (2010). Ce mécanisme de transformation de la pensée verbale abstraite en signal mental audible me semble un point clef dans l'explication de la métacognition et de la prise de conscience de soi et je tente d'en rendre compte dans la modélisation du langage intérieur proposée dans la section I.2.2.

Il faut noter cependant que des formes de pensée peuvent être perçues consciemment alors qu'elles ne sont pas verbales. Le témoignage d'Einstein cité ci-dessus (cf. I.1.2.4.4), ou les données et hypothèses sur le codage dual de la pensée, verbal et visuel (*Dual coding theory*, Paivio, 1990), invitent à nuancer l'hypothèse selon laquelle seule la production de parole intérieure rendrait la pensée consciente. Je fais l'hypothèse qu'un même mécanisme de transformation est à l'œuvre dans la pensée sur un mode visuel que celui à l'œuvre dans la pensée qui devient verbale. C'est parce que nous transformons nos pensées en représentations phonologiques audibles que nous pouvons y prêter attention et c'est probablement parce que nous sommes aussi capables, du moins certains d'entre nous, de les transformer en représentations visuelles que nous pouvons en prendre conscience.

Toutefois, la construction auto-néotique, l'élaboration de récits sur soi passent probablement uniquement par le langage qui permet, par l'étendue du lexique, par les nombreux dispositifs syntaxiques – comme l'utilisation des pronoms personnels pour désigner différents types de relations entre les personnes, l'emploi différencié des temps des verbes (passé, présent, futur), de leurs aspects (passé simple, imparfait), de leurs modes (indicatif, subjonctif, impératif, conditionnel), de leurs voix (active, passive) – de structurer la narration, de la complexifier, de la détailler. Ainsi la pratique du langage intérieur renforce et enrichit notre conscience auto-néotique. Je rejoins donc Jackendoff (2011) et surtout Martínez-Manrique & Vicente (2010) dans l'idée que c'est la pratique du langage intérieur qui nous permet de focaliser notre attention sur notre pensée, de suivre et contrôler nos processus cognitifs avec une perspective personnelle, individuelle, mais aussi de placer le soi comme objet de réflexion, de nourrir les récits sur nous-mêmes (le soi narratif), de renforcer notre sentiment de nous-mêmes, dans le temps.

C'est ainsi la capacité d'inhiber notre production de parole et de la simuler mentalement qui serait la clef du développement de la cognition tel qu'on l'observe chez l'humain. Je me permets donc une conjecture :

La spécificité humaine ne réside pas dans la capacité de communiquer (*a minima*), ni dans celle de penser (rudimentairement), mais dans la capacité de simuler mentalement l'acte de parole.

Une forme de contrôle inhibiteur existe chez les primates, capables de supprimer une pulsion et retarder une action si une gratification plus importante est attendue (Amici *et al.*, 2008). Une forme de simulation mentale des actions d'autrui a été démontrée également. Les neurones miroirs découverts par l'équipe de Rizzolatti chez le singe macaque rhésus, sont activés lorsque le singe exécute une action ou voit un congénère effectuer cette même action. Ils semblent fournir la base d'un comportement de simulation motrice qui serait à l'œuvre chez ces singes pour interpréter le comportement de leurs congénères. Whiten (2013) discute de l'existence de capacités de théorie de l'esprit et de lecture des états mentaux d'autrui chez les primates et montre que certains

résultats ont été sur-interprétés. Il admet toutefois que les chimpanzés notamment sont capables d'estimer ce que voit leur congénère, et d'adapter leur action en fonction de ce qu'ils estiment être la perspective de l'autre. Les comportements d'inhibition et de prise en compte de la perspective visuelle d'autrui constituent les pré-requis d'une simulation mentale volontaire (et non pas automatique comme dans le cas de l'observation d'action) de l'action. Toutefois, je ne connais pas d'études permettant de conclure que les animaux sont capables délibérément de simuler mentalement une action. Ce serait donc là notre spécificité humaine.

Une grande partie de notre temps est consacrée à simuler et imaginer des actions, pour les revivre, pour les prédire, pour les répéter, pour nous entraîner à les faire mieux. Parmi ces actions simulées, celle de langage nous occupe beaucoup et nous l'avons vu, joue un rôle crucial dans la métacognition et la conscience auto-noétique. L'implication du langage intérieur dans la construction de soi est d'ailleurs bien reconnue dans les travaux récents en sciences cognitives. S'appuyant sur des données objectives comportementales et neurophysiologiques ainsi que sur des rapports introspectifs de patients cérébro-lésés ou sur des questionnaires et des entretiens guidés, Morin (2005, 2018), Perrone-Bertolotti *et al.* (2014), Alderson-Day & Fernyhough (2015), Alderson-Day *et al.* (2018), Fernyhough (2016) et McCarthy-Jones & Fernyhough (2011) ont répertorié les différents rôles que le langage intérieur, joue dans diverses fonctions cognitives qui relèvent de la conscience de soi et/ou de la conscience d'autrui. Ainsi, il a été suggéré que le langage intérieur permet de raisonner sur autrui et d'attribuer à autrui des pensées, différentes des nôtres (théorie de l'esprit, Newton & De Villiers, 2007). Il participe donc ainsi à l'élaboration du soi par contraste avec la constitution de représentations d'autrui. Il est également établi que le langage intérieur est au centre de la mémoire autobiographique (Morin & Hamper, 2012), il interagit avec la mémoire de travail afin d'améliorer l'encodage de nouvelles informations (Baddeley & Hitch, 1974 ; Marvel & Desmond, 2012). Il permet le voyage mental dans le temps, par l'évocation de souvenirs passés et la planification de situations futures (Meacham, 1979). Il contribue également à la métacognition (cf. I.1.3.1 ci-dessus), l'auto-évaluation, l'auto-régulation, l'auto-motivation (Hardy, 2006 ; Clowes, 2007), l'estime de soi, la connaissance de soi, de nos émotions et sensations ainsi que l'auto-description (Morin, 2018).

La pratique du langage intérieur se fait avec plusieurs degrés de métacognition (cf. I.2.2.1). Il peut être volontaire, et l'on peut utiliser le langage intérieur délibérément pour de courts énoncés, lorsque l'on compte dans sa tête, que l'on décide de faire mentalement une liste de courses ou que l'on répète mentalement un numéro de téléphone pour le composer immédiatement après. On peut même délibérément produire des énoncés plus longs, lorsque l'on prépare une conférence, ou que l'on imagine une conversation future. Mais, lorsque notre esprit vagabonde, le langage intérieur peut survenir de façon involontaire, spontanée, presque passive (cf. Bonald, cité par Egger, 1881, ou Saint-Paul, 1892). Il paraît alors évanescent, insaisissable, c'est ce que nous avons nommé le "vagabondage mental verbal" (Perrone-Bertolotti *et al.*, 2014). Cette forme plus passive a été entrevue par Platon qui faisait une distinction entre opinion et imagination :

« le dialogue intérieur de l'âme avec elle-même et sans la voix s'appelle pensée [...]; quand cela se fait en silence dans l'âme par la pensée, il faut l'appeler opinion et imagination quand cet état de l'âme n'est pas l'ouvrage de la pensée, mais de la sensation » (Platon, Sophiste, 263e-264a)

Le vagabondage mental (VM) ou *mind wandering* en anglais, peut se faire de diverses façons et peut contenir de l'imagerie visuelle, en plus du langage intérieur. Smallwood & Schooler (2015) ont repéré divers apports bénéfiques du vagabondage mental (voir aussi Mooneyham & Schooler, 2013 ; Schooler *et al.*, 2014 ; Smallwood & Andrews-Hanna 2013). Selon eux, le VM permet tout d'abord la prospection. Les pensées qui surviennent durant le VM sont souvent focalisées sur le futur, engageant le sujet dans la prospection et la planification de l'avenir et permettant ainsi d'améliorer la vie quotidienne. Un deuxième apport bénéfique est la créativité. Le VM semble lié à la capacité de générer des pensées novatrices, créatives. Un troisième bénéfice du VM est de conférer un sens à l'expérience personnelle. Le VM incite en effet au voyage mental dans le temps et nous permet d'intégrer nos expériences passées ou anticipées dans un récit de vie cohérent et porteur de sens. D'autres apports positifs du VM évoqués par Smallwood & Schooler incluent la pause mentale, soulageant l'ennui procuré par des activités monotones, et la rêverie diurne, qui permet de se préparer à des obstacles ou menaces à venir. Selon moi, parmi les diverses formes de VM, c'est le langage intérieur non-délibéré qui est le plus susceptible de remplir ces trois premiers rôles : prospection, créativité et sens donné à l'expérience personnelle. La pratique du langage intérieur, délibérée ou vagabonde, est donc essentielle dans la conscience de soi et le sentiment de cohésion du soi.

Encadré I.14. La spécificité humaine : capacité délibérée de simulation de la parole

La pratique (délibérée ou non) du langage intérieur – en nous permettant de placer le soi comme objet de réflexion, d'enrichir notre soi narratif, de nous projeter dans le passé et l'avenir, de créer de nouvelles pensées – renforce le sentiment de soi, dans le temps. C'est parce que la production de langage à voix haute peut être inhibée et simulée intérieurement que les êtres humains peuvent passer une grande partie de leur temps à construire et développer mentalement leur ego, sans embarrasser la terre entière.

La plus importante spécificité humaine ne réside ni dans la capacité de communiquer (certains animaux communiquent *a minima*), ni dans celle de penser (certaines personnes privées de langage sont capables de raisonnement conceptuel), mais dans celle de simuler mentalement l'acte de parole, délibérément.

I.1.3.2.7. La vertu thérapeutique du langage et du langage intérieur

Le rôle du langage dans la construction de soi m'amène à évoquer les vertus bénéfiques de la pratique du langage et en particulier du langage intérieur.

L'apport bienfaisant du langage a été repéré de façon universelle et ancestrale : de la pratique de la poésie (Meillier, 1982) à la récitation de prières ou de mantras (Shah, 1983), les humains utilisent le langage comme un remède pour calmer l'agitation mentale, atténuer les douleurs ou les souffrances psychiques et même physiques.

La vertu thérapeutique de la parole pour soigner les troubles psychologiques a été utilisée et théorisée par Freud qui a développé le concept de « *talking cure* ». Dans la cure psychanalytique telle qu'il l'a conçue, le récit, l'expression des troubles, des symptômes permet le soin, provoque la délivrance. Le patient, en formulant ses maux, en prend conscience et peut s'en libérer. Selon Freud & Beuer (1895/1956), la narration est « dépuratoire », l'énonciation a un pouvoir curatif. Si dans la cure psychanalytique, la présence de l'auditeur-riche est nécessaire, pour instaurer la narration, la vertu

thérapeutique du langage peut s'exercer aussi sans la présence d'autrui, dans la pratique du discours privé (intérieur ou à voix haute). Comme mentionné ci-dessus, le discours privé, le langage adressé à soi-même, a un rôle bénéfique dans la construction et le renforcement de la conscience de soi et il a été mentionné par Vygotski (1934/1985) qu'il contribue à l'auto-régulation, chez l'adulte bien sûr, mais également chez le jeune enfant. La pratique dialogale du langage intérieur peut apporter un nouvel éclairage sur soi (Bergounioux, 2004 ; Lacan, 1966 ; Philippe, 2001 ; Rosenthal, 2012), elle augmente notre estime de soi et nous permet de nous encourager (Morin, 2009, 2018), de nous reconforter, de nous consoler (Pavlenko, 2014).

Dans le cadre d'une collaboration avec Stéphanie Smadja, porteuse du projet *Monologuer* (<https://cerilac.univ-paris-diderot.fr/monologuer>) j'ai eu l'occasion de mener des paires d'entretiens phénoménologiques auprès de participants de tous âges et tous horizons, entrecoupés par la pratique par ces participants de l'écoute de leur propre parole intérieure, via la tenue d'un carnet, sur une période de quelques semaines (cf. II.2.1.4). Au deuxième entretien, la plupart des personnes ayant participé à cette étude souligne le caractère bénéfique et plaisant de cette pratique introspective régulière. Certains précisent même qu'elle a été pour eux thérapeutique. Cette vertu exutoire de l'écriture pour soi est celle que l'on retrouve dans la tenue d'un journal intime, dans le récit de soi, dont on devine en l'écrivant qu'il pourra être lu d'autrui, à notre insu, après notre mort, et même qu'il pourra prendre, le statut d'autobiographie ou de mémoires, publiés à titre posthume, qu'on l'ait désiré ou non (on pense au journal intime posthume de Maine de Biran, aux mémoires d'outre-tombe de Chateaubriand, aux cahiers de Simone de Beauvoir, au journal d'Anne Frank ou... aux carnets de Marc Masson, cités plus haut).

J'ai également eu la chance de participer à un projet art-science du groupe n+1 de la compagnie *Les ateliers du spectacle*, sur le thème des recherches scientifiques sur le langage et le cerveau. Ce projet a, entre autres, conduit à la création d'un « impromptu scientifique » sur le thème de la parole intérieure. Ce spectacle art-science, intitulé « des voix dans la tête » (<http://www.ateliers-du-spectacle.org/spectacle/les-impromptus-scientifiques/#voix>) invite l'audience à écouter son propre langage intérieur par le biais d'une mise en scène participative. Lors des échanges qui ont suivi les diverses représentations de cet impromptu, il a été régulièrement affirmé par le public que cet exercice de scrutation de la parole intérieure est intéressant, enrichissant et bénéfique.

Il me semble donc que la pratique de l'énonciation, à destination d'un·e thérapeute ou de soi-même, à voix haute ou silencieuse, écrite ou non, peut constituer une piste de recherche clinique intéressante, à développer.

Mais je n'oublie pas que le langage intérieur, qu'on le couche dans un carnet ou pas, peut parfois jouer un rôle néfaste. Derrida (1972) a montré la duplicité de l'écriture, qui peut être thérapeutique ou nocive. Il cite Platon qui expliquait que l'écriture empêche la connaissance directe de soi :

« Socrate compare à une drogue (Pharmakon) les textes écrits que Phèdre a apportés avec lui. Ce Pharmakon, cette "médecine", ce philtre, à la fois remède et poison, s'introduit déjà dans le corps du discours avec toute son ambivalence. Ce charme, cette vertu de fascination, cette puissance d'envoûtement peuvent être – tour à tour ou simultanément – bénéfiques et maléfiques. » Derrida (1972, p. 87).

Le langage intérieur peut en effet s'emballer, devenir négatif et excessif, comme dans le cas de la rumination mentale (Nalborczyk, *et al.*, 2017). Il peut aussi devenir délétère, comme dans le cas des hallucinations auditives verbales, une forme de parole intérieure non délibérée et intrusive, attribuée à une source externe (Rapin *et al.*, 2013, cf. I.2.3). Mais même dans ce cas, il semble que l'exercice délibéré du langage intérieur puisse être bénéfique. Charles Fernyhough – qui a mené des entretiens avec de nombreuses personnes vivant des hallucinations verbales, les « entendeurs de voix » (dont beaucoup ne souffrent pas de troubles psychiatriques) – rapporte dans son ouvrage « *The Voices Within* » que, chez certains, la pratique délibérée du langage intérieur permet de dialoguer avec les voix irruptives, de les pacifier, les rendre moins perturbantes. Là encore s'annoncent des perspectives de recherche avec des applications cliniques : comment réguler la parole intérieure dysfonctionnante, comment gérer ces voix produites de façon non délibérée, les rendre moins envahissantes et invalidantes ?

Encadré I.15. Langage et auto-noëse

On peut ainsi considérer que le langage est à l'œuvre, à côté d'autres dispositifs cognitifs non verbaux tels que la monstration et le pointage, lors des premières étapes de la construction du soi et de l'individuation soi/autrui chez le nourrisson et l'enfant.

Intériorisé, il joue ensuite un rôle essentiel, dans la construction d'un soi narratif plus élaboré et dans la maintenance de la perception (illusoire) de la permanence et de l'intégrité du soi. Il contribue à l'auto-régulation, renforce la connaissance et l'estime de soi, nous permet de nous encourager, nous reconforter, nous consoler, nous soigner.

Ce que l'on peut résumer en : « je parle, donc je communique, je pense, je suis ».

I.1.4. La triade Communication-Pensée-Auto-noëse

Ainsi à la question en préambule, sur l'origine de la diversité des formes sonores, gestuelles et écrites des langues du monde et les causes de la complexité de la machinerie cérébrale mise en œuvre pour émettre des sons et des gestes langagiers, je réponds que les langues subissent des pressions évolutives, liées à trois fonctions essentielles à l'être humain : la communication, la pensée et l'auto-noëse.

La fonction de communication avec autrui requiert d'être capable d'une pensée élaborée adaptée à des situations et environnements complexes d'échanges avec autrui. Elle nécessite également la conscience de soi pour se concevoir comme l'agent d'actions et de pensées à échanger, ainsi que la conscience d'autrui, pour reconnaître en nos congénères des partenaires de communication, avec des états mentaux et des intentions propres. C'est ainsi que nous pouvons développer des comportements verbaux culturellement et socialement adaptés (cf. Frith, 2010). Elle requiert enfin de savoir distinguer une situation communicationnelle réelle d'une situation imaginée, ce qui passe également par la conscience de soi, de son état. En effet, la fonction communicative nécessite de savoir si l'on est effectivement en train de parler et écouter ou si l'on est dans une phase d'imagination, en train de se remémorer une conversation ou de planifier un dialogue à venir (Gregory, cité par Frith, 2010).

La fonction cognitive d'élaboration de la pensée s'appuie, elle, sur la maîtrise des fondamentaux de la communication pour s'auto-questionner et nourrir nos dialectiques

personnelles. Comme le défend Fernyhough (2016), c'est par le dialogue (à la Platon !) internalisé, par des conversations imaginaires, que l'être humain peut envisager plusieurs hypothèses, considérer plusieurs points de vue et s'auto-réguler. Il n'est pas nécessaire de postuler l'existence d'un chef d'orchestre intérieur, un homoncule dirigeant notre pensée, celle-ci progresse naturellement au cours de nos dialogues intérieurs. La fonction d'élaboration de la pensée est donc dépendante de la fonction de communication. Elle met également en jeu la mémoire de soi et de ses raisonnements passés, la mémoire épisodique et la conscience auto-noétique – qui permettent de construire, affiner progressivement nos réflexions. Elle implique aussi la méta-conscience de soi qui permet de nous refocaliser sur une tâche en cours, interrompue lorsque notre esprit se met à vagabonder (Schooler *et al.*, 2011), ce qui arrive lorsque l'on lit un texte long et ennuyeux (et ce qui est donc, hélas, peut-être le cas des lecteur·rices de ce texte, en ce moment...).

Enfin, *la fonction d'auto-noèse* nécessite de reconnaître un autre, de communiquer avec autrui pour s'identifier soi comme singulier. Elle s'appuie aussi sur la capacité de mener une pensée structurée et élaborée, nous permettant de nous remémorer des épisodes passés et d'anticiper des événements futurs avec précision et élaboration, ces événements passés et futurs constituant le fondement de notre auto-noèse.

Encadré I.16. La triade systémique qui a mené à la complexité et la diversité du langage

Ainsi, il apparaît que le langage remplit trois fonctions majeures : la fonction communicative, qui nous permet d'échanger avec nos congénères, la fonction cognitive, qui nous permet d'élaborer et de structurer notre pensée, notre raisonnement et enfin la fonction métacognitive d'auto-noèse qui contribue à l'ipséité, l'élaboration du tandem soi/autrui et la conscience de soi dans le temps.

Ces trois fonctions sont intriquées, de façon systémique, chacune s'appuyant sur les deux autres, avec des interactions dynamiques complexes.

Ce sont, selon moi, les pressions conjuguées de la complexification de la communication et du développement de la pensée, associée à une troisième pression, celle du renforcement de la conscience de soi, qui ont conduit au langage humain actuel, dans toute sa complexité, sa diversité, sa richesse.

Chaque fonction est toutefois susceptible de perturbation, mettant en danger la triade.

La fonction communicative peut à tout moment dévier, à cause d'erreurs de syntaxe, de lexique, nous empêchant d'atteindre notre objectif initial d'échange d'idées, de sentiments, d'émotions avec autrui. Il peut nous arriver de ne pas nous faire comprendre, ni même de ne pas comprendre autrui. Parfois, dans le cas de l'ironie ou de la métaphore par exemple, nous jouons intentionnellement à rendre notre message difficilement compréhensible pour autrui (en tout cas non transparent), tout en espérant que notre interlocuteur·rice saura finalement le décoder.

La fonction cognitive peut elle-même dérailler, nous amenant à utiliser des structures linguistiques inappropriées et à faire des déductions illogiques, des raisonnements fallacieux, à ne pas nous comprendre nous-mêmes.

Enfin, la fonction auto-noétique peut achopper, nous conduisant à attribuer à autrui notre propre pensée verbale. Nous pouvons alors avoir la sensation de ne pas être l'auteur·e de nos pensées ou de nos actes, de ne pas nous reconnaître nous-mêmes.

Pour que ces trois fonctions s'exécutent normalement, une cybernétique verbale est nécessaire, avec plusieurs composantes. Il faut, d'une part, des dispositifs de *monitoring* (i.e. de contrôle, d'ajustement, de correction) des messages verbaux et, d'autre part, des mécanismes d'attribution de l'agentivité, que j'aborde maintenant.

I.2. UNE CYBERNETIQUE ADAPTEE A CES FONCTIONS

Comme l'exprime Martinet (1970) :

« quel que soit l'emploi que nous fassions du langage, qu'il nous serve à ordonner ou clarifier notre pensée, ou que nous parlions pour nous exprimer au sens propre du terme, nous nous comporterons toujours comme s'il fallait nous faire comprendre d'autrui ».

Cette motivation essentielle, celle de rendre notre message compréhensible par autrui, mais aussi par nous-mêmes, requiert une cybernétique verbale, un mécanisme général de prédiction, de régulation et de contrôle, qui se décline en plusieurs composantes – *monitorings*¹⁴ verbaux et attribution de l'agentivité – impliquées de façon différenciée dans chacune des trois fonctions du langage. En effet, pour que les fonctions communicative, cognitive et auto-étiologique du langage puissent se dérouler de façon efficace et adaptée, il faut que les erreurs possiblement commises par les systèmes de production et perception du langage soient détectées et éventuellement corrigées, d'une part, et que ces systèmes soient en mesure de distinguer les événements auto-générés de ceux produits par autrui, d'autre part (cf. e.g. Houde *et al.*, 2002 ; Tian & Poeppel, 2015 ; Tourville *et al.*, 2008). Je détaille à présent les dispositifs de *monitoring* verbal et d'attribution de l'agentivité pour chacune des trois fonctions essentielles du langage.

I.2.1. Le *monitoring* verbal dans la communication

La fonction de communication nécessite des processus de *monitoring* des flux verbaux produits et reçus. En m'appuyant sur des travaux antérieurs en psycholinguistique et neurosciences, je propose de modéliser comment le *monitoring* verbal se déploie au cours de la communication. Comme énoncé dans Schwartz & Lœvenbruck (2012) :

« 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. »

¹⁴ Le terme anglais de *monitoring* est utilisé ici à dessein, car il recouvre les notions de contrôle, surveillance, supervision, observation, régulation, guidage, ce qu'aucun terme français ne permet.

Les modélisations psycholinguistiques classiques du processus de communication contiennent trois niveaux symétriques de production et de réception de message (Caron, 1989) : un niveau supérieur consistant en la représentation conceptuelle, incluant l'intention communicative du·de la locuteur·rice, un niveau médian consistant en la représentation lexico-syntaxique et un dernier niveau, consistant en des représentations articulatoires et acoustiques. Dans cette vision classique de la communication, illustrée sur la Figure I-6, le contenu de pensée est encodé sous forme lexico-syntaxique puis phonologique, donnant ainsi un signal de parole, transmis à l'allocutaire. Celui-ci décode le signal de parole reçu pour reconstruire un contenu de pensée aussi proche que possible du concept initial. Les trois niveaux s'enchaînent donc de façon descendante (*top-down*) avec des processus d'encodage pour la production du langage et de façon ascendante (*bottom-up*) avec des processus de décodage pour la réception.

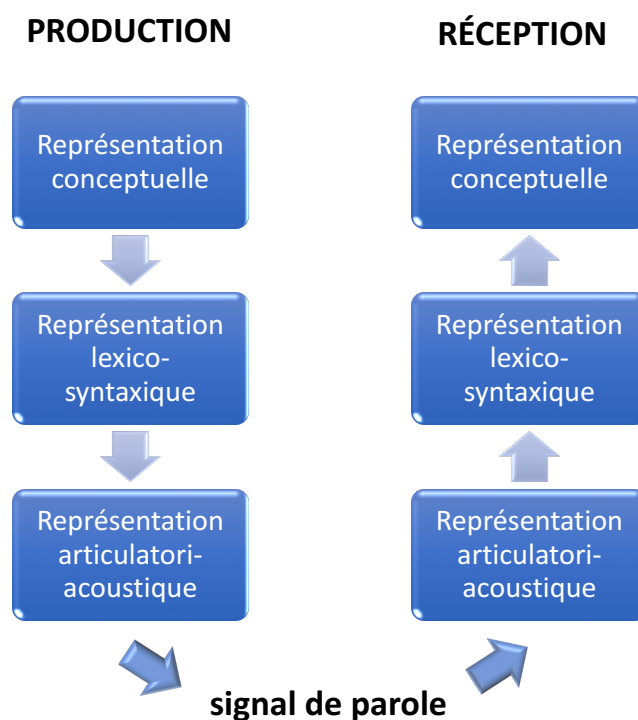


Figure I-6. Schéma classique de la communication linguistique avec la symétrie locuteur·rice / allocutaire des trois niveaux fondamentaux. La production y est naïvement conçue comme impliquant des processus descendants et la perception, des processus ascendants.

Cette vision est réductrice, car elle ne tient pas compte de la variabilité interne du·de la locuteur·rice, ni de l'auditeur·rice, ni de celles du contexte et de la situation dans lesquels s'inscrit la communication. Elle simplifie aussi les interactions entre les niveaux. Elle a donc été débattue et de nombreux modèles psycholinguistiques différents ont été proposés (cf. François & Nespoulous, 2011). En particulier, en ce qui concerne les interactions entre niveaux, les observations expérimentales sur l'autocorrection immédiate lors de la production du langage, ont conduit les psycholinguistes à introduire des processus ascendants, constituant des boucles de rétroaction d'un niveau à l'autre. De façon symétrique, lors de la perception, les connaissances et les *a priori* de

l'auditeur·rice introduisent des processus descendants des niveaux supérieurs vers les niveaux inférieurs.

En effet, lorsque, en tant que locuteurs, nous produisons du langage, nous disposons de mécanismes de *monitoring* de notre production verbale, nous permettant de nous corriger en cas d'erreur (cf. *self-editing*, Skinner, 1957). Cette capacité fondamentale, dénommée *speech monitoring* ou *verbal self-monitoring* (e.g. Levelt, 1983 ; Huettig & Hartsuiker, 2010 ; Postma, 2000 ; Riès *et al.*, 2011) et que l'on peut traduire par auto-contrôle verbal, nous permet de suivre et réguler nos propres productions langagières, de les vérifier, de détecter des erreurs par rapport à l'objectif initial et éventuellement les corriger immédiatement, en ligne, parfois avant même qu'elles soient complètement réalisées ou juste après. Elle nous permet également de tenir compte des signaux de l'environnement, du bruit ambiant par exemple, et d'adapter notre énonciation afin qu'autrui puisse nous comprendre.

Sur le versant perception, lorsqu'en tant qu'auditeur·rice, nous percevons le langage d'autrui, nous mettons également en œuvre des mécanismes de *monitoring* des productions verbales d'autrui, à partir de nos connaissances sur le monde, sur nos interlocuteurs et sur le langage, afin de suivre et comprendre nos interlocuteurs. Ces mécanismes nous permettent également de tenir compte des signaux de l'environnement et d'ajuster notre interprétation du message d'autrui en conséquence.

Je représente sur la Figure I-7 ces mécanismes de *monitoring* des productions verbales de soi-même et d'autrui, ou *monitorings* verbaux égocentré et allocentré, qui tiennent compte des signaux de l'environnement et qui introduisent des boucles de rétroaction dans la production et la réception du langage. Cette architecture parallèle peut être rapprochée de celle proposée et défendue par Jackendoff (2002, p. 199, Figure 7.1).

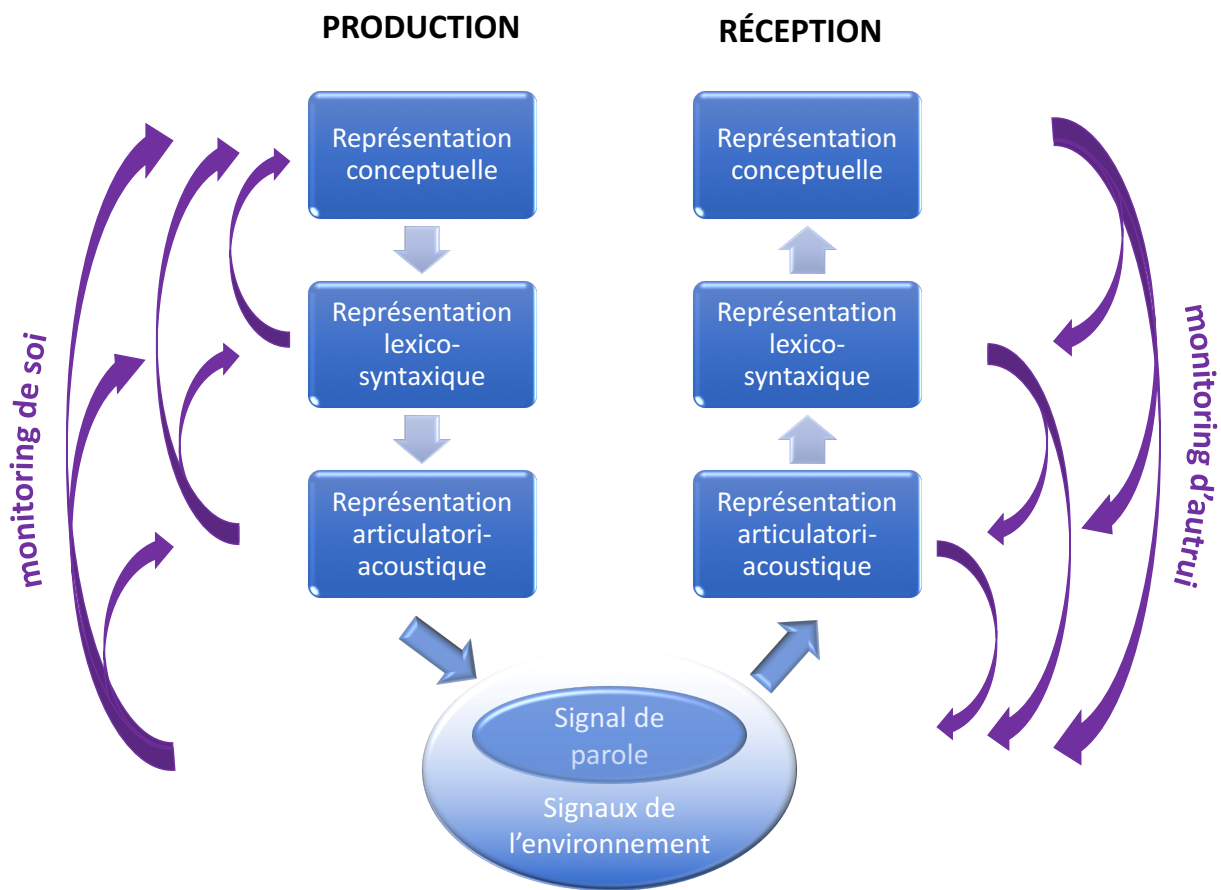


Figure I-7. Boucles de rétroaction induites par le monitoring verbal de soi et d'autrui

Je détaille par la suite comment se met en place le *monitoring* verbal dans chacun des versants de la fonction de communication du langage. Pour faciliter la description, je sépare production et perception du langage, même s'il est évident que les deux sont indissociables. Dans ce cadre, deux modèles bayésiens complémentaires me paraissent particulièrement pertinents. Le modèle COSMO développé par Schwartz et collègues, permet de rendre compte de l'implication des processus de production de la parole dans la perception (cf. Moulin-Frier, Diard, Schwartz & et Bessière, 2015). Et de façon réciproque, la modélisation couplée des processus de production et perception, intégrant des mécanismes de *monitoring* de la production proposée par Patri, Perrier et Diard (2018) est tout particulièrement intéressante. Ces deux modèles s'appliquent pour l'instant au niveau de la phonologie, de la production et de la perception des phonèmes, mais ils me paraissent extrêmement prometteurs. On peut d'ailleurs imaginer qu'ils puissent être intégrés à des modèles plus complets de la production et de la perception de la parole. Ces deux modèles font partie des premières tentatives d'intégrer perception et production dans une même architecture et sont donc très inspirants. La théorie intégrative de Pickering & Garrod (2013) offre, elle aussi, un cadre permettant d'expliquer l'imbrication des processus de production et perception du langage. Comme nous le verrons par la suite, elle est moins détaillée dans l'implémentation que les deux modèles précédents, mais elle prend en compte les niveaux syntaxiques et sémantiques en plus du niveau phonologique et est donc, à ce titre, particulièrement éclairante.

I.2.1.1. La production du langage et le *monitoring verbal* égo-centré

Comme introduit ci-dessus, la production du langage engage le *monitoring verbal égo-centré*, c'est-à-dire un mécanisme permettant de suivre, réguler, ajuster nos propres productions langagières, en tenant compte des signaux de l'environnement, et en adaptant notre énonciation pour permettre à autrui de nous comprendre. Avant de détailler ce dispositif de *monitoring verbal*, rappelons quels sont les différents niveaux de traitements impliqués, le *monitoring* pouvant en effet s'appliquer à divers niveaux.

I.2.1.1.1. Les différents niveaux de traitement impliqués dans la production du langage

La production langagière (orale ou signée), telle qu'elle est classiquement décrite dans la plupart des modèles psycholinguistiques ou neurolinguistiques (notamment dans le modèle de la production de phrases avec niveaux de traitement interactifs de Bock, 1987 ; la théorie des activations réciproques inter-niveaux (ou le modèle connexionniste de la production de la parole) de Dell, 1986 ; le modèle sériel à 5 niveaux de Fromkin, 1971 ; le modèle standard sériel à niveaux de traitements indépendants de Garret, 1980 ; la Grammaire Procédurale Incrémentale de Kempen & Hoenkamp, 1987 ; le modèle neurolinguistique de Laver, 1980 ; le modèle de Levelt, 1989 ; le modèle WEAVER ++ de Levelt, Roelofs & Meyer, 1999), inclut trois composantes principales, plus ou moins autonomes : la conceptualisation (ou idéation), la formulation, qui inclut les traitements lexico-syntaxique et morpho-phonologique, et l'articulation, qui inclut la planification et l'exécution des mouvements articulatoires. Les modèles diffèrent sur la façon dont ces trois composantes et leurs sous-composantes s'articulent, ainsi que sur leur organisation et leur interdépendance, mais ils s'accordent sur leur existence (cf. e.g. Dell, 2013 pour une revue). L'organisation peut ainsi être modulaire ou non, hiérarchique ou non, en cascade « de haut en bas » ou interactive (les informations d'un niveau inférieur de traitement pouvant influencer en retour les niveaux supérieurs).

Les modèles les plus repris dans la littérature décrivent le langage oral, mais peuvent s'appliquer au langage signé. En particulier, la notion de morphème peut être remplacée par celle de kinème, celle de phonème peut être remplacée par celle de chérème ou gestème, celle d'articulation orale par celle d'articulation manuo-facio-corporelle et celle de son par celle de geste (Stokoe, 1960 ; Nève, 1996 ; Kervajan, 2006). Pour simplifier, je restreins la description aux langues orales, d'une part, et je reprends les principaux éléments du modèle de Levelt et collègues (dans la lignée de celui de Garret, 1980), d'autre part. Ces éléments incluent : conceptualisation, encodages grammatical et phonologique (parfois appelés formulation) et articulation (dont la planification articulatoire). Dans le modèle de Levelt, ces éléments sont organisés de façon modulaire, en cascade d'un niveau hiérarchique à l'autre, alors que dans d'autres modèles (notamment celui de Dell et collègues, cf. plus bas), des interactions bidirectionnelles existent d'un niveau à l'autre. Malgré des conceptions différentes des interactions inter-niveaux, les modèles s'accordent toutefois sur les traitements linguistiques effectués à chaque niveau. Je décris donc ici ces différents traitements, tels qu'ils sont conçus généralement.

Au cours de la conceptualisation, le-la locuteur·rice élabore ce qu'il-elle veut dire, à partir de son intention communicative, et construit un message pré-verbal (qui peut tenir compte de la situation contextuelle et inclure des informations pragmatiques). Dans l'étape de formulation, le message pré-verbal est traduit dans une forme linguistique

propre à être énoncée. Cette étape est elle-même divisée en deux opérations distinctes, l'encodage grammatical et l'encodage phonologique. Le premier encodage consiste à sélectionner dans le lexique les lemmes¹⁵ correspondant aux éléments conceptuels du message préverbal et à les organiser dans un ordre syntaxique approprié. Le deuxième consiste à produire, à partir de la structure de surface fournie par l'encodage grammatical, les structures morphologiques des mots de l'énoncé puis la structure phonologique de l'ensemble, intégrant la structure phonémique (dans le cas des langues orales : les consonnes, les voyelles, les groupes consonantiques, etc.) des mots (autrement dit les lexèmes) et la structure syllabique, métrique et intonative (avec éventuelles resyllabifications trans-frontières lexicales, comme dans « ce modèle est original » -> sə.mo.dɛ.le.to.bi.zi.nal). Le découpage en deux opérations d'encodage distincts est justifié par des données expérimentales, notamment la typologie des erreurs de production (e.g. Fromkin, 1971 ; Nootboom, 1980), les études de chronométrie de tâches d'interférence lors de la dénomination (Levelt, 1989), le phénomène du mot sur le bout de la langue (Brown & McNeill, 1966) ou la multiplicité des troubles de la production de parole chez les patients cérébro-lésés (Dell, Schwartz, Martin, Saffran, Gagnon, 1997).

Enfin, un objectif phonétique/articulatoire est obtenu, à partir de la représentation phonologique (incluant la prosodie) ainsi construite, en récupérant les gestes articulatoires dans un syllabaire mental. Ces gestes, spécifiés sous formes d'objectifs articulatoires (e.g., dans le cas des langues orales, friction alvéolaire, occlusion labiale) constituent l'objectif phonétique/articulatoire qui, selon Levelt et collègues, correspond au langage interne ou intérieur (la petite voix dans la tête). Cet objectif est fourni en entrée à l'appareil articulatoire (appareil vocal ou brachio-manuel) pour produire les mouvements articulatoires qui créeront des sons ou des gestes manuels. Dans le modèle de Levelt (1989), une mémoire-tampon articulatoire (*articulatory buffer*) est décrite, entre l'objectif phonétique/articulatoire et l'appareil articulatoire, pour rendre compte des instances où la formulation a lieu pendant que l'appareil articulatoire est encore occupé avec l'énoncé précédent. Les parties de l'objectif phonétique qui ne sont pas encore articulées sont ainsi maintenues actives dans cette mémoire tampon jusqu'à ce que l'appareil articulatoire soit prêt.

1.2.1.1.2. Le *monitoring verbal égocentré* : suivi des verbalisations auto-produites

Pour que le versant productif de la fonction de communication se déroule avec succès, c'est-à-dire pour que le message produit soit perçu et compris par nos interlocuteurs, un mécanisme de *monitoring verbal égocentré* est nécessaire, engageant des processus de perception. Plusieurs travaux sur la production de parole montrent en effet que la perception de soi-même y joue un rôle majeur (Perkell, 2012 ; Turgeon *et al.*, 2015) et que des dispositifs d'auto-perception et d'auto-évaluation sont en jeu. En particulier, les études sur les erreurs de production de parole et leurs corrections spontanées (cf. e.g. Fromkin, 1980) indiquent que les locuteurs sont capables de s'apercevoir qu'une erreur s'est produite ou est sur le point de se produire, et donc que des mécanismes de

¹⁵ Dans la terminologie de Levelt et collègues, le lemme est une entité abstraite, désincarnée de la forme articulatoire et sonore. Il contient des informations sémantiques et syntaxiques : le sens du mot, sa catégorie syntaxique (nom/ verbe/ adjectif) et des propriétés telles que le temps, la personne et le nombre pour les verbes ou le genre et le nombre pour les noms. Le lexème est déterminé à un stade ultérieur et est constitué de la structure morphologique et phonologique (segmentale) du mot. Le recouvrement exact des notions de lemme et lexème est débattu, nous simplifions ici, par souci de concision.

monitoring verbal sont effectivement en jeu. Il existe toutefois des erreurs non détectées et on estime que seule la moitié des erreurs verbales est corrigée par les locuteurs (Nooteboom, 1980). L'étude détaillée des erreurs et des processus de correction suggère que le *monitoring* de la production de langage oral peut s'effectuer à différents niveaux : sémantique, syntaxique, prosodique, lexical et phonologique. De nombreux chercheurs ont en effet recueilli divers types d'erreurs observées et leurs corrections (e.g. Blackmer et Mitton 1991 ; Fromkin, 1980 ; Cutler, 1983 ; Levelt, 1983 ; Oomen and Postma, 2001 ; Rossi, 2001 ; Stemberger, 1982). Les locuteurs peuvent par exemple contrôler que l'étape de conceptualisation s'est bien déroulée comme dans l'exemple suivant (tiré de Blackmer & Mitton, 1991) :

*Exemple 1 : « because [most]... a lot of people are »
(parce que [la plupart des]... beaucoup de gens sont)*

Ils peuvent également vérifier que les choix lexicaux sont appropriés, comme dans cet exemple (tiré de Levelt, 1983) :

*Exemple 2 : « Entrance to [yellow] eh to grey. »
(L'entrée dans le [jaune] euh le gris)*

Ils peuvent vérifier la structure syntaxique de leur énoncé, comme dans cet exemple (tiré de Stemberger, 1982) :

*Exemple 3 : « I hope the professor didn't notice what [was I] ... what I was doing.»
(J'espère que le professeur n'a pas remarqué [qu'est-ce que je] ... ce que je faisais)*

Et de façon cruciale, l'erreur peut être détectée avant même que l'énoncé soit entièrement produit, comme dans les deux exemples suivants (tirés de Levelt, 1983) :

*Exemple 4 : We can straight on to the [ye]... , to the orange node
(Nous pouvons aller directement jusqu'au nœud [jau]... orange)*

*Exemple 5 : ... is a [v] ... a horizontal line
(...est une ligne [v]... horizontale)*

Lorsque les erreurs sont détectées et corrigées, différentes méthodes de réparation peuvent être appliquées. Dans les réparations avec reprise (*retracing repairs*, cf. van Wijk & Kempen, 1987), les locuteurs interrompent l'énonciation, repartent en arrière et reprennent l'énonciation, avec une correction partielle ou totale, comme dans l'exemple 6, tiré de Levelt (1983) :

*Exemple 6 : And [over] the grey sphere a, or right of the grey sphere a purple sphere
(Et [au-dessus] de la sphère grise une, ou à droite de la sphère grise une sphère violette)*

Dans les réparations sans reprise (*nonretracing repairs*), la partie de l'énoncé à réparer est remplacée par la correction, sans retour en arrière (exemple 1).

Le *monitoring* verbal opère aussi bien à voix haute qu'en parole intérieure (cf. I.2.2.2). La façon dont il s'opère a été modélisée de diverses manières (voir e.g. Blackmer & Mitton, 1991 ; Epting & Critchfield, 2006 ; Hartsuiker & Kolk, 2001 ; Levelt, 1989 ; Postma, 2000). Dans certains modèles, les mécanismes de contrôle des erreurs incluent des critères de *production*, dans d'autres ce sont des critères de *perception* qui sont privilégiés. La régulation des erreurs peut s'appuyer sur des méthodes de correction des erreurs (*editing*) dans certains modèles mais reposer sur un simple renforcement des sorties correctes (*boosting*) dans d'autres. Enfin, les modèles diffèrent sur le degré

d'automaticité du *monitoring* verbal, avec détection et correction délibérée et conscience ou bien rapide et automatique. Ces différences donnent lieu à des descriptions variées du déroulement temporel de la détection et de la correction des erreurs. Je résume ici les stratégies de *monitoring* verbal implémentées dans quatre types de modèles, le modèle à boucle perceptive (*Perceptual Loop*) de Levelt et collègues (Levelt, 1989), le modèle neurolinguistique de contrôle de Laver (1980), le modèle de production de Kempen et collègues (Van Wijk & Kempen, 1987) et les modèles connexionnistes de Dell et collègues (Dell, 1986) ou de MacKay (1990).

Dans le modèle à boucle perceptive de Levelt (1989), trois dispositifs de *monitoring* sont décrits, recrutant un même centre de contrôle (nommé *monitor*), situé au niveau du module de conceptualisation. Tout d'abord, après l'étape de conceptualisation, le locuteur·rice vérifie que le message pré-verbal correspond bien à l'intention communicative initiale. Ensuite, lorsque la formulation est effectuée, le système utilisé pour compréhension de la parole d'autrui intervient, pour décoder l'objectif phonétique (le langage intérieur) et s'assurer qu'il correspond à l'intention communicative initiale. Il peut donc, selon Levelt, y avoir une correction de l'erreur avant même que l'énoncé soit articulé (cf. aussi Levelt, 1983, p. 52). Enfin, après l'articulation (la planification articulatoire puis les mouvements articulatoires qui en résultent), le son produit est lui-même décodé par le système de compréhension et comparé à l'intention communicative. Ainsi dans ce modèle, le *monitoring* est centralisé, avec un seul centre de contrôle, délibéré (conscient), situé au niveau du module de conceptualisation, permettant de réguler, sur la base de critères perceptifs, les erreurs dans le message pré-verbal, dans l'objectif phonétique/articulatoire ou dans la réalisation acoustico-articulatoire finale. Un des avantages de ce modèle est qu'il prévoit que l'objectif phonétique (le langage intérieur) soit inspecté, corrigé et recalculé avant même d'être réalisé, sans interrompre le processus de production en cours. Cette surveillance du produit langagier interne permet ainsi, lorsqu'une erreur est produite, qu'un nouvel objectif phonétique soit prêt très rapidement, une production corrigée pouvant ainsi être émise quasi-immédiatement, ce qui est conforme aux observations expérimentales sur la rapidité des processus de correction en ligne des erreurs (moins de 150ms entre l'erreur produite et sa correction). Elle permet d'autre part d'expliquer la détection et la correction des erreurs lors de la production de parole intérieure. Une objection majeure faite à ce modèle, cependant, est que le *monitoring* requiert l'implication du système de compréhension de la parole (MacKay, 1992a). Or des études de patients aphasiques, par exemple, semblent indiquer une dissociation entre les capacités de compréhension du langage et celles de *monitoring* des erreurs lors de la production (pour une revue, cf. Postma, 2000). Il existe des cas de *monitoring* déficient avec des capacités de compréhension préservées (cf. e.g. Nickels & Howard, 1995). Et, plus problématique, il existe des cas de compréhension altérée n'entraînant pas de problème de *monitoring* (Marshall, Rappaport and Garcia-Bunuel, 1985).

Dans le modèle neurolinguistique de contrôle de Laver (1980), le *monitoring* est au contraire distribué, avec de multiples centres de contrôle locaux autonomes, essentiellement automatiques, intervenant à chaque étape de la production de langage, avec des critères de production, et un centre de contrôle final délibéré (conscient), qui évalue la sortie sensorielle effective par rapport à l'intention initiale (avec des critères perceptifs). Dans ce modèle, la correction des erreurs est ainsi effectuée au niveau de l'étape à laquelle l'erreur a été détectée, interrompant la suite du processus de

production pour corriger le défaut. Ainsi, certaines des erreurs peuvent être corrigées avant même que l'énoncé soit effectivement articulé. Ce modèle prédit, de façon conforme aux observations, que les erreurs sont corrigées rapidement, puisqu'elles sont détectées et réparées au niveau où elles apparaissent. Toutefois, il prédit de façon incorrecte que les erreurs sont rares (puisque elles sont détectées à plusieurs niveaux) et que la production du message est longue puisqu'elle est soumise à des vérifications et interruptions en cas d'erreur.

Dans le modèle de production de la parole de Kempen et collègues (Van Wijk & Kempen, 1987), le mécanisme de *monitoring* est centralisé. Il est localisé dans un centre situé dans le système de production de la parole, mais externe aux modules de traitement. Ce moniteur central vérifie les sorties incrémentales de chacun des modules, sur des critères de production et interrompt la production lorsqu'une erreur a été détectée à un niveau. Toutefois, le modèle n'est pas clair sur les détails de l'implémentation de ce mécanisme de *monitoring*. En particulier, il n'est pas précisé si les vérifications des sorties à chaque étape se font de façon sérielle ou parallèle, ni quand le moniteur central interrompt la production (si la production est interrompue dès qu'une erreur est détectée, alors il ne devrait jamais y avoir d'erreur produite).

Enfin, dans les modèles connexionnistes – tels que le modèle à activation interactive de Dell et collègues, ou le modèle connexionniste *Node Structure Theory* de MacKay (1990) – le mécanisme de *monitoring* se fonde sur le renforcement (sans correction explicite des erreurs). La combinaison de l'activation et de l'inhibition des unités (nœuds) des différents niveaux induit un biais vers la sortie correcte, ce qui minimise les erreurs de production. Le *monitoring* dans ce type de modèle met en œuvre des critères de production et peut être automatique (dans le modèle de Dell et collègues) ou avec prise de conscience de l'erreur et correction délibérée (dans celui de MacKay). Toutefois, si ce type de modèle explique bien comment les erreurs sont évitées, la façon dont celles-ci sont réparées n'est pas détaillée.

Un panorama des différentes implémentations des dispositifs de *monitoring* proposées dans la littérature est proposé par Postma (2000). Il en ressort qu'un modèle de production adéquat doit rendre compte des cinq contraintes suivantes :

- (1) les erreurs de production sont corrigées rapidement, et peuvent même être détectées et réparées avant même la fin de l'énonciation (cf. exemples 4 et 5) ;
- (2) la production de la parole est en général fluide et rapide ;
- (3) les erreurs sont rares ;
- (4) la production de la parole peut se faire avec un système de compréhension déficient ;
- (5) la réparation des erreurs peut se faire en reprenant une partie de l'énoncé (*retracing*) ou en poursuivant directement avec la correction, après une halte (*nonretracing*).

Comme nous l'avons noté ci-dessus, le modèle de Levelt ne permet pas de rendre compte de la contrainte (4), puisque le système de *monitoring* s'appuie sur le système de compréhension de la parole, qui transforme les unités à évaluer (la forme sonore externe ou interne de la parole) en unités pré-verbales pouvant être comparées, par un module dédié (non spécifié d'ailleurs), à l'intention communicative initiale. Les modèles de Laver

et Kempen, avec un *monitoring* à multiple niveaux vérifiant chaque étape de la production, ne permettent pas de rendre compte de la fluidité de la parole (contrainte 2). Les modèles connexionnistes n'expliquent pas comment les haltes et les reprises d'énoncé peuvent se mettre en place (contrainte 5).

Monitoring par contrôle prédictif de l'étape d'articulation

Il a été proposé que des mécanismes de prédiction puissent mettre en œuvre les dispositifs de *monitoring* verbal de façon efficace, en répondant aux quatre contraintes ci-dessus, c'est-à-dire avec rapidité, sans perturber le décours de la production de parole, sans nécessiter l'intervention du processus de compréhension entier et en autorisant les haltes et les reprises (Hickok, 2012 ; Pickering & Garrod, 2013 ; Vicente & Martínez-Manrique, 2016). Les mécanismes de prédiction invoqués s'appuient sur la notion de copie d'efférence – décrite en détail dans Rapin *et al.* (2013, 2016) et Lœvenbruck *et al.* (2018) – ou sur un signal inhibiteur équivalent, dans le cas du modèle de Hickok (2012). Ces mécanismes ont été introduits dans le cadre de la robotique et des théories sur le contrôle moteur adaptatif de l'action, sous le terme de « prédicteur de Smith » (e.g. Miall, Weir, Wolpert, & Stein, 1993 ; Miall & Wolpert, 1996 ; Wolpert & Kawato, 1998). Tout comme le cas particulier de la parole, le contrôle de l'action en général nécessite un *monitoring*, avec détection et correction des erreurs, et donc doit inclure un mécanisme de rétroaction (ou *feedback*, cf. Wiener, 1948) qui permet de réduire l'écart entre l'action effectivement produite et celle désirée. Mais le contrôle uniquement par rétroaction pose un problème de fluidité de l'action : le délai entre la détection de l'erreur par l'appareil sensoriel et la réponse motrice à fournir pour la corriger engendre des instabilités incompatibles avec un système de contrôle dynamique. L'introduction de mécanismes de prédiction a constitué une avancée significative dans les théories du contrôle moteur chez l'humain. Ces mécanismes font un compromis entre le contrôle de l'action de type *feedforward* – fondé uniquement sur des objectifs correspondants à des états à atteindre, contrôle stable, mais ne permettant pas de *monitoring* – et le contrôle de l'action de type *feedback* – fondé sur des signaux de rétroaction-correction (*control*) mais engendrant des instabilités liées aux délais de détection des erreurs. Les mécanismes prédictifs remédient aux problèmes de délais dans les boucles de *feedback*, par l'introduction d'une prédiction des états futurs, qui permet une correction des commandes motrices avant même que l'état final soit atteint. En effet, un système prédicteur inclut un simulateur, un modèle interne de l'appareil moteur, qui prédit un signal de *feedback* (rétroaction) interne rapide qui permet d'ajuster les commandes motrices avant que l'action soit effectuée. Plus précisément, tel qu'illustré sur la Figure I-8, le *monitoring* de l'action par contrôle prédictif (MCP) met en œuvre deux modèles internes : un modèle direct et un modèle inverse. Le « modèle interne direct » (ou prédicteur, simulateur, émulateur, modèle prédictif) est une représentation interne (une simulation) de l'appareil moteur qui permet de prédire les conséquences de l'émission d'une commande motrice. A partir de l'état en cours de l'appareil moteur et de commandes motrices qui lui seraient appliquées, le modèle direct fournit une estimation du nouvel état de l'appareil moteur, sous la forme de sensations prédites (Jordan & Rumelhart, 1992 ; Miall & Wolpert, 1996). La production d'actions requiert en amont un deuxième modèle interne, le « modèle interne inverse » (ou contrôleur) qui effectue l'opération inverse du modèle direct, c'est-à-dire que ce modèle calcule les commandes motrices à appliquer pour atteindre l'état désiré de l'appareil moteur. Ainsi, lors de la production d'une action, des commandes motrices sont générées par le modèle inverse

pour être envoyées à l'appareil moteur. Pour corriger les erreurs le plus rapidement possible, un duplicata de ces commandes motrices, nommé copie d'efférence, est fourni au modèle interne direct, qui génère des signaux sensoriels simulés. Ces signaux simulés, sont nommés « prédits » car le simulateur, ne mettant pas en œuvre toute la complexité effective de la musculature (avec des délais de transmission synaptique et les inerties musculaires et articulaires), est capable de fournir des signaux de façon quasi immédiate, avant les signaux effectifs, en sortie de l'appareil moteur lui-même. Les signaux simulés (également appelés « décharge corollaire ») sont donc générés avant que les mouvements aient effectivement lieu. Ils constituent donc une prédiction de l'état futur de l'appareil moteur. Une comparaison peut donc se faire entre l'état sensoriel prédit et l'état désiré, avant que les commandes motrices soient effectivement appliquées. En cas de non-conformité, les commandes motrices sont ajustées. Le mécanisme de prédiction par copie d'efférence permet d'expliquer comment, en cas d'erreurs ou d'inexactitude dans la génération des commandes motrices initiales (par le contrôleur), celles-ci peuvent être corrigées avant même que le mouvement ait lieu, ce qui rend la production d'actions efficace, stable et fluide.

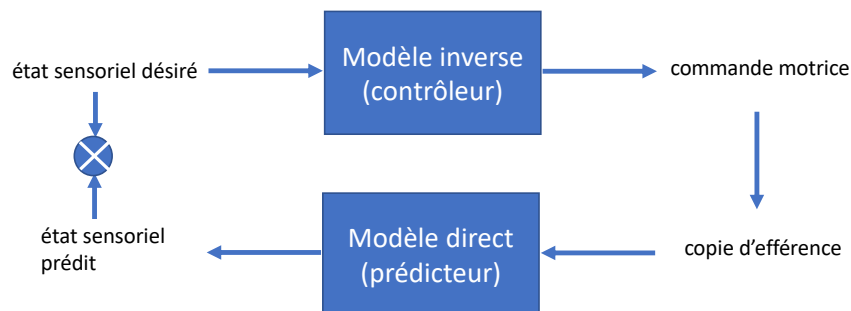


Figure I-8. Le monitoring des actions par contrôle prédictif, avec modèles internes inverse (contrôleur) et direct (prédicteur) et copie d'efférence

Le *monitoring* par contrôle prédictif (MCP) a été appliqué au cas particulier des actions de parole, à partir d'objectifs phonétiques (l'étape d'articulation des modèles psycholinguistiques), par certaines équipes (notamment Guenther *et al.*, 1998 ; Guenther *et al.*, 2006 ; Hickok, Houde, Rong, 2011 ; Hickok, 2012 ; Houde & Nagarajan, 2011, voir aussi la modélisation bayésienne de Patri, Diard & Perrier, 2016, qui prend en compte des informations auditives et somatosensorielles). Des propositions ont même été émises quant aux substrats neuroanatomiques des différentes composantes du contrôle prédictif de l'action de parole (cf. Lœvenbruck *et al.*, 2018).

Pour résumer nos propres propositions, comme illustré sur la Figure I-9, toute action de parole correspond à des objectifs phonétiques désirés, qui peuvent être exprimés en termes à la fois articulatoires et acoustiques (cf. discussion plus bas) : les sensations dans les organes de la parole et les sons associés (flèche 1). Nous avons expliqué dans Lœvenbruck *et al.* (2018) que des objectifs visuels sont peut-être également à l'œuvre, la parole mettant en jeu des représentations visuelles, dans le geste manuel l'accompagnant, mais aussi pour permettre (en perception) la lecture labiale, notamment. Mais ceci reste à mieux comprendre. Par ailleurs, ces objectifs phonétiques sont probablement sous une forme supramodale (fusion de plusieurs modalités). Je fournis ci-dessous quelques arguments pour justifier ce format supramodal,

correspondant à la fusion de représentations auditives et somatosensorielles, probablement établie au cours du développement du langage.

Je modifie le schéma initial proposé dans Løevenbruck *et al.* (2018) pour tenir compte, d'une part, de ce nouveau format des objectifs phonétiques et, d'autre part, de résultats récents sur les modèles internes dans le contrôle de l'action (cf. implémentation neuronale et Figure I-10 ci-dessous). J'ajoute deux phases d'intégration multisensorielle. La première s'applique aux retours multisensoriels effectifs (auditifs, somatosensoriels et peut-être visuels) qui sont fusionnés pour être comparés aux signaux supramodaux désirés. La seconde s'applique à l'expérience multisensorielle prédite, elle-même intégrée pour être comparée aux objectifs supramodaux. J'ajoute également une phase de programmation au cours de laquelle la spécification motrice fournie par le modèle interne inverse est transformée en commandes motrices coordonnées, utilisables par l'appareil moteur. Cette phase de programmation se rapproche conceptuellement du *buffer* articulatoire proposé dans le modèle de Levelt (cf. plus haut) introduit pour maintenir en mémoire l'objectif phonétique le temps que l'appareil articulatoire, occupé avec l'énoncé précédent, soit prêt.

Ainsi, comme on le voit sur la Figure I-9, dans le cadre du MCP appliqué à la parole, les objectifs phonétiques supramodaux sont envoyés au contrôleur pour être convertis en une spécification motrice (flèche 2). Cette spécification motrice est transmise à un programmeur (flèche 3) pour coordonner les différents objectifs de parole en cours et générer des commandes motrices. Ces commandes motrices sont ensuite envoyées à l'appareil articulatoire (flèche 4'), afin de produire des mouvements articulatoires, résultant en des sensations proprioceptives et en des sons audibles : l'expérience sensorielle effective. Cette expérience est intégrée (cf. argumentaire ci-dessous) pour fournir un signal dans un format supramodal, comparable à l'objectif initial.

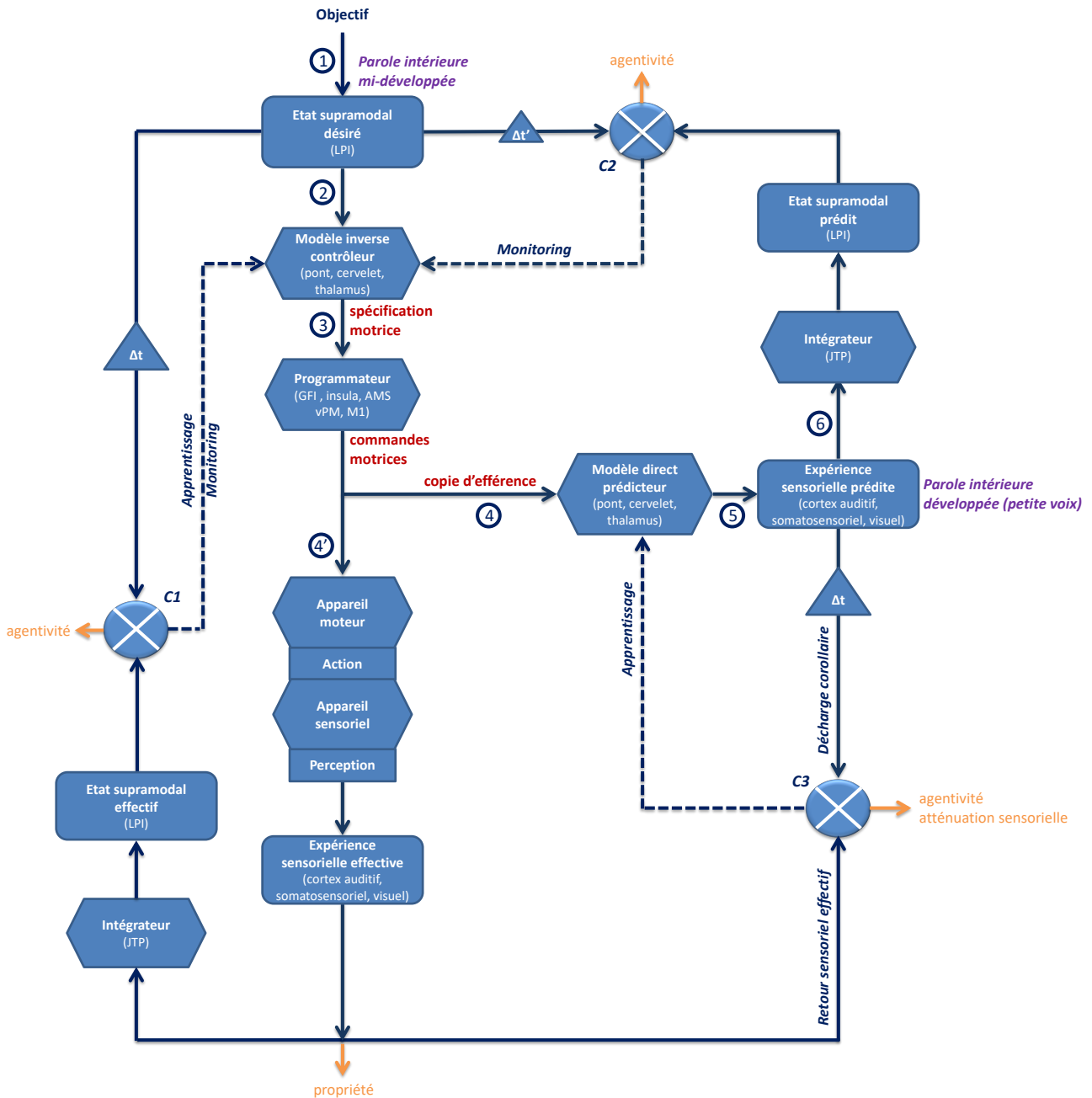


Figure I-9. Le monitoring de l'étape d'articulation par contrôle prédictif
(traduit et révisé de Lævenbruck et al., 2018)

La comparaison (notée C1 sur la Figure I-9) entre le signal de *feedback* supramodal afférent (qui correspond à l'expérience sensorielle effective après intégration) et l'état phonétique désiré (auquel un délai Δt a été ajouté, comme cela est classique en contrôle moteur, pour que les signaux désirés et effectifs soient synchronisés) permet d'ajuster les paramètres de l'action et ainsi d'améliorer le contrôleur. Si la comparaison n'est pas nulle, i.e. si l'expérience sensorielle (après intégration) ne correspond pas à l'état désiré, le contrôleur en est informé (trait pointillé) et les commandes motrices seront ajustées lors d'une prochaine action. La comparaison C1 est donc cruciale pour l'apprentissage moteur (ou en présence de perturbation de l'appareil moteur, qui nécessite un ajustement des commandes motrices). De plus, les signaux ascendants de *feedback* vécu (ainsi que les signaux descendants d'état désiré ou intentionnel) nous fournissent le sens

de la propriété corporelle. Ils permettent de savoir que c'est bien notre corps qui est en mouvement (Gallagher, 2000 ; Franck & Thibaut, 2003 ; Tsakiris *et al.*, 2007). Un premier type de *monitoring* de la production de parole est donc effectué au niveau de *C1*.

Mais ce type de *monitoring* est tardif, puisqu'il est fondé sur l'état effectivement atteint par l'appareil sensori-moteur. Or comme expliqué ci-dessus, nous sommes capables de nous corriger avant même d'avoir produit notre énoncé. Pour corriger les erreurs le plus rapidement possible, une copie d'efférence des commandes motrices (flèche 4) est fournie au modèle direct prédicteur, qui génère des signaux sensoriels prédits, de nature proprioceptive et auditive, qui seront traités par le système sensoriel (flèche 5). Ces signaux multisensoriels sont intégrés en une représentation supramodale (cf. argumentaire ci-dessous). Une comparaison (nommée *C2* sur la Figure I-9) peut alors être effectuée entre les signaux supramodaux prédits et désirés. Un délai de temporisation Δt est appliqué aux signaux désirés pour l'alignement temporel avec les signaux prédits, mais ce délai est minimal (il est représenté en plus petit sur la figure). Si les objectifs et les prédictions ne concordent pas, le contrôleur est ajusté en ligne (trait pointillé) et de nouvelles commandes motrices plus appropriées aux objectifs désirés peuvent être générées. Ainsi, un deuxième type de *monitoring* de la production de parole est effectué au niveau de *C2* : celui-ci, fondé sur notre capacité de prédire les conséquences d'un acte moteur, permet un ajustement rapide et régulier de l'action en cours par rapport aux intentions initiales. La comparaison *C2* est également à l'œuvre lors de la simulation mentale d'une action, pour l'entraînement mental par exemple (Frith, 2005).

Pour parfaire encore le contrôle de nos actions, une troisième comparaison (*C3*) est à l'œuvre, celle entre l'état effectivement atteint, qui correspond à l'expérience multisensorielle vécue (avant fusion supramodale), et l'état multisensoriel prédit à partir de la copie d'efférence des commandes motrices (auquel un délai Δt est appliqué pour l'alignement temporel avec les signaux de *feedback* effectifs). Si cette comparaison entre expérience multisensorielle prédite (décharge corollaire) et *feedback* afférent n'est pas nulle, le prédicteur est ajusté (trait pointillé) ce qui améliorera les prédictions ultérieures.

Ainsi, grâce à ces 3 comparaisons, le *monitoring* par contrôle prédictif permet un suivi et un contrôle fluide de nos actions de parole en cours, en ajustant et corrigeant les commandes motrices émises au cours de l'action. La mise à jour des modèles contrôleur et prédicteur permet aussi d'améliorer nos performances lors de l'apprentissage de nouvelles actions, en généralisant les ajustements pour les productions futures.

Nous verrons dans la section I.2.3 que la comparaison *C1* fournit le sens de la *propriété* et que *C2* et *C3* (peut-être aussi *C1*) jouent un rôle dans le sens de *l'agentivité*, ces deux sens étant des composantes essentielles de la conscience de soi (Gallagher, 2000).

Nature des objectifs phonétiques : somatosensoriels, auditifs, supramodaux ?

Un détail important est souvent omis ou négligé dans les MCP appliqués à la parole, ce qui conduit parfois à des imprécisions ou des contradictions. Il s'agit de la nature précise des signaux désirés, prédits et comparés. Dans les premiers modèles de contrôle de l'action fondés sur les prédicteurs de Smith (e.g. Miall & Wolpert, 1996), il est précisé que le signal d'entrée est l'état désiré de l'appareil moteur musculo-squelettique. Cet état configurationnel correspond à la position et à la vitesse désirées de l'appareil moteur (de la main par exemple). Il est fourni en entrée au contrôleur qui, à partir de cet état désiré et d'une estimation de l'état courant, calcule les commandes motrices. Le point crucial

est que le prédicteur se décompose lui-même en (i) un prédicteur de la dynamique de l'appareil moteur – qui à partir de commandes motrices, fournit l'état prédit de l'appareil moteur – et (ii) un prédicteur de l'appareil sensoriel – qui, à partir de l'état courant de l'appareil moteur, prédit des signaux sensoriels réafférents. L'état courant (position/vitesse) n'étant pas connu du contrôleur, il est estimé, à partir des signaux sensoriels (proprioception par exemple) prédits (en sortie du modèle prédictif de l'appareil sensoriel), des signaux sensoriels réafférents effectifs et de l'état configurationnel prédit (par le modèle prédictif) de l'appareil moteur. L'estimation optimale de l'état de l'appareil moteur peut être modélisée avec un filtre de Kalman (Kalman, 1960), un système dynamique qui permet d'optimiser l'estimation de l'état de l'appareil à l'instant $t+1$, en combinant la prédiction interne de l'état à l'instant $t+1$ (elle-même fondée sur l'estimation de l'état à l'instant t), la prédiction sensorielle à l'instant t et le signal sensoriel effectif à l'instant t (Wolpert *et al.*, 1995 ; Grush, 2004).

Cela signifie que ce qui est *in fine* contrôlé par le modèle est un état configurationnel désiré de l'appareil moteur. Ceci convient pour le contrôle d'actions comme le mouvement du bras ou de la main pour atteindre un objet. Dans le cas du contrôle d'actions de parole, il semble peu probable que l'objet du contrôle par les locuteurs soit la position d'un ou plusieurs articulateurs. Pour la production de certains phonèmes, si l'on peut concevoir que l'objectif contrôlé soit la position des articulateurs (pour /i/ et /u/ par exemple les lèvres sont respectivement en position étirée et arrondie), pour d'autres phonèmes, par contre, il semble inapproprié d'envisager un contrôle reposant uniquement sur la position articulaire (quelle est la position de la lèvre pour /ʌ/ ?). De nombreux travaux suggèrent que la cible visée par les locuteurs est de nature à la fois auditive et articulaire (pour des cibles auditives, cf. par exemple Perkell *et al.*, 1997 ou Guenther *et al.*, 2006 ; pour des cibles articulaires, cf. par exemple Saltzman & Munhall, 1989, Browman & Goldstein, 1989 ou Tremblay, Shiller & Ostry, 2003 ; pour des cibles auditivo-articulaires ou auditivo-somatosensorielles, cf. Løevenbruck, 1996 ; Perrier, Løevenbruck & Payan, 1996 ou Perkell, 2012). Pour résumer le débat, les données sur l'équivalence motrice par exemple (un même son de parole peut être produit par différents gestes articulaires) suggèrent que les locuteurs visent un résultat auditif. Mais les données sur la variabilité acoustique ou sur la production de parole des personnes sourdes suggèrent que les locuteurs ont (également) en tête un objectif articulaire. Toutefois, contrairement à la confusion ou simplification souvent faite (cf. par exemple Hickok *et al.*, 2011, p. 411-414), « articulaire » ne signifie pas ici « moteur ». Ce que les locuteurs visent, c'est une configuration articulaire du conduit vocal qui correspond à un état proprioceptif et tactile (donc somatosensoriel) et non pas à des commandes motrices. Plusieurs possibilités sont donc envisageables. On peut concevoir un objectif en deux temps, d'abord auditif puis articulaire. Le modèle rudimentaire de contrôle de la production de parole proposé dans ma thèse incluait justement une étape intermédiaire, avec un modèle inverse transformant une intention perceptive (auditive) en des trajectoires articulaires désirées, elles-mêmes ensuite converties en commandes motrices par un deuxième modèle inverse (Løevenbruck, 1996). On peut proposer aussi que l'objectif soit de nature auditive pour les sons vocaliques et somatosensorielle pour les sons associés à un contact articulaire, notamment les consonnes occlusives (cf. des arguments dans ce sens dans Perkell, 2012). On peut également envisager que l'objectif soit d'emblée de nature auditivo-somatosensorielle, ce que nous proposons dans Løevenbruck *et al.* (2018), ce que défendent Hickok *et al.* (2011), en confondant, selon moi, les termes « moteur » et

« articulatoire », et ce qu'explicitent Patri, Perrier et Diard (2018) dans un cadre bayésien. Cette position me semble compatible avec la théorie PACT de Schwartz *et al.* (2012), dans laquelle les unités de parole perçues (et produites) sont de nature auditivo-articulatoire (Schwartz *et al.* font l'hypothèse supplémentaire que des représentations motrices sont en jeu lors de la construction des unités de parole, mais leur modèle reste agnostique sur la nature motrice ou non des représentations elles-mêmes, i.e. une fois construites).

Hickok *et al.* 2011 ne s'avancent pas sur la façon dont les deux types de représentation (auditive et somatosensorielle) sont gérées, ni sur l'éventuelle primauté d'un type de représentation sur l'autre. Patri *et al.* (2018) introduisent des variables de contraintes qui modélisent le compromis entre représentations auditives et somatosensorielles. De plus amples recherches sont nécessaires pour avancer sur ce point.

Il me semble, à la relecture de ces différents travaux, mais également de Rauschecker et Scott (2009), qu'une conjecture intermédiaire est plausible. Rauschecker et Scott ont émis l'hypothèse que, dans le processus de production de la parole, le lobule pariétal inférieur est le siège de comparaisons entre des signaux de *feedback* sensoriels afférents, issus du *Planum Temporale*, et de signaux *feedforward* de type copie d'efférence issus des régions prémotrices. Ces signaux prendraient la forme d'ébauche des événements sensoriels en cours (« *a sparse but fast primal sketch of ongoing sensory events* », Rauschecker & Scott, 2009, p. 722). Le modèle de Rauschecker et Scott ne tient pas compte des éléments récents sur les modèles internes et le rôle du cervelet (cf. ci-dessous) et ne peut donc être intégralement repris. Mais il présente l'intérêt d'avoir introduit cette notion d'ébauche des événements sensoriels, impliquant le lobule pariétal inférieur, que je reprends ici. L'implication du lobule pariétal inférieur à ce stade précoce de la production de parole (en amont des modèles internes et de la génération de commandes motrices) est en accord avec des études de neurostimulation. En particulier, Desmurget *et al.* (2009) ont montré que la stimulation électrique directe du lobule pariétal inférieur (BA 40) chez des patients peut produire l'intention de produire un mouvement orofacial, voire l'intention et la sensation de parler (alors qu'aucun mouvement n'est effectué et que l'activité électromyographique orofaciale est celle du repos). Les auteurs ont relié ce résultat au rôle du LPI dans l'intention motrice. Ceci corrobore en tout cas l'hypothèse que le LPI est activé précocement dans le processus de production de la parole et bien en amont des régions prémotrices et motrices.

Encadré I.17. La nature supramodale des objectifs phonétiques

Il ressort de cette parenthèse sur la nature des signaux désirés et prédits que, dans le domaine du contrôle de la parole, les objectifs doivent être, selon moi, définis en termes supramodaux. Je fais l'hypothèse qu'ils constituent une intégration des divers signaux multisensoriels fusionnés, une ébauche supramodale, fusionnant des représentations auditives, somatosensorielles et peut-être visuelles. Ce mélange de toutes les modalités sensorielles serait probablement sous-tendu par le cortex associatif, au niveau du lobule pariétal inférieur. Cette représentation supramodale ébauchée serait construite au fur et à mesure de nos apprentissages langagiers.

Une deuxième parenthèse est importante. En appliquant ce type de modèle à la simulation mentale d'actions et en particulier à la production de parole intérieure (cf. I.2.2.2), certains chercheurs ont noté que la notion d'état prédit conduit à un paradoxe

(Gallagher, 2004 ; Langland-Hassan, 2008). En effet, si, comme le suggère par exemple Frith (2005), la comparaison C2 est à l'œuvre lors de la simulation mentale d'une action (pour l'entraînement mental) alors on doit postuler que l'action imaginée est représentée deux fois : au niveau de l'état désiré et au niveau de l'état prédit. En d'autres termes, si lorsque l'on souhaite produire de la parole intérieure (parler dans sa tête), on génère un état sensoriel désiré (en termes proprioceptifs et auditifs) et si cet état désiré correspond à la petite voix dans la tête, pourquoi faire en plus une prédiction, qui sera redondante ?

Ma réponse est que l'état sensoriel désiré est justement sous forme d'ébauche supramodale, avec fusion des différentes modalités sensorielles. Il n'est pas vécu comme une petite voix articulée mentalement, mais plutôt comme un plan schématique, une esquisse de la succession d'états multisensoriels à atteindre. Il ne s'agit donc pas de l'expérience de parole complète, avec le décours temporel précis des mouvements musculaires et de la propagation du son. Ce qui est contrôlé est une ébauche supramodale fusionnant les différentes modalités sensorielles. Mais l'appareil de parole produit des gestes articulatoires dynamiques qui résultent en une expérience multisensorielle qui se déroule dans le temps. J'introduis donc la notion d' « intégrateur » qui fusionne les signaux multisensoriels en une représentation supramodale schématisée. Comme on le voit sur la Figure I-9, le prédicteur est un modèle de l'appareil sensorimoteur, il fournit une expérience sensorielle simulée. C'est cette expérience dynamique qui correspond à la petite voix dans la tête, dans le cas de la production de parole intérieure délibérée (cf. I.2.2). Cette expérience sensorielle dynamique est, selon moi, intégrée pour fournir ensuite un état fusionné supramodal. Cet état supramodal schématique est alors dans le même format que l'ébauche supramodale désiré. De même, l'expérience sensorielle effective est intégrée en un état supramodal effectif qui peut être comparé à l'état supramodal désiré. Je rajoute que l'état supramodal désiré, avant la génération d'une prédiction (donc d'une petite voix), constitue probablement une forme de parole intérieure plus condensée (ou semi-développée, cf. I.2.2.1). Cette forme de parole intérieure apparaît sous la forme d'une ébauche supramodale schématique, plutôt que sous la forme d'une réalisation multisensorielle complète. Elle est donc probablement perçue comme atténuée sensoriellement, peu audible.

Encadré I.18. Distinction entre objectifs supramodaux et petite voix audible

Les objectifs phonétiques seraient donc définis sous forme d'ébauche supramodale. Ces objectifs correspondraient à une forme semi-condensée de parole intérieure. La prédiction issue de ces objectifs donnerait, elle, lieu à une expérience sensorielle dynamique, la perception d'une petite voix dans la tête, audible, sensible, une forme plus développée de parole intérieure (cf. I.2.2).

Cette différenciation entre l'objectif désiré, de nature schématique et supramodale, et l'expérience sensorielle intermédiaire, de nature développée et dynamique, est en accord avec d'autres propositions dans le domaine du contrôle moteur des actions. Comme le notent Sebanz & Knoblich (2009), citant Prinz (1997), de nombreux travaux montrent que les actions sont planifiées en termes des effets *finaux* visés plutôt qu'en termes proximaux (i.e. qu'en termes de la dynamique des mouvements à accomplir).

Implémentation neuronale du contrôle prédictif de l'étape d'articulation

Ce modèle computationnel du MCP de la parole, limité au niveau de l'articulation, peut être mis en lien avec les réseaux neuronaux physiques qui pourraient sous-tendre les différents processus évoqués. Le réseau cérébral qui sous-tend la production de mots, que ce soit la dénomination de mots à partir d'images ou à partir de stimuli audio, est relativement bien décrit, même si tous les auteurs ne s'accordent pas sur la séquence dynamique des activations et si le rôle du cervelet est souvent négligé (Baciu *et al.*, 1999 ; 2005 ; Baciu & Perrone-Bertolotti, 2015 ; Duffau *et al.*, 2014 ; Grappe *et al.*, 2018 ; Guenther & Vladusich, 2012 ; Hickock, 2012 ; Kell *et al.*, 2011 ; Keller & Kell, 2016 ; Llorens *et al.*, 2011 ; Moritz-Gasser & Duffau, 2013 ; Munding *et al.*, 2016 ; Ueno *et al.*, 2011 ; Price, 2012 ; Riès *et al.*, 2013 ; Roelofs, 2014 ; Tian & Poeppel, 2013). Ce réseau inclut des régions cérébrales de l'hémisphère gauche – le pôle temporal, le gyrus temporal moyen, le gyrus temporal supérieur postérieur, l'hippocampe, le gyrus angulaire, le gyrus supramarginal, la jonction temporo-pariétale, le gyrus frontal inférieur, les régions motrices et prémotrices – ainsi que le cervelet droit.

Je résume ci-dessous l'implémentation neuronale que nous avons envisagée dans Lœvenbruck *et al.* (2018) pour l'étape d'articulation de mot isolé (après l'accès lexical), en nous appuyant sur des données expérimentales et des hypothèses théoriques détaillées (pour les arguments : Perrone-Bertolotti *et al.* (2014, 2016), Rapin *et al.* (2013, 2016) ainsi que II.2.1). Cette implémentation partage certaines hypothèses avec les modèles cités ci-dessus. J'apporte toutefois quelques modifications à notre proposition initiale, pour tenir compte de réflexions plus abouties sur le format des représentations en jeu (représentation supramodale dans le LPI), ainsi que de données récentes sur les liens entre cervelet et cortex cérébral dans la spécification des modèles internes (cf. ci-dessous). Cette implémentation révisée est schématisée sur la Figure I-10.

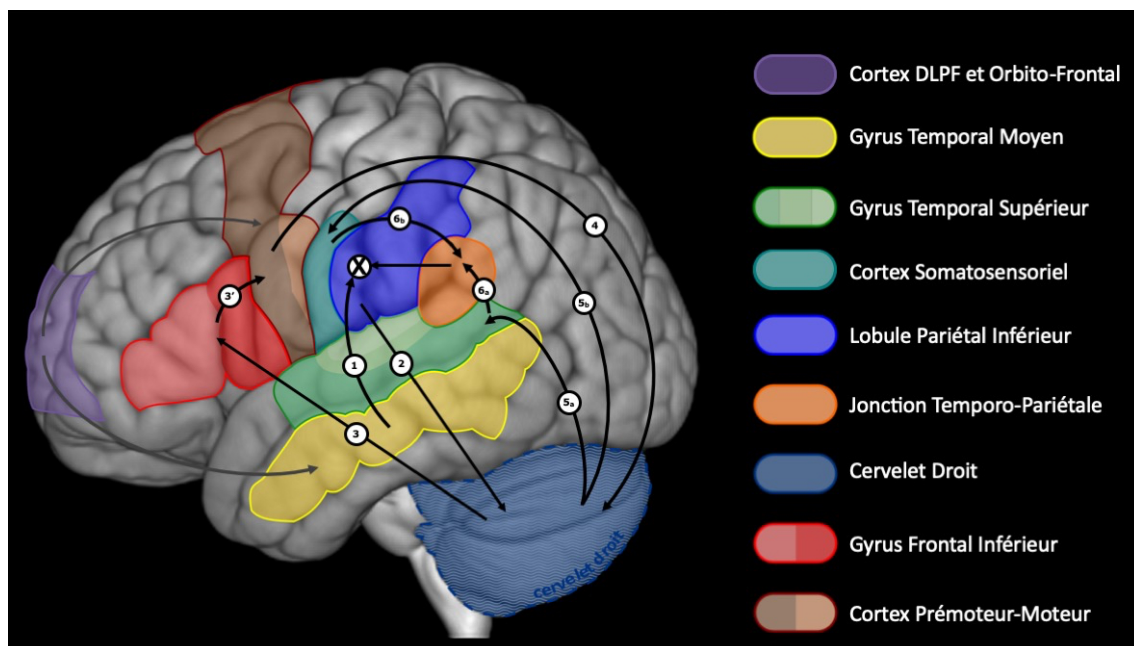


Figure I-10. Réseau neuronal impliqué dans l'étape d'articulation par contrôle prédictif, incluant la production de parole intérieure

*Du lemme, issu du gyrus temporal moyen, aux objectifs phonétiques supramodaux dans le gyrus supramarginal, puis à la petite voix perçue sensoriellement, via la copie d'efférence de la commande motrice générée dans le cortex frontal inférieur et inhibée par le cortex DLPF et orbito-frontal (révision de Lœvenbruck *et al.*, 2018, cf. Grandchamp *et al.*, soumis).*

Comme illustré sur la Figure I-10, lors de la production d'un mot isolé, un lemme est d'abord récupéré, ce qui correspond à l'activation du gyrus temporal moyen gauche. Ce lemme est converti en objectifs phonétiques, dans un format supramodal activant le lobule pariétal inférieur gauche (flèche 1). Contrairement à ce que nous avons initialement proposé dans Løevenbruck *et al.* (2018) dans la lignée de Hickok (2012), je pense que la spécification des objectifs phonétiques ne se fait pas *via* deux voies, l'une auditive et l'autre somatosensorielle, qui seraient intégrées en aval, mais que la spécification est supramodale, dès ce stade initial. En effet, comme mentionné à propos du modèle computationnel de la Figure I-9, cet objectif est supramodal, intégrant des représentations auditives et somatosensorielles. Il est transmis directement au cervelet (flèche 2). Les arguments pour cette modification de notre implémentation initiale ont été détaillés ci-dessus, en ce qui concerne la nature supramodale des représentations de parole initiales. En ce qui concerne le rôle du cervelet dans les modèles internes et les connexions bidirectionnelles entre le cervelet et les cortex moteur, sensoriel et associatif, la nouvelle implémentation s'appuie sur des publications récentes, présentées plus loin.

Une fois la spécification supramodale définie, le cervelet droit (probablement cortex cérébelleux antérieur, cf. ci-dessous) met en œuvre un modèle interne inverse qui transforme les objectifs phonétiques en une spécification motrice. Cette spécification est transmise au cortex frontal (flèches 3 et 3'). Plus précisément, les programmes moteurs sont générés, en coordonnant la spécification motrice, issue du cervelet, avec les actions de parole en cours. Les régions impliquées sont d'abord le gyrus frontal inférieur gauche (GFI, région de Broca) et l'insula (flèche 3), puis le cortex prémoteur ventral gauche (vPM), l'aire motrice supplémentaire (AMS) et le cortex moteur primaire (M1) (flèche 3'). Des commandes motrices sont alors générées et transmises à l'appareil moteur.

Ainsi, cette nouvelle proposition se démarque de celle de Løevenbruck *et al.* (2018). Notre implémentation initiale comportait deux voies de traitement. Dans la première, le cervelet générait des commandes motrices à partir d'une spécification somatosensorielle. Ces commandes étaient directement transmises aux aires prémotrices et motrices. Dans la seconde voie, la spécification auditive était transformée en un programme moteur, via la JTP et le GFI. Mais compte tenu des données récentes sur le rôle du cervelet (cf. ci-dessous), je reformule nos hypothèses en intégrant les deux voies dès le départ (en amont du cervelet) et en distinguant un modèle inverse contrôleur (dans le cervelet) et un programmeur (dans les régions frontales). Ma nouvelle proposition tient également compte des données en électrocorticographie (ECoG) de Flinker *et al.* (2015) lors de la répétition de mots monosyllabiques. Ces données ECoG temporelles précises permettent de détailler le décours temporel des activations dans le lobe temporal, dans la région de Broca et dans les régions motrices. Elles indiquent que les pics d'activités cérébrales démarrent dans le GTS et le STS (il s'agit de répéter un mot), puis se propagent à la région de Broca (GFI) avant d'atteindre les cortex prémoteur et moteur. L'activation dans la région de Broca est antérieure à la production et s'arrête avant la production de parole, alors que celle des régions prémotrices et motrices continue pendant la production. Selon Flinker *et al.* (2015), le GFI semble donc impliqué dans la planification pré-articulatoire plutôt que dans la coordination en ligne des articulateurs, mais il joue bien un rôle dans l'encodage articulatoire (et pas uniquement dans les processus phonologiques de plus haut niveau, comme la syllabification, ce qui avait été proposé par Indefrey & Levelt, 2004). Je propose donc une programmation en deux étapes (flèche 3 puis 3' sur la Figure I-10).

Les commandes motrices sont ensuite transmises aux muscles du larynx et du conduit vocal, qui permettent la mise en mouvement de l'appareil vocal et la production de sons. Les mouvements générés et les sons produits sont perçus par les systèmes périphériques somesthésiques et auditifs qui fournissent des retours sensoriels, traités par les cortex somatosensoriel et auditif.

En parallèle, une copie d'efférence des commandes motrices est envoyée au cervelet (probablement cortex cérébelleux postérieur, cf. ci-dessous) pour générer une prédiction de l'état sensoriel qui résulterait de ces commandes motrices (flèche 4). Le cervelet contient en effet un deuxième modèle interne, celui-ci direct (prédicteur), qui permet de générer un signal sensoriel prédit, lui-même transmis aux cortex sensoriels via le thalamus (flèches 5a et 5b). C'est ce signal multisensoriel prédit (auditif et somatosensoriel, voire peut-être aussi visuel) qui correspond à la parole intérieure, la petite voix que l'on entend dans sa tête lorsque l'on parle intérieurement et que l'on inhibe la sortie motrice effective (cf. I.2.2 et II.2.1). Après l'application d'un délai, le signal multisensoriel prédit pourra être comparé au signal sensoriel effectivement reçu, suite à la production vocale. Cette comparaison se fait possiblement au niveau de l'olive inférieur, mais elle pourrait avoir lieu dans le cortex auditif et dans le cortex somatosensoriel (cf. Sokolov *et al.*, 2017). Les signaux multisensoriels prédits et effectifs sont ensuite transformés chacun, via l'activation de la jonction temporo-pariétale, en une représentation supramodale, qui peut être comparée aux objectifs désirés, au niveau du cortex associatif (LPI). La comparaison entre les états désirés et prédits ainsi que désirés et effectifs se fait possiblement à ce niveau (cf. ci-dessous, discussion sur le rôle du cervelet). En l'état actuel des connaissances, il est difficile de se prononcer sur la localisation de ces différentes comparaisons et je ne les ai pas indiquées sur le schéma.

Enfin, l'ensemble de ces activations est sous contrôle du système exécutif, notamment du cortex rostral préfrontal (BA 10) et du cortex cingulaire antérieur (BA 32), qui permettent d'enclencher le processus de production de parole ou au contraire d'inhiber la sortie motrice, dans le cas de la parole intérieure (flèches grises non numérotées).

Rôle du cervelet dans l'étape d'articulation : modèles contrôleur et prédicteur

Les modifications que j'apporte ici à notre proposition initiale sur l'architecture neurale de l'étape d'articulation du processus de *production* de la parole (Løevenbruck *et al.*, 2018), permettent ainsi d'être plus conforme aux propositions récentes sur le rôle du cervelet dans la coordination des actions en général et du langage en particulier (Buckner *et al.*, 2011 ; De Smet *et al.*, 2013 ; Diedrichsen & Zotow, 2015 ; Imamizu & Kawato, 2009 ; Leiner *et al.*, 1986 ; Mariën *et al.*, 2014 ; Raichle *et al.*, 1994 ; Sokolov *et al.*, 2017 ; Stoodley *et al.*, 2012 ; Tamada *et al.*, 1999). Les travaux en neurosciences cognitives des trente dernières années, issus de données de neuroimagerie et de connectivité (fonctionnelle et de repos), ou de données neuropsychologiques sur des patients cérébro-lésés indiquent que le cervelet est bidirectionnellement connecté, via le thalamus, le pont et l'olive inférieure, non seulement au cortex moteur, mais également au cortex préfrontal, pariétal et temporal et qu'il joue un rôle majeur non seulement dans des tâches de coordination motrice mais également dans de nombreux processus affectifs et cognitifs, dont le langage (voir Buckner *et al.*, 2011, De Smet *et al.*, 2013, Diedrichsen & Zotow, 2015, Mariën *et al.*, 2014, ainsi que Sokolov *et al.*, 2017, pour des revues de la littérature et/ou des schémas des connexions). Il semble donc admis que le cervelet reçoive des informations des régions corticales impliquées dans la production

du langage et les influence en retour. Le rôle exact du cervelet reste cependant débattu et fait l'objet de nombreux travaux de recherches, Mariën *et al.* (2014) évoquent même une « énigme en cours ».

Dans le cadre du contrôle moteur, Imamizu & Kawato (2009) ont fourni des arguments convaincants (fondés sur des données de neuroimagerie et de neurophysiologie très précises) sur l'implication du cervelet dans l'acquisition des deux types de modèles internes : inverses (contrôleurs) et directs (prédicteurs). Selon Imamizu & Kawato (2009), les modèles *inverses* (contrôleurs) seraient donc spécifiés dans le cervelet, qui fournirait en sortie des signaux (commandes motrices) vers les régions prémotrices dorsales (PMd), l'aire motrice supplémentaire (AMS) et les régions motrices primaires (M1), ainsi que, probablement, vers les régions prémotrices ventrales (PMv) dont la région de Broca. La sélection des modèles internes appropriés à l'action en cours se ferait au niveau des régions pariétales (Lobule Pariétal Supérieur, LPS et Lobule Pariétal Inférieur, LPI) et du cortex préfrontal dorsolatéral (cortex DLPF). Imamizu & Kawato font l'hypothèse d'une boucle impliquant le DLPF, le LPI, le LPS et le cervelet. Le DLPF serait bidirectionnellement connecté au LPI, lui-même influençant le LPS. La sélection de modèles internes (directs et inverses) dans le cervelet se ferait de façon prédictive par le LPS et le LPI. A la sortie du modèle *direct* (prédicteur) une erreur de prédiction serait calculée et fournie au LPI qui modulerait ainsi, de façon post-dictive, les sélections de modèle interne ultérieures. Enfin, la prédiction sensorielle issue du modèle direct sélectionné serait envoyée au cortex occipito-temporal latéral et au sillon temporal supérieur (STS).

Imamizu et Kawato (2009) montrent également que ces boucles cortico-cérébelleuses impliquant des modèles internes directs et inverses ne sont pas uniquement en jeu dans la motricité, mais également dans de nombreuses fonctions cognitives, dont le langage, tant en production qu'en compréhension. Le rôle cognitif du cervelet est désormais démontré, mais les mécanismes des interactions entre le cortex et le cervelet dans les fonctions cognitives, et notamment dans le langage, restent encore inconnus. En plus de son rôle dans la construction et la mise à jour de modèles internes (e.g. Gomi *et al.*, 1998 ; Ito, 2008 ; Imamizu & Kawato, 2009), il a été proposé que le cervelet soit impliqué dans le traitement du temps (Braitenberg, 1967 ; Ivry & Spencer, 2004 ; Ivry *et al.*, 2002), dans le séquençage et l'ordonnancement des unités du langage (Ackermann, 2008 ; Molinari *et al.*, 2008), dans la sélection de commandes motrices pour la production de syllabes (Guenther & Perkell, 2004 ; Guenther *et al.*, 2006), dans la modulation et l'atténuation du comportement, tant moteur, que cognitif ou affectif (« *dysmetria of thought hypothesis* », Schmahmann, 2004).

Une proposition intégrative a été formulée récemment par Sokolov, Miall et Ivry (2017), qui ont rassemblé les données récentes dans ce domaine. Selon eux, deux principes computationnels généraux sont à l'œuvre dans les boucles cérébro-cérébelleuses (connexions bidirectionnelles) : la prédiction et l'apprentissage fondé sur l'erreur. Sokolov *et al.* (2017) montrent l'existence de circuits cortico-ponto-cerebello-dento-thalamo-corticales qui forment une série de boucles parallèles, se terminant dans les régions corticales étant elles-mêmes la source de signaux en direction du cervelet. Les régions corticales envoyant des signaux au cervelet (via les noyaux pontiques) et recevant des signaux de la part du cervelet (via les noyaux profonds du cervelet (notamment noyaux dentelés) puis le thalamus) sont le cortex préfrontal, le cortex temporal et le cortex pariétal.

Selon Sokolov *et al.* (2017) un rôle majeur joué par le cervelet est celui de modèle interne direct *prédicteur*. Le cervelet reçoit une copie d'efférence des commandes motrices pour prédire les conséquences sensorielles des actions. L'erreur de prédiction sensorielle (la différence entre la sortie prédite et la sortie effective) serait, selon eux, calculée au niveau de l'olive inférieure, puis renvoyée au cervelet (ainsi qu'aux noyaux profonds du cervelet, dont les noyaux dentelés, et au thalamus) pour mettre à jour le modèle interne prédicteur. L'olive inférieure est en effet considérée comme un composant essentiel du processus de comparaison entre les signaux prédits et les signaux effectifs afférents, car elle reçoit des signaux convergents, signaux excitateurs issus des structures corticales, sous-corticales, spinales et signaux inhibiteurs issus du cervelet. Sokolov et collègues notent qu'il reste encore à vérifier si la sortie du cervelet est une véritable estimation de l'état prédit ou un simple signal de mise à jour de l'estimation de l'état, elle-même générée dans le cortex pariétal ou moteur. Ils notent également qu'il n'est pas encore certain que l'erreur soit calculée au niveau de l'olive inférieure plutôt que dans des régions corticales ou sous-corticales (l'olive n'étant alors qu'un relai du signal d'erreur, de ces régions vers le cervelet). De plus, la comparaison vraiment pertinente dans le cadre du contrôle prédicteur n'est pas celle entre signaux prédits et effectifs, que Sokolov *et al.* (2017) décrivent en détail, mais plutôt celle entre signaux prédits et désirés. Je fais donc des propositions ci-dessous (cf. Figure I-11) sur l'implémentation de cette comparaison, que des travaux ultérieurs pourront vérifier ou amender.

Sokolov et collègues ajoutent que le cervelet contribue probablement également à la génération des commandes motrices, c'est-à-dire qu'il jouerait également un rôle de modèle interne inverse *contrôleur*, en accord avec les propositions de Imamizu & Kawato (2009) décrites ci-dessus.

Enfin, Sokolov et collègues montrent que les circuits cortico–ponto–cerebello–dento–thalamo–corticaux impliqués dans la motricité – mettant en jeu ce qu'ils nomment le « cervelet moteur » – sont également à l'œuvre dans le langage, la cognition sociale et le traitement affectif – mettant alors en jeu le « cervelet cognitif ». Tout comme le cervelet moteur prédit des changements d'état sensoriel, le cervelet cognitif prédirait des changements d'état perceptifs ou mentaux et fournirait ces prédictions aux régions corticales associatives. En se référant aux travaux de Pickering et Clark (2014) sur les systèmes prédicteurs dans la production et la compréhension du langage (cf. aussi la discussion sur le modèle de Pickering & Garrod, 2013, ci-dessous), Sokolov *et al.* proposent que le rôle de prédicteur et de régulateur des erreurs joué par le cervelet pourrait ainsi expliquer sa contribution aux aspects lexical, syntaxique et sémantique du langage, tant en production qu'en compréhension. Ils étendent cette contribution du cervelet aux traitements affectifs et sociocognitifs, en argumentant que prédiction et détection/régulation des erreurs sont également en jeu dans la compréhension des intentions et émotions d'autrui.

Sokolov et collègues fournissent ainsi un cadre unifié permettant de rendre compte de façon intégrative de la contribution du cervelet dans différentes fonctions motrices, cognitives et affectives, en montrant qu'il permet, via la prédiction d'états et le suivi de signaux d'erreurs, d'optimiser et fluidifier les traitements et comportements cognitifs.

Parmi les nombreuses questions qu'il reste à élucider concernant les interactions entre le cervelet et le reste du cerveau dans la cognition, Sokolov et collègues notent que la double (voire triple) représentation du cortex dans le cervelet est une énigme. En effet,

les travaux de Buckner *et al.* (2011), ainsi que Diedrichsen & Zotow (2015) notamment, ont montré, à partir de données de neuroimagerie fonctionnelle de tâche et de repos, que le cervelet possède une représentation en deux exemplaires (voire trois) de la quasi-totalité du télencéphale (cortex et structures sous-corticales). La latéralisation de ces représentations est inversée, le cortex cérébral gauche étant principalement représenté dans le cortex cérébelleux droit et réciproquement (Stoodley *et al.*, 2012). Les régions somato-motrices sont représentées une première fois dans le lobe cérébelleux antérieur, en respectant l'organisation topologique du cerveau, mais en l'inversant, et une deuxième fois (cette fois à l'endroit) dans le lobe postérieur. Il en va de même pour les cortex cérébraux prémoteurs et associatifs, qui sont représentés intégralement une première fois, de façon homotopique mais à l'envers, dans le cervelet antérieur (s'étendant jusqu'à Crus I/II) et une seconde fois en miroir, de façon homotopique et à l'endroit, dans le cervelet postérieur (partant de Crus I/II).

J'ose ici une spéculation, qui me permet de répondre à ma propre interrogation sous-jacente aux travaux sur le rôle du cervelet, et pourtant jamais clairement énoncée. La plupart des travaux récents s'accordent sur le fait que le cervelet joue un rôle dans les mécanismes de prédiction qui semblent cruciaux dans la cognition. Qui dit prédiction, dit prédicteur et donc modèle interne direct. Mais pour que la prédiction ait lieu, il faut une copie d'efférence des commandes motrices et donc un système de génération de ces commandes motrices, soit un modèle interne inverse, qui, selon Kawato et collègues (cf. plus haut), est lui-même sous-tendu par le cervelet. En d'autres termes, comme il a été simplement esquissé ci-dessus, certains auteurs (et notamment Kawato et collègues) défendent l'idée que le cervelet soit à la fois le siège des modèles prédicteurs et contrôleurs (direct et inverse). Cette idée est séduisante, mais elle se heurte à un problème computationnel si la même structure (le cervelet) doit gérer des signaux (des commandes motrices) à la fois en sortie (du contrôleur) et en entrée (du prédicteur) : une boucle critique. Le fait que le cervelet dispose d'une double association avec le cerveau permet justement de sortir de cette impasse théorique.

Encadré I.19. Conjecture sur les rôles respectifs des lobes antérieur et postérieur du cervelet dans l'étape d'articulation : contrôleur et prédicteur

Je fais donc ici la conjecture – vérifiable empiriquement en utilisant deux tâches, impliquant l'une plus les modèles inverses et l'autre plus les modèles directs – que la partie antérieure (ou postérieure) du cervelet est connectée avec le cerveau (via les noyaux pontiques) pour générer les commandes motrices, dont une copie d'efférence est fournie à la partie postérieure (ou antérieure) du cervelet pour calculer une prédiction sensorielle. Cette prédiction sensorielle pourrait elle-même être comparée (via l'olive inférieure, cf. ci-dessus) aux états désirés et effectifs, à travers les connexions de cette partie du cervelet avec le cortex et les structures sous-corticales et spinales. Ceci expliquerait la rapidité de ces processus de contrôle et de prédiction.

J'ai avancé que le modèle contrôleur serait plutôt sous-tendu par la partie antérieure du cervelet et le modèle prédicteur par la partie postérieure, mais je n'ai pas d'argument pour cela, à part peut-être la distance plus courte entre le pont et le cervelet pour le modèle qui doit être utilisé en premier. L'organisation inverse est donc tout à fait envisageable également.

Cette conjecture est illustrée dans la Figure I-11. La voie correspondant à l'objectif désiré est représentée en vert, celle correspondant aux prédictions est en rouge, celle correspondant aux états effectifs en noir. Les signaux d'erreur sont en gris. Je fais l'hypothèse que l'état désiré (ou spécification supramodale), issu du cortex associatif (et en amont du cortex temporal moyen), est transmis au cervelet antérieur (contrôleur) via les noyaux pontiques. Le contrôleur transforme cette spécification supramodale en une spécification motrice, envoyée aux cortex prémoteur et moteur (après passage par le GFI, cf. ci-dessus), via les noyaux cérébelleux profonds (NCP) et le thalamus (Th). Des commandes motrices sont générées et envoyées à l'appareil moteur lui-même. La mise en mouvement de l'appareil moteur résulte en des sensations auditives et somatosensorielles dans l'appareil sensoriel. Ce retour sensoriel effectif est traité par les cortex sensoriels correspondant aux diverses modalités (auditive, somatosensorielle, visuelle) et est transmis à la jonction temporo-pariétale pour conduire à une représentation intégrée traitée par le cortex associatif. La représentation supramodale qui en résulte est comparée à la spécification supramodale initiale (à laquelle un délai Δt a été ajoutée (comparaison C1, cf. aussi Figure I-9). Cette comparaison peut donner lieu à une erreur de contrôle, qui est transmise au cervelet antérieur pour améliorer les contrôles ultérieurs. En parallèle, une copie d'efférence des commandes motrices est transmise au cervelet postérieur (prédicteur) via les noyaux pontiques. Le prédicteur fournit une prédiction sensorielle, transmise aux cortex sensoriels via les NCP et le Th, ce qui donne lieu à un état multisensoriel (percept) prédit (la petite voix dans la tête). Après ajout d'un délai Δt (cf. aussi la Figure I-9), ce signal prédit est également fourni à l'olive inférieure (OI). A ce niveau, la prédiction est comparée avec l'état sensoriel effectif (comparaison C3, cf. aussi sur la Figure I-9 ci-dessus), ce qui donne éventuellement lieu à une *erreur de prédiction*, transmise au modèle interne prédicteur (cervelet postérieur) pour le mettre à jour. L'état multisensoriel prédit est également intégré en une représentation supramodale prédite qui est traitée par le cortex associatif et comparée à la spécification supramodale désirée (comparaison C2, cf. aussi sur la Figure I-9 ci-dessus). Cette comparaison donne éventuellement lieu à une *erreur de contrôle*, transmise au modèle interne contrôleur pour ajuster les commandes motrices avant même que l'action soit effectuée.

Je ne fais pas d'hypothèse sur le lieu précis des comparaisons entre état désiré et effectif (C1) et désiré et prédit (C2). Je propose qu'elles se fassent après transit par le cortex associatif, mais d'autres possibilités sont envisageables et ceci reste à clarifier. Par ailleurs, je note qu'un argument supplémentaire pour le rôle du cervelet dans le contrôle prédictif de la parole est qu'il a été montré qu'il permet des traitements temporels extrêmement précis, adaptés à différentes situations de contrôle moteur (de Zeeuw & Ten Brinke, 2015). L'introduction de délais dans le modèle de la Figure I-11, dispositif classique en robotique pour synchroniser les signaux, est donc biologiquement plausible.

Par conséquent, contrairement à ce que nous avons proposé dans Løevenbruck *et al.* (2018), il ne me paraît pas adapté d'impliquer le cervelet uniquement dans la voie somatosensorielle et d'envisager le circuit de façon aussi sérielle que nous l'avions décrit, avec des activations corticales sensorielles suivies d'activations cérébelleuses, elles-mêmes suivies d'activations corticales motrices. Il me semble qu'il est plus juste d'envisager un circuit comportant des boucles cérébello-corticales parallèles, intégrant des délais pour synchroniser les signaux et permettre les comparaisons à point nommé.

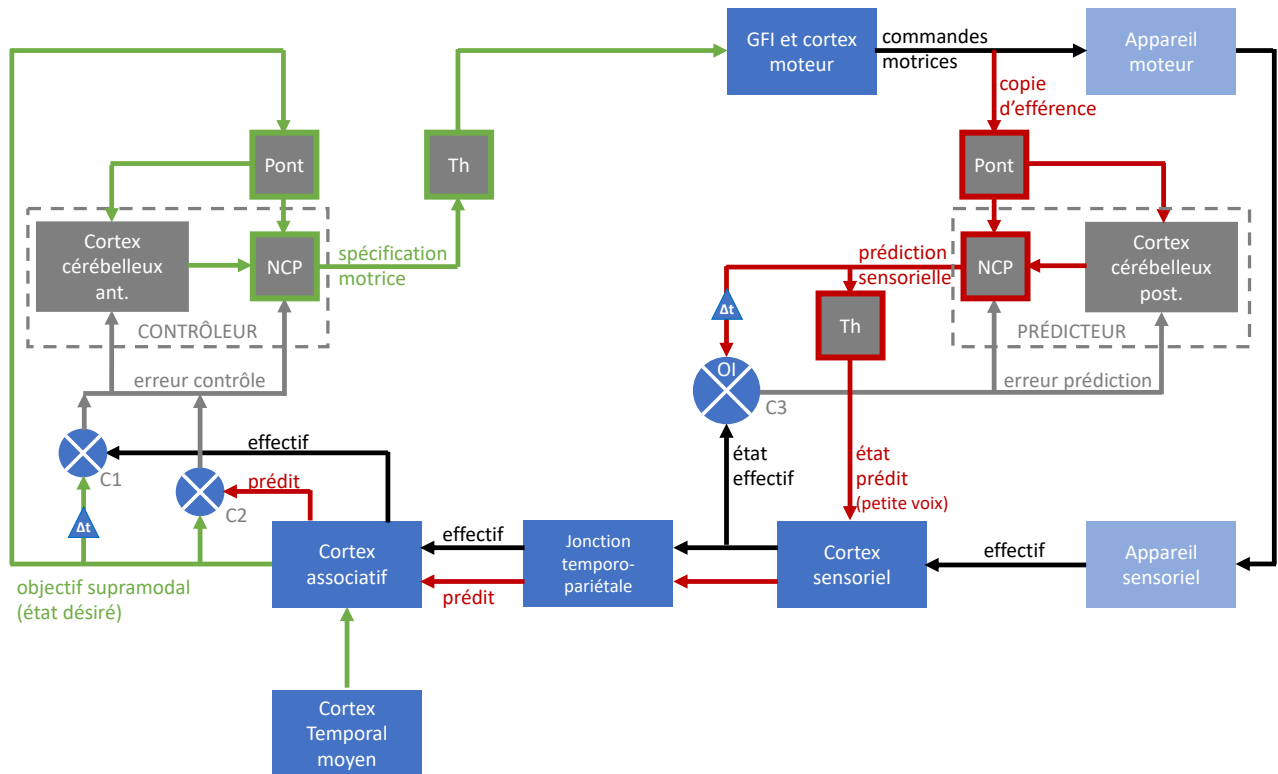


Figure I-11. Conjecture (non publiée) sur l'implémentation dans le cervelet des modèles internes inverse (contrôleur) et direct (prédicteur) nécessaires au MCP de la parole, et sur les connexions entre le cervelet et les cortex cérébraux sensoriel et moteur. Cette conjecture est inspirée et complétée de Sokolov et al. (2017). Pour des raisons de lisibilité, les noyaux pontiques, les noyaux profonds cérébelleux et le thalamus ont été dupliqués, mais il s'agit bien des mêmes structures. Abréviations : Pont : noyaux pontiques, NCP : noyaux cérébelleux profonds, Th : thalamus, OI : olive inférieure

Ainsi, le dispositif de *monitoring* correspondant au dernier niveau de la production de parole, celui de l'articulation (après la conceptualisation et la formulation), peut être mis en place de façon efficace grâce à un MCP qui met en œuvre des objectifs phonétiques désirés, des prédictions de signaux de parole et des signaux de parole effectifs. Ces trois entités peuvent être mises en correspondance avec les différentes formes de *verbum* déjà entrevues de façon originale par Thomas d'Aquin dans *De Veritate* (1256-1259). Ainsi que l'explique Rico (2008), Thomas d'Aquin distingue en effet le *verbum vocis* (parole prononcée), l'*exemplar exterioris verbi* (modèle de la parole extérieure) et le *verbum cordis* (parole du cœur). Il me semble que le *verbum vocis*, la parole extérieure, perceptible, exprimée oralement correspond aux signaux de parole effectifs. D'après Rico (2008), l'*exemplar exterioris verbi* pré-existe dans l'esprit de celui qui prononce le verbe extérieur, c'est une parole intérieure, une image acoustique du mot prononcé. Je l'associe donc à la sortie prédite, simulée, qui constitue la parole intérieure. Enfin le *verbum cordis*, intérieur et premier dans l'ordre de la nature (i.e. avant les deux autres), constitue selon Rico « l'intellection intérieure », ce qui est signifié par la parole intérieure. Il est immatériel et incorporel, autrement dit, en termes contemporains, abstrait. Il ne peut donc être assimilé à l'objectif phonétique désiré, qui nous l'avons vu ci-dessus s'exprime déjà en termes articulatoires et acoustiques. Il paraît correspondre, selon Rico, à l'idée ou au concept non verbalisés, et précède ainsi l'objectif phonétique. Thomas d'Aquin me semble avoir mis le doigt sur le fait que la production de parole requiert un

contrôle enchâssant différents niveaux. Le modèle par contrôle prédictif décrit ci-dessus est insuffisant, car le-la locuteur·rice doit confronter les sorties effective (*verbum vocis*) ou prédite (*exemplar exterioris verbi*) à ce qu'il·elle avait l'intention de signifier (*verbum cordis*), au niveau phonétique mais également au niveau conceptuel. Le modèle par contrôle prédictif permet simplement aux locuteurs de s'assurer que les commandes motrices générées à partir de leur objectif phonétique fourniront bien les mouvements articulatoires et sons désirés. En cas d'erreur de prédiction, les commandes motrices peuvent être corrigées. Qu'en est-il du *monitoring* aux autres niveaux de la production de la parole ? Comment le-la locuteur·rice s'assure-t-il·elle que le signal de parole produit correspond au message conceptuel désiré ? Un enchâssement des systèmes de *monitoring* doit donc être envisagé avec une hiérarchie du contrôle, depuis le niveau du concept jusqu'au niveau du son, les différents niveaux dépendant les uns des autres.

Monitoring hiérarchique par contrôle prédictif

Comme mentionné ci-dessus, Sokolov *et al.* (2017) ont proposé que le rôle de prédicteur du cervelet dans le contrôle moteur s'étende au langage, conférant au cervelet un caractère cognitif. Ce cervelet cognitif pourrait prédire les changements d'état mentaux et fournirait ces prédictions aux régions corticales associatives. Il pourrait ainsi contribuer aux aspects lexical, syntaxique et sémantique du langage, tant en production qu'en compréhension. Il me semble que cette proposition est peu tenable car elle nécessiterait un mécanisme complexe de régulation des interactions entre le cervelet et le cortex à chacun des niveaux linguistiques, pour que les signaux ne soient pas confondus. Il me paraît plus approprié de concevoir un enchâssement hiérarchique, dans lequel le cervelet jouerait un rôle au niveau où on l'attend, i.e. au niveau du contrôle moteur.

La notion de hiérarchie dans le contrôle moteur de l'action a été introduite et justifiée dès les années 1980 (cf. pour les travaux sur l'écriture : Keele *et al.*, 1990, ou sur la saisie d'objet : Jeannerod, 1984). Dans ce cadre, Haruno, Wolpert & Kawato (2003) ont proposé un modèle de contrôle des actions complexes comprenant une hiérarchie structurée d'actions élémentaires, le modèle HMOSAIC (*Hierarchical Modular Selection and Identification for Control*). Dans ce modèle, des paires de modèles directs et inverses (des modules) sont organisées de façon hiérarchique pour relier les intentions de plus haut niveau avec les opérations motrices bas niveau. Chaque niveau est constitué de plusieurs modules comprenant chacun un modèle direct (modèle prédictif) et un modèle inverse (modèle de contrôle). A chaque état, des probabilités sont calculées représentant la part de responsabilité (ou l'implication) de chaque module à chaque niveau. Le plus haut niveau de la hiérarchie reçoit en entrée une trajectoire symbolique (abstraite) désiré ainsi que les probabilités *a posteriori* des modules du niveau subordonné. Le modèle de contrôle du plus haut niveau génère des probabilités *a priori* pour les modules du niveau subordonné, c'est-à-dire qu'il décrit quels sont les modules subordonnés qui doivent être prioritairement sélectionnés. Le modèle prédictif associé prédit, lui, les probabilités *a posteriori* des modules du niveau subordonné. L'erreur de prédiction permet de pondérer les modèles prédictifs et de contrôle.

Grafton et collègues (Grafton & Hamilton, 2007 ; Grafton *et al.*, 2009) ont fourni des données de neuroimagerie en faveur d'une hiérarchie distribuée de la représentation de l'action, avec des pistes de modélisation anatomo-fonctionnelle. Leurs conclusions sont en accord avec les propositions computationnelles de Haruno, Wolpert & Kawato.

Pacherie (2008) a elle-même formulé une version étendue du MCP permettant de ne pas limiter la description des actions à celle des mouvements du corps, mais bien de remonter, de façon hiérarchique, jusqu'aux intentions sous-jacentes aux actions. Son modèle est compatible avec les propositions de Haruno *et al.* (2003). Selon elle, les intentions sont de trois niveaux hiérarchiques : les intentions distales (D-intentions), les intentions proximales (P-intentions) et les intentions motrices (M-intentions). Ces trois niveaux sont organisés en cascade, les D-intentions générant les P-intentions de façon causale et les P-intentions produisant elles-mêmes des M-intentions. Les D-intentions correspondent à des intentions de faire quelque chose dans le futur (e.g. « finir de rédiger mon HDR »). Les P-intentions sont concomitantes avec les actions en cours et participent à leur supervision (e.g. « taper sur le clavier de l'ordinateur »). Les M-intentions consistent en des représentations motrices (représentations de l'objet, de l'environnement et des mouvements à faire pour atteindre l'objet compte tenu de l'environnement et des contraintes biomécaniques propres à l'agent qui effectue l'action). Ce dernier niveau d'intention est celui qui est traité par les modèles de contrôle moteur prédictif évoqués ci-dessus, permettant, à partir d'objectifs désirés, de spécifier des commandes motrices et de prédire les conséquences de leur exécution par l'appareil moteur. Le *monitoring* de ce dernier niveau met donc en œuvre un modèle interne direct, qui simule le comportement de l'appareil moteur, permettant de passer des commandes motrices aux mouvements (prédits), et un modèle interne inverse, qui représente la transition inverse, i.e. des mouvements (désirés) aux commandes motrices à exécuter.

Pacherie propose qu'un *monitoring* fonctionnant sur le même principe s'applique aux deux niveaux d'intention supérieurs, P- et D-intentions. Bien sûr, les modèles internes impliqués dans ces deux niveaux sont d'un tout autre ordre. Ils ne simulent pas la dynamique de l'appareil moteur. Au niveau des D-intentions, les modèles internes correspondent à des modèles des connaissances de l'agent sur le monde, les objets et les agents qui l'entourent ainsi que sur elle-lui-même. Au niveau des P-intentions, les modèles internes intègrent des informations conceptuelles et perceptives sur la situation courante de l'agent, sur l'objectif de l'action en cours ainsi que sur son contexte. Une première proposition de Pacherie qui me paraît particulièrement adaptée à la conception du *monitoring* que je défends, est que des modèles internes soient impliqués à chacun des trois niveaux, avec des modèles inverses qui, à partir d'objectifs désirés, calculent des moyens pour les atteindre (des commandes motrices dans le cas du dernier niveau) et des modèles directs qui prédisent les conséquences de la mise en œuvre de ces moyens. Le *monitoring* dédié à chaque niveau s'appuie donc sur un couplage entre modèles inverse et direct, comme cela a été proposé à l'origine, dans le cadre du contrôle moteur des actions, à partir d'intentions motrices concrètes. La deuxième proposition intéressante de Pacherie est que deux types de dynamique de *monitoring* s'appliquent : les dynamiques locales des transitions à l'intérieur de chaque niveau d'intention et une dynamique globale pour les transitions d'un niveau d'intention à l'autre. Le *monitoring* peut s'effectuer à la fois localement, en ajustant le modèle interne au niveau des M-intentions par exemple et en recalculant ainsi simplement des commandes motrices qui ne fourniraient pas le mouvement désiré. Mais il peut se faire aussi plus globalement, de deux façons. Tout d'abord le *monitoring* global se met en place en faisant remonter en cascade le résultat de la comparaison entre l'état final atteint effectivement et les sorties prédites à chacun des trois niveaux, de façon cumulative, pour ajuster les modèles directs à chacun des trois niveaux (lignes pointillées du graphe de la Figure 3 de Pacherie, 2008). Ensuite, il intervient à travers les comparaisons entre l'état final et les objectifs désirés à

chacun des trois niveaux pour ajuster les modèles inverses (ligne de rétroaction à l'extrémité droite du graphe de la Figure 3 de Pacherie, 2008). Cette proposition me paraît cependant comporter une faille. En effet, les états prédits à chacun des trois niveaux ne sont pas dans le même format. Au niveau des D-intentions, l'état prédit doit pouvoir être comparé à l'objectif général (« *overarching goal* » : rédiger mon HDR), alors que l'état prédit au niveau des P-intentions doit pouvoir être comparé au but instantané (« *instantaneous goal* » : taper sur le clavier). Ce dernier état prédit doit lui-même pouvoir être comparé à l'état effectif final (« *actual state* »). Ainsi l'état prédit au niveau des M-intentions est dans un format sensoriel et ne peut être dans le même format que l'état prédit au niveau des D-intentions. Il n'est donc pas concevable que l'état prédit au niveau des M-intentions puisse être comparé aux autres états prédits (traits pointillés du modèle). Il n'est pas non plus concevable que l'état effectif puisse être directement comparé aux deux états prédits de niveaux supérieurs, ni aux deux objectifs supérieurs. Les quatre premières comparaisons (en partant du haut) figurant sur ce modèle sont donc inappropriés. On peut cependant concevoir qu'un signal d'erreur (et non la comparaison) remonte lui-même en cascade et soit pris en compte à chaque niveau.

Je propose d'adapter ce modèle de MCP hiérarchique – incluant une double dynamique, globale et locale avec hiérarchie de l'action sur trois niveaux – à la production de la parole, avec les trois niveaux conçus par les psycholinguistes, soit conceptualisation, formulation (regroupant syntaxe, prosodie et phonologie) et articulation. Comme noté ci-dessus, il faudrait ajouter les gestes dans l'étape de formulation, pour tenir compte des langues des signes mais aussi pour les langues orales, souvent accompagnées de gestes ponctuant et complétant les énoncés (de Ruiter, 2007). Je propose en outre, dans cette adaptation, de corriger la faille concernant la remontée en cascade des signaux d'erreurs (de prédiction et de contrôle). Ce modèle adapté a été esquissé dans Lœvenbruck (2018) et présenté dans Grandchamp *et al.* (soumis). Il est schématisé sur la Figure I-12. Dans mon adaptation des propositions de Pacherie (2008) au cas de la parole, j'ai, à dessein, ajouté une boîte, celle de l'appareil articulatoire. Grâce à cet ajout, la dernière étape du modèle, celle de l'articulation, est conforme au modèle prédictif tel qu'il est habituellement formulé (cf. Figure I-9), avec une sortie prédite produite *avant* la sortie effective, compte tenu de l'inertie de l'appareil articulatoire physique. Toutefois, pour les étapes de niveaux hiérarchiques supérieurs, il n'est pas utile d'ajouter de boîte avec l'appareil physique, puisque ces étapes sont toutes mentales. Les modèles internes à ces deux niveaux ne sont donc pas des simulateurs *d'appareils physiques*. Ce sont des modèles (ou des procédures computationnelles) permettant de passer d'un type de représentation mentale (e.g. pour les modèles contrôleurs, intention communicative en entrée de l'étape de conceptualisation ou message pré-verbal en entrée de la formulation) à un autre (toujours pour les contrôleurs : message pré-verbal en sortie de l'étape de conceptualisation ou objectif phonétique en sortie de la formulation). Ces modèles internes permettent d'implémenter le mécanisme général de prédiction à l'œuvre dans le *monitoring* de la production de la parole.

Au niveau de l'articulation, j'ai montré ci-dessus comment le contrôle prédictif peut être implémenté, avec, d'une part, un modèle interne direct (ou « émulateur de l'appareil articulatoire ») fournissant un état sensoriel prédit à partir d'une copie des commandes motrices et, d'autre part, un modèle interne inverse, le « contrôleur », calculant une spécification motrice à partir d'un objectif phonétique (état supramodal désiré), spécification à son tour transformée en commandes motrices par le programmeur.

Au niveau de la formulation, le modèle interne direct (nommons-le « prédicteur » à la suite de Haruno *et al.*, 2003 ou Pacherie 2008) ne serait pas une simulation d'un appareil moteur mais implémenterait le lien entre un objectif phonétique spécifié et le message pré-verbal qui en découlerait, par prédiction (ou simulation). Le modèle inverse (nommons-le « contrôleur d'objectif phonétique ») permettrait de générer un objectif phonétique à partir d'un message pré-verbal désiré. Les messages pré-verbaux désirés et prédits pourraient ainsi être comparés et toute erreur commise lors de la génération de l'objectif phonétique pourrait être corrigée quasi-instantanément. Pacherie n'explique pas comment les signaux prédits sont générés avant les signaux effectifs. Je propose ici une solution à ce manque. Tout d'abord, au niveau de l'articulation, la sortie du modèle interne inverse (contrôleur) est fournie à l'appareil articulatoire, qui lui-même fournira des mouvements et des sons, avec un certain délai. Les commandes motrices peuvent donc être corrigées par le système de prédiction avant d'avoir été effectuées. De plus, rappelons qu'un programmeur, en sortie du contrôleur, coordonne les objectifs moteurs et permet l'agencement des commandes motrices. Ce programmeur pourrait aussi induire un délai permettant que les erreurs de formulation (et de conceptualisation) soient corrigées avant que la spécification motrice soit transformée en commandes motrices. Comme il a été mentionné plus haut, le niveau de la formulation pourrait être sous-divisé en encodages grammatical et phonologique. Dans ce cas, il faut le remplacer par deux sous-étages en cascade fonctionnant sur le même principe.

De la même manière, au niveau de la conceptualisation, on peut imaginer que le modèle inverse (nommons-le « contrôleur de message ») générerait un message pré-verbal désiré à partir d'une intention communicative désirée et que le prédicteur implémenterait le lien entre un message pré-verbal désiré et l'intention communicative qui en résulterait. Les intentions communicatives désirées et prédites pourraient ainsi être comparées et toute erreur commise lors de la génération du message pré-verbal par le modèle inverse pourrait être corrigée quasi-instantanément. Il n'est pas prévu, dans le modèle psycholinguistique de Levelt *et al.*, de mémoire-tampon permettant d'ajouter un délai avant l'étape de formulation. Il reste à déterminer si cette temporisation est nécessaire ou si l'étape de formulation peut être initiée sans délai, quitte à être interrompue (sans conséquence puisque les étapes ultérieures ne seraient pas encore exécutées) pour être corrigée. On peut imaginer que le programmeur soit suffisant pour temporiser (attendre que les *monitorings* des étapes de conceptualisation et de formulation soient achevés) avant l'envoi des commandes motrices aux articulateurs.

Avec cette architecture et ces paires de prédicteur et contrôleur, le MCP hiérarchique appliqué à la production de parole peut s'envisager comme suit.

Au niveau de l'articulation, l'état sensoriel effectif est comparé, d'une part, à l'état sensoriel prédit (supramodaux phonétique) et, d'autre part, à l'état phonétique désiré. Ces deux derniers états sont temporisés, avec un système de *buffer*, qui permet qu'ils soient synchrones avec l'état effectif (cf. les délais Δt sur la Figure I-12). La première comparaison fournit un signal d'erreur de prédiction, qui permet de corriger l'émulateur de l'appareil vocal (le modèle direct, prédicteur). La deuxième comparaison permet de vérifier que la génération de commandes motrices était correcte. En cas de non-conformité entre ces deux états, le signal d'erreur (de prédiction et de contrôle) permet d'ajuster le contrôleur. Ce signal d'erreur peut (cf. modèle HMOSAIC, Haruno *et al.*, 2003) être proportionnel à la différence entre les deux états. Comme expliqué précédemment, une troisième comparaison peut avoir lieu, toujours au niveau de l'articulation : l'état

phonétique prédit peut lui-même être comparé à l'objectif phonétique désiré. En cas de non adéquation entre ces deux états, le contrôleur peut être ajusté, avec un signal d'erreur de prédiction et de contrôle, calculé de façon similaire. Pour ne pas obscurcir le schéma, les flèches depuis les comparateurs vers le contrôleur et le prédicteur n'ont pas été mentionnées sur la Figure I-12. Elles sont explicitées sur la Figure I-9.

Au niveau de la formulation, le message pré-verbal prédit par la paire contrôleur - prédicteur de ce niveau est comparé au message pré-verbal désiré (auquel un délai Δt est ajouté, probablement minime). En cas de non adéquation entre les deux, le contrôleur d'objectif phonétique peut être ajusté (avec un signal d'erreur calculé à partir de l'écart entre les deux états) et le niveau inférieur est affecté.

Au niveau de la conceptualisation, l'intention communicative prédite par la paire contrôleur - prédicteur de ce niveau est comparée à l'intention communicative désirée (à laquelle un délai Δt est ajouté, comme pour la formulation). En cas de non adéquation, le contrôleur de message pré-verbal peut être ajusté (avec un signal d'erreur correspondant à l'écart entre les deux états) et les niveaux inférieurs sont affectés.

Un point de vue similaire est défendu dans le modèle de la production du langage de Pickering & Garrod (2013). Ces chercheurs affirment en effet que le système de *monitoring* par prédiction s'applique à chaque niveau de la production langagière : ainsi la sortie effective du niveau sémantique est comparée à sa prédiction, la sortie effective du niveau syntaxique est comparée à sa prédiction et la sortie effective du niveau phonologique à sa prédiction. Selon eux, les erreurs commises à chacun des niveaux peuvent ainsi être détectées et corrigées et permettent des productions fluides, avec peu d'erreur effective. Les auteurs ne détaillent pas la façon dont le *monitoring* est mis en place à chaque niveau, mais leur commentaire sur la redondance des différents systèmes de prédiction révèle une faille dans leur modèle. Selon moi, ils appliquent directement le modèle prédictif tel qu'il est habituellement formulé – avec à la fois un appareil moteur physique et un modèle interne simulateur de cet appareil physique – et ceci à chaque niveau de la hiérarchie. Cette implémentation est correcte pour l'articulation, puisqu'il existe un appareil articulatoire qui peut être simulé de façon interne, mais elle est inadaptée pour les autres niveaux, puisqu'il n'existe pas d'appareil physique. Pickering & Garrod s'en sortent en invoquant la notion de prédiction appauvrie :

« the predictions are not the same as the implemented production representations, but are easier-to-compute "impoverished" representations. They leave out (or simplify) many components of the implemented representations »¹⁶

Or, comme le commentent de Ruiters & Cummins (2013) ainsi que Strijkers, Runnqvist, Costa, & Holcomb (2013) et Hartsuiker (2013), la notion de prédiction appauvrie n'est pas compatible avec le fait que les prédictions sont comparées à des objectifs désirés, qui doivent eux-mêmes inclure des informations riches et détaillées pour que la production de langage corresponde à l'intention initiale.

¹⁶ « Les prédictions ne sont pas identiques aux représentations implémentées à chaque étape de la production mais sont des représentations appauvries plus faciles à calculer. Elles n'incluent pas (ou simplifient) de nombreuses caractéristiques des représentations implémentées »

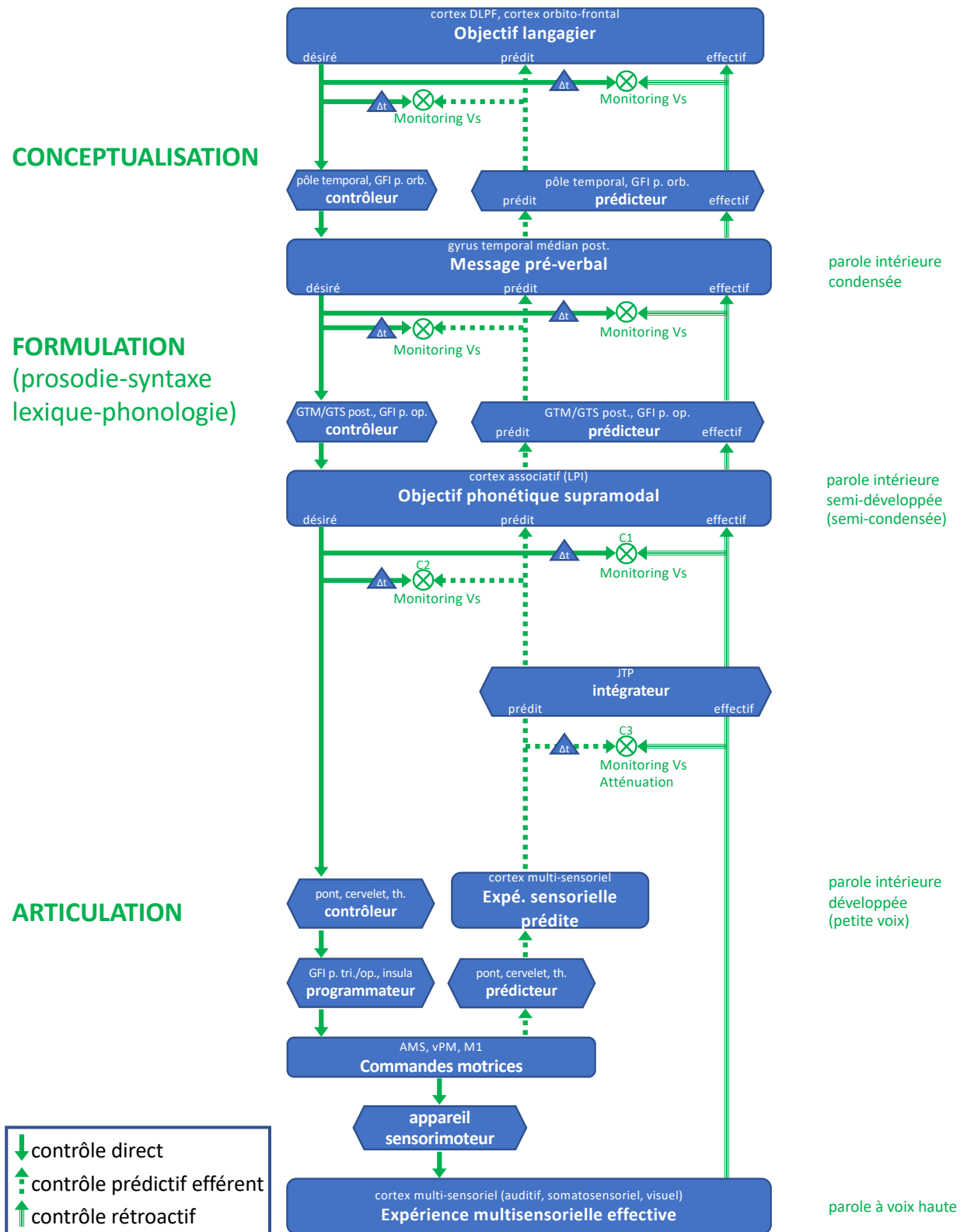


Figure I-12. Modèle hiérarchique du monitoring prédictif de la production de langage. Ce modèle est traduit et adapté de Lævenbruck (2018), inspiré de Haruno et al. (2003), Pacherie (2008) et Pickering & Garrod (2013), ainsi que du modèle de Duffau et al. (2014). Des mécanismes probabilistes sont en jeu dans la transmission des signaux d'erreur de prédiction et de contrôle, à différents niveaux. Les structures neurales (cérébrales et cérébelleuses) possiblement impliquées sont indiquées. Les faisceaux cortico-corticaux et sous-corticaux-corticaux ne sont pas indiqués pour ne pas obscurcir le schéma.

J'en reste donc sur cette adaptation de la proposition de Pacherie (2008), implémentant un *monitoring* global et local de la production du langage, compatible avec le modèle HMOSAIC de Haruno *et al.* (2003) et ne nécessitant pas d'ajouter la notion de prédiction appauvrie. Vicente & Martínez-Manrique (2016) font par ailleurs la même hypothèse en la justifiant théoriquement et en montrant qu'elle permet d'expliquer les liens entre pensée pré-verbale (*unsymbolized thinking*) et la parole intérieure, mais sans en donner une implémentation explicite. Je détaille donc ci-dessous l'implémentation.

En plus des *monitorings* propres à chaque niveau, une dynamique de *monitoring* globale, en cascade, peut être envisagée, comme Pacherie le propose pour les actions en général (mais en corrigeant la faille concernant les comparaisons de signaux qui ne sont pas dans le même format), en utilisant les comparaisons entre l'état final effectif et les sorties prédites et désirées à chaque niveau, pour ajuster les différents modèles directs et inverses. Toutefois, cela nécessite d'ajouter des mécanismes de « conversion » pour que les différents signaux à comparer soient dans le format adéquat et pour que l'erreur résultante puisse être prise en compte. Par exemple, si l'état sensoriel effectif n'est pas conforme à l'état sensoriel (objectif phonétique) désiré, faut-il, en plus du contrôleur de commandes motrices, ajuster le contrôleur d'objectif phonétique ? Et dans ce cas quel ajustement apporter ? Dans les modèles de contrôle classiques, c'est l'écart entre les deux états qui est utilisé pour corriger le modèle défaillant (avec par exemple un filtre de Kalman, cf. Grush, 2004). Mais si le modèle à corriger est situé à un niveau supérieur dans la hiérarchie, alors cet écart n'a pas de sens. Il faut donc le convertir, ce qui nécessite un mécanisme *ad-hoc*, qui reste à expliciter.

Encadré I.20. *Monitoring localisé à chaque niveau ou monitoring global ?*

Ainsi, une question que je laisse en suspens est celle de savoir si l'état sensoriel effectif est vérifié par rapport à l'intention communicative initiale, en d'autres termes si la comparaison se fait directement du niveau inférieur au niveau supérieur ou si les comparaisons à chaque niveau sont suffisantes pour détecter et corriger les erreurs éventuelles des différents contrôleurs. Dans le modèle actuel, je n'indique pas de remontée directe du niveau sensoriel effectif au niveau de l'objectif initial.

Levelt (1989) soutient qu'il existe une boucle (« *auditory loop* ») faisant intervenir le système de compréhension de la parole, et permettant de remonter directement du signal auditif à l'intention communicative pour éventuellement corriger une production défaillante. Dans le modèle de Pacherie (2008), il est prévu une remontée directe du niveau de l'état effectif vers chacun des niveaux supérieurs, dont l'objectif global. Dans le modèle HMOSAIC de Haruno *et al.* (2003), il n'est pas clairement décrit comment s'organisent les flux d'information entre les différents niveaux de la hiérarchie. Il est précisé (p. 581) que seul le niveau le plus bas prédit et contrôle la dynamique de l'appareil moteur et de l'environnement. Les niveaux supérieurs ne prédisent et contrôlent que la dynamique de leurs niveaux immédiatement subordonnés. Il semble donc qu'il y ait des échanges en cascade d'un niveau à l'autre, mais pas de flux direct du niveau le plus bas au niveau le plus élevé de la hiérarchie. Ce que je propose provisoirement ici (pas de remontée directe du niveau sensoriel effectif vers l'intention communicative) est donc en accord avec le modèle HMOSAIC sur ce point. De même, dans le modèle de Pickering & Garrod (2013), la façon dont les différents niveaux linguistiques sont agencés n'est pas non plus complètement spécifiée. Toutefois, il est précisé (p. 10) que le *monitoring* se fait

séparément à chaque niveau, chacun des percepts sémantique, syntaxique et phonologique étant comparé à sa propre prédiction, avec dans l'ordre la sémantique puis la syntaxe puis la phonologie. Ce que je propose, provisoirement, en l'état des connaissances, est donc également en accord (sur ce point) avec le modèle de Pickering & Garrod (2013). Mais on peut envisager qu'un signal d'erreur (de prédiction et de contrôle) résultant de la comparaison entre état sensoriel effectif et objectif phonétique désiré soit remonté aux niveaux supérieurs pour pondérer le contrôleur d'objectif phonétique ainsi que le contrôleur de message pré-verbal.

Une deuxième question que je laisse en suspens est celle de l'éventuel ajustement des différents prédicteurs. Au niveau le plus bas de la hiérarchie, le prédicteur est en fait un émulateur de l'appareil vocal, un véritable modèle interne au sens classique du contrôle moteur, qui peut être ajusté en fonction de signaux physiques (écart entre l'état sensoriel effectif et l'état sensoriel prédit). Aux niveaux supérieurs, quels peuvent être les critères selon lesquels ces prédicteurs devraient être ajustés ? Quelles sont les comparaisons à effectuer ? Au niveau de la formulation, le prédicteur prédit un message pré-verbal. A quel message pré-verbal effectif faut-il le comparer ? Il n'y a pas de signal externe effectif qui corresponde à ce signal interne. Dans le modèle HMOSAIC, les modèles prédictifs (ou prédicteurs) des niveaux supérieurs sont ajustés non pas en fonction d'un *feedback* sensoriel (qui n'existe qu'au niveau le plus bas de la hiérarchie) mais en fonction de la façon dont le niveau subordonné a été efficient. Dans notre cas, on peut donc envisager qu'un signal d'erreur remontant de la comparaison entre objectif phonétique prédit et objectif phonétique désiré soit pris en compte pour pondérer le prédicteur du message pré-verbal. De même, au niveau de la conceptualisation, on peut envisager qu'un signal d'erreur remontant de la comparaison entre message pré-verbal prédit et message pré-verbal désiré pondère le prédicteur d'intention communicative. Ces signaux de correction d'un niveau à l'autre, pourraient d'ailleurs s'appliquer également aux contrôleurs (pas uniquement aux prédicteurs).

Encadré I.21. Ajustement probabiliste des prédicteurs aux différents niveaux ?

Ainsi, une deuxième question ouverte est celle de l'ajustement des différents prédicteurs, en l'absence de signaux sensoriels de référence. Comme suggéré par Grush (2004) ou par Haruno *et al.* (2003) dans le cas du contrôle de l'action en général, la façon dont le signal d'erreur est transmis à un modèle ou un autre pourrait dépendre de façon probabiliste de la confiance que les locuteurs ont envers leurs différents contrôleurs (de commandes motrices, d'objectif phonétique et de message pré-verbal) et prédicteurs (d'objectif phonétique, de message pré-verbal et d'intention communicative).

Une autre question en suspens concerne le déroulement temporel des différentes transmissions de signaux d'un niveau à l'autre. Certains transferts d'information utilisant des modèles de contrôle et des prédicteurs peuvent se faire de façon quasi-instantanée, les modèles pouvant être considérés (comme dans Haruno *et al.* 2003 par exemple) comme des réseaux de neurones dont les poids sont ajustés au fur et à mesure de l'apprentissage et de l'expérience du locuteur. Mais, comme expliqué ci-dessus, certains échanges d'informations, notamment entre les niveaux abstraits et le niveau physique, nécessitent une temporisation, un *buffer*, pour que les commandes puissent être prises en compte à bon escient et au bon moment par l'appareil moteur physique.

Encadré I.22. Alignement temporel des signaux à comparer ?

Ainsi, une troisième question sans réponse est celle de l'alignement temporel des signaux à comparer, entre états désirés, prédits et effectifs. Ces temporisations sont nécessaires, mais il reste à expliquer comment elles sont mises en œuvre, d'autant plus que nous verrons plus loin (I.2.3) qu'elles sont cruciales dans l'attribution de l'agentivité.

Une dernière question ouverte est celle des réseaux cérébraux impliqués. Les réseaux associés à l'étape d'articulation ont été explicités ci-dessus (Figure I-9, Figure I-10, Figure I-11). Comme expliqué ci-dessus, les données expérimentales dont nous disposons, suggèrent que l'état phonétique supramodal désiré, au dernier niveau de la hiérarchie (l'articulation), correspond à des activations du cortex associatif (lobule pariétal inférieur), tout comme les états supramodaux effectif et prédit (après intégration des états multisensoriels). Les états multisensoriels prédit et effectif correspondent à des activations des cortex sensoriels (auditif, somatosensoriel et peut-être visuel). Comment ces différents états sont-ils gérés au plan cérébral ? Dans les cortex multisensoriels, les états effectifs et prédits sont générés à des temps différents, il n'y a donc pas lieu de les distinguer. Dans le cortex associatif, l'état supramodal désiré est généré avant l'état prédit, lui-même avant l'état effectif, or ces états doivent être comparés pour le *monitoring* (les erreurs éventuelles étant transmises au contrôleur, cervelet). Des mécanismes de temporisation (avec des équivalents neuronaux des filtres de Kalman) peuvent donc être envisagés pour synchroniser ces signaux. Mais cela reste à décrire. Et la localisation des comparateurs reste à préciser.

S'il est possible de faire quelques spéculations sur les patrons d'activations cérébrales associés au niveau de l'articulation (cf. ci-dessus, p.67), qu'en est-il des réseaux cérébraux impliqués aux autres niveaux ?

Les travaux de recherche des vingt dernières années convergent vers l'idée que la conception traditionnelle modulaire de la production du langage doit être abandonnée (voir par exemple Pickering & Garrod, 2014 PNAS). Il existe peu de modèles neurocomputationnels exhaustifs de la production du langage, depuis la conceptualisation d'un message jusqu'à son articulation. Parmi les plus complets, le modèle de Duffau *et al.* (2014) défend une architecture neurale dynamique, avec une conception hodotopique, contre la vision classique modulaire, localisationniste ou topologique. Selon Duffau (2008, 2013), le langage résulte en effet de traitements parallèles, effectués par différents groupes de neurones distribués, connectés et synchronisés, plutôt que par des centres localisés et modulaires. Ces sous-réseaux parallèles distribués à large échelle interagissent ensemble et peuvent se compenser après une lésion cérébrale. Les voies de connexions entre ces groupes de neurones jouent donc un rôle essentiel, d'où l'introduction du terme de « hodotopie » (du grec : *hodos* : voies, *topos* : localisation). Duffau et collègues (2014) ont posé des hypothèses précises, fondées sur des données de patients cérébro-lésés ainsi que sur des données issues de la stimulation électrique directe chez des patients éveillés. Une cartographie corticale et sous-corticale des interactions entre les sous-traitements articulatoires, phonologiques, syntaxiques et sémantiques est proposée. Dans ce modèle, les réseaux sont de deux types : des régions corticales interconnectées par des faisceaux cortico-corticaux (connectivité « horizontale ») et des structures sous-corticales modulatrices (telles que les noyaux gris centraux) elles-mêmes connectées au cortex par des fibres

« verticales ». Une même aire peut être impliquée dans différents sous-réseaux et donc sous-fonctions. Selon Duffau (2013 ; Duffau *et al.*, 2014), le réseau de la production de la parole (autrement dit, l'étape d'articulation) inclut le cortex prémoteur ventral (partie latérale du gyrus précentral), le gyrus supramarginal, la partie latérale du faisceau longitudinal supérieur (connectant ces deux aires corticales entre elles), l'insula antérieure et le putamen (reliés au cortex prémoteur ventral et supramarginal par des fibres verticales). Le traitement phonologique impliquerait la partie profonde du faisceau longitudinal supérieur (ou faisceau arqué). Le traitement syntaxique activerait le gyrus frontal inférieur gauche, la partie postérieure du lobe temporal (gyri supérieur et moyen) et le faisceau longitudinal supérieur. Le traitement sémantique serait sous-tendu par deux sous-réseaux direct et indirect. La voie directe met en jeu le faisceau occipito-frontal inférieur, la voie indirecte le faisceau longitudinal inférieur, avec un relais au niveau du pôle temporal, avant la connexion avec les régions orbito-frontales via le faisceau unciné (Moritz-Gasser, Herbet & Duffau, 2013, 2015).

Si des recherches sont encore nécessaires pour mieux décrire comment se déroule l'intégration de ces différents sous-traitements, leurs corrélats neuroanatomiques et leur éventuelle hiérarchisation, on peut donc faire quelques conjectures, fondées sur des travaux ou des revues de la littérature antérieurs (notamment Baciú *et al.*, 1999 ; 2005 ; Baciú & Perrone-Bertolotti, 2015 ; Caplan *et al.*, 2000 ; Duffau *et al.*, 2014 ; Gernsbacher & Kaschak, 2003 ; Guenther & Vladusich, 2012 ; Haller *et al.*, 2005 ; Hickok, 2009 ; Indefrey *et al.*, 2001 ; Indefrey & Levelt, 2004 ; Kell *et al.*, 2011 ; Keller & Kell, 2016 ; Llorens *et al.*, 2011 ; Lœvenbruck *et al.*, 2005 ; Price, 2012 ; Rauschecker & Scott, 2009 ; Tian & Poeppel, 2013 ; Tremblay & Dick, 2016). Dans Grandchamp *et al.* (soumis), nous avons proposé l'implémentation neuronale hiérarchique suivante, mentionnée dans la Figure I-12.

Le niveau de la conceptualisation implique le cortex dorsolatéral préfrontal, le cortex orbito-frontal et le pôle temporal et le gyrus temporal médian postérieur, les connexions entre les régions frontales et temporales se faisant via une voie ventrale impliquant le faisceau occipito-frontal inférieur (non mentionné pour ne pas obscurcir le schéma).

La formulation syntaxico-phonologique recrute la partie postérieure du lobe temporal (gyri supérieur et moyen) et le gyrus frontal inférieur (GFI) gauche, avec des connexions dorsales, via le faisceau longitudinal supérieur et le faisceau arqué. A ce stade, je fais l'hypothèse que le lobule pariétal inférieur (LPI) gauche est également activé pour former l'objectif phonétique supramodal. Cette hypothèse est compatible avec le rôle associatif reconnu du LPI. Mais il reste encore une question ouverte cependant, celle de savoir si l'activation du GFI gauche est en amont de celle du LPI ou si le LPI fournit lui-même des informations efférentes au GFI. J'opte actuellement pour la première solution.

Comme décrit plus haut, l'articulation met en œuvre un transfert d'information, depuis le LPI (ou le GFI si les connexions entre LPI et GFI sont dans l'ordre inverse) vers le cervelet (probablement le lobule antérieur), via le pont. Une spécification motrice est envoyée aux régions frontales gauche via le thalamus. Des programmes moteurs sont alors émis, en coordonnant la spécification motrice issue du cervelet, avec les actions de parole en cours, ce qui correspond probablement à l'activation du GFI gauche, de l'insula, puis du cortex prémoteur ventral, de l'aire motrice supplémentaire et du cortex moteur primaire (via le faisceau frontal oblique). Il existe des arguments pour l'hypothèse que l'activation du GFI précède celle du cortex prémoteur ventral (cf. e.g. l'étude ECoG de la production

de la parole de Flinker *et al.*, 2015, citée plus haut) et que le LPI (gyrus supramarginal) a des connexions efférentes vers le cortex prémoteur ventral, via la partie antérieure du faisceau longitudinal supérieur (Duffau *et al.*, 2014). Il existe aussi des données en faveur de connexions depuis le LPI vers le cervelet (Imamizu & Kawato, 2009 ; Miall, 2003) ainsi que depuis le cervelet vers les régions motrices et prémotrices frontales, incluant probablement le GFI (Imamizu & Kawato, 2009 ; Murdoch, 2010). Il reste cependant à éclaircir si la connexion pariéto-frontale directe (ne passant pas par le cervelet) est associée avec le niveau de l'articulation ou si elle ne concerne que le niveau de la formulation. S'il s'avère qu'elle concerne le niveau de l'articulation, alors il faudra ajouter à ce modèle une flèche de connexion directe supplémentaire, depuis l'objectif phonétique supramodal jusqu'au niveau de l'articulation (sans passer par le cervelet). Quoi qu'il en soit, les commandes motrices qui résultent de la coordination effectuée dans les régions frontales inférieures sont envoyées à l'appareil moteur. Une copie de ces commandes est envoyée, via le pont, au cervelet (probablement le lobe postérieur). Il en résulte une prédiction multisensorielle transmise, via le thalamus au cortex auditif (gyrus temporal supérieur) et somatosensoriel (gyrus post-central). Je fais la spéculation que ces réponses auditive et somatosensorielle sont intégrées en une représentation supramodale, via la jonction temporo-pariétale. La prédiction phonétique supramodale résultante peut ainsi être comparée à l'objectif phonétique au niveau du LPI, le lieu du *monitoring* de cette étape articulatoire.

Ainsi, dans ce modèle, il est proposé que le GFi est impliqué à deux niveaux. Je m'appuie sur des données de neuroimagerie pour proposer qu'il joue un rôle dans la formulation, notamment dans le suivi des rôles thématiques (qui-fait-quoi-à-qui), suivi crucial pour les traitements morphosyntaxiques (Caplan & Hanna, 1998 ; Caplan *et al.*, 2000 ; Indefrey *et al.*, 2001 ; Løevenbruck *et al.*, 2005). D'autres données indiquent qu'il est également à l'œuvre dans la coordination et le séquençement des gestes articulatoires, lors de l'étape d'articulation (Blank *et al.*, 2002 ; Indefrey & Levelt, 2004).

Je préfère éviter d'indiquer les fibres de connexion pour ne pas obscurcir le schéma. Il reste à mieux comprendre et décrire comment se font justement les connexions, et leur agencement temporel (quels sont les traitements en parallèle, quels sont les traitements en séquence, quels sont les retours ou les boucles).

Pour conclure, ma proposition répond aux cinq contraintes listées précédemment. Grâce aux mécanismes de *monitoring* locaux, les erreurs de production peuvent être détectées et corrigées rapidement, même avant la fin de l'énonciation. La production de la parole peut être fluide et rapide car elle s'appuie sur des prédictions et non sur des boucles de rétroactions longues. Les erreurs sont rares car les mécanismes de vérification s'appliquent à plusieurs niveaux. La production de la parole peut se faire avec un système de compréhension déficient car la supervision n'implique pas tout le système de compréhension, simplement des prédicteurs locaux. La réparation des erreurs peut se faire en reprenant une partie de l'énoncé, par remontée en cascade jusqu'aux niveaux supérieurs de la hiérarchie ou en poursuivant directement avec la correction, après une halte, par correction locale à un niveau donné.

Encadré 1.23. $M_S(v_S)$: monitoring verbal égocentré

Pour distinguer ce sous-système de *monitoring* des autres sous-systèmes impliqués dans le langage, j'utiliserai par la suite (cf. 1.2.4) le terme de *monitoring par soi-même des verbalisations auto-produites* ou $M_S(v_S)$, plus simplement *monitoring verbal égocentré*.

1.2.1.2. La perception du langage et le *monitoring verbal allocentré*

Sur le versant perception et compréhension, lorsqu'en tant qu'auditeur·rice, nous percevons le langage d'autrui, nous mettons également en œuvre des mécanismes de *monitoring* des productions verbales d'autrui, à partir de nos connaissances sur le monde, sur nos interlocuteurs et sur le langage, tout en tenant compte des signaux de l'environnement. Ces mécanismes nous permettent d'ajuster notre interprétation du message d'autrui au fur et à mesure de son énonciation. Avant de détailler les dispositifs de *monitoring verbal* impliqués dans la perception du langage, examinons quels sont les différents niveaux de traitements impliqués.

1.2.1.2.1. Les différents niveaux de traitement impliqués dans la perception du langage

Comme cela a été esquissé ci-dessus (cf. introduction de la section 1.2.1), il est généralement admis que la réception du langage (oral ou signé) comprend trois processus : le décodage de l'entrée sensorielle (signaux auditifs et visuels) pour identifier les sons et gestes produits par l'interlocuteur·rice, la reconnaissance d'unités lexicales et l'analyse de la structure syntaxique, prosodique et pragmatique (*parsing*) et l'interprétation sémantique (cf. e.g. François & Nespoulous, 2011). Les spécialistes de la parole se sont attachés à décrire la perception et l'identification des sons du langage, les phonèmes ou les syllabes isolées, en rendant compte au mieux des faits observés. Les psycholinguistes ont, eux, privilégié la description de la reconnaissance des mots dans les énoncés et la compréhension de phrases. Ce n'est que récemment que des modèles intégratifs de la perception du langage incluant des modélisations de ces trois processus ont été proposés (e.g. Friederici, 2012 ; Hickok & Poeppel, 2007 ; Skeide & Friederici, 2016).

En ce qui concerne la perception des phonèmes, l'un des faits à expliquer concerne la façon dont les auditeur·rice·s gèrent la coarticulation. Dans la parole continue, il n'existe pas de frontière entre les mots ni les phonèmes. Il n'existe pas non plus de correspondance unique entre un signal acoustique et un phonème donné. Dans la séquence « il y a ici », un·e locuteur·rice qui doit gérer l'enchaînement des voyelles [i a i] utilisera des stratégies d'économie articulatoires pour éviter de trop déplacer la langue, qui doit être antérieure et élevée pour [i] et centrale et abaissée pour [a]. La voyelle [a] médiane sera donc teintée de [i] et le signal acoustique correspondant au [a] sera finalement proche de celui d'un [ɛ], ce que l'on nomme la « réduction vocalique ». Pourtant les auditeurs, sont capables de retrouver dans ce signal acoustique modifié des traces du [a] visé par le·la locuteur·rice (cf. Løevenbruck, 1996). Cette influence d'un phonème sur un autre, la coarticulation, peut se faire de façon progressive (anticipatoire) ou régressive. Ainsi, compte tenu des nombreuses possibilités combinatoires d'associer un phonème à un autre, il existe de multiples signaux acoustiques correspondant au même phonème. Il est donc impossible d'identifier directement dans le signal acoustique les phonèmes prononcés. Certains modèles de perception du langage proposent donc

que l'unité sur laquelle repose la reconnaissance des énoncés soit la syllabe, le diphone, le triphone ou même le mot, plutôt que le phonème. Toutefois, la coarticulation se fait parfois sur un empan large, plus large que la syllabe et même le mot (e.g. West, 2000).

Une autre source de difficulté dans la reconnaissance des unités linguistiques à partir du signal acoustique est la variabilité interlocuteur. La taille des conduits vocaux est extrêmement variable d'un-e locuteur-riche à l'autre, en fonction du sexe et de l'âge, ce qui induit notamment des modifications majeures des fréquences de résonance du conduit et, partant, du signal acoustique produit lors de l'articulation. Il a donc été proposé que les auditeurs sont capables au cours du processus de perception, de normaliser le signal acoustique pour y retrouver le message linguistique standardisé, idéalisé. Différentes théories de normalisation ont été proposées, intrinsèques et extrinsèques (pour des revues voir Løevenbruck, 1996 ; voir aussi Apostol *et al.*, 2004). Les théories de normalisation intrinsèque suggèrent que les auditeurs s'appuient sur une représentation paramétrique adéquate des propriétés spectrales propres à chaque son. Il est suggéré qu'avec certaines transformations sur la fréquence fondamentale et les formants (pics spectraux), on peut par exemple identifier les voyelles. Les théories de normalisation extrinsèque supposent que les auditeurs sont capables de rétablir un cadre de classification des phonèmes (voyelles notamment) à partir d'informations disséminées sur l'ensemble des productions d'un-e locuteur-riche : la qualité phonétique d'une voyelle pourrait ainsi être inférée à partir de la relation entre les valeurs formantiques de cette voyelle et celles des autres voyelles produites. Les auditeurs ne porteraient donc pas leur attention uniquement sur les valeurs formantiques d'une voyelle donnée, mais en s'appuyant sur les relations entre ces formants et la gamme de fréquence caractéristique du-de la locuteur-riche, ils pourraient recalibrer l'ensemble des productions du-de la locuteur-riche et identifier ainsi les phonèmes produits.

Les auditeurs seraient donc capables de soustraire la variabilité acoustique liée aux informations indexicales de sexe, de taille du conduit vocal et d'âge, par exemple, pour reconnaître le message transmis. Il a été suggéré que les auditeurs font mieux que soustraire ce type d'information, ils exploitent tous types d'informations indexicales sur l'identité et la condition du-de la locuteur-riche et ces informations interagissent avec les processus mnésiques pour guider la perception (Johnson & Mullennix, 1997, Cleary & Pisoni, 2001). La perception du langage consiste donc à extraire des informations d'ordre linguistique sur le message lui-même, mais également des informations indexicales sur le-la locuteur-riche. On voit donc ici que le système de *monitoring* verbal, qui nous permet de communiquer, joue également un rôle dans la construction de la représentation d'autrui, de l'altérité (cf. I.2.4).

Une autre source de difficulté pour retrouver les unités linguistiques dans le signal acoustique est la variabilité intralocuteur, liée aux différents styles de parole. Si le but essentiel de la communication est bien de faire percevoir un message, il existe toutefois une exigence qui lui est opposée, celle de minimiser l'effort de la part des locuteurs. Ainsi l'objectif acoustique sera plus ou moins parfaitement atteint car les locuteurs sont capables de sur- ou sous-articuler, en fonction des contraintes de la situation (Lindblom, 1988). L'espace acoustique couvert par les voyelles par exemple peut par exemple rétrécir en parole spontanée « décontractée » (*casual speech*), que Lindblom a dénommé *hypospeech*, et au contraire s'étendre lors d'une diction claire et soignée (*clear speech* ou *hyperspeech*) pour un auditoire non spécialiste ou dans un environnement bruyant. Les contrastes de durée entre voyelles (notamment en anglais) ont tendance à s'estomper

en *hypospeech*. Les contrastes sur le délai d'établissement du voisement entre consonnes voisées et non voisées sont plus prononcés en *hyperspeech*. Cette variabilité phonétique est de nature systématique et peut se définir selon un axe hyper/hypo, qui s'ajuste en fonction de la situation communicationnelle (Lindblom, 1988). Comme expliqué ci-dessus, le signal de parole n'est pas traité de façon isolée : des connaissances supplémentaires sur le·la locuteur·rice, mais aussi sur la situation, lui sont ajoutées par l'auditeur·rice et sont présumées par le·la locuteur·rice. Comme le résume Lindblom (1988), les locuteurs adaptent leur énonciation aux conditions ambiantes :

« *Intra-speaker variation is genuine and arises as a consequence of the speaker's adaptation to his judgment of the need of the situation* ».

Les locuteurs jouent de la variabilité pour optimiser leur production ou pour communiquer une information supra-linguistique (niveau du discours, emphase, émotion), dans le cadre d'une négociation locuteurs-auditeurs assumée.

Il est donc crucial de mieux comprendre les stratégies utilisées par les auditeurs pour passer du signal acoustique variable au phonème (ou à la syllabe, au mot) et de nombreux débats ont lieu dans ce domaine. Pour résumer, deux types principaux de théories s'opposent. Les théories « auditives » – pour lesquelles la perception des phonèmes se fait sur la base d'apprentissage de processus de catégorisation exploitant des connaissances statistiques sur des séquences sonores (Diehl *et al.*, 2004 ; Stevens, 1989) – et les théories « motrices » selon lesquelles l'auditeur·rice dispose de moyens pour retrouver les gestes articulatoires à l'origine des sons (Liberman, Cooper, Harris & MacNeilage, 1962 ; Liberman & Mattingly, 1985 ; Fowler, 1986 ; Best, 1995). Il existe des données en faveur de l'un et l'autre type de théorie. Différentes théories hybrides ont donc été proposées. Parmi celle-ci, la théorie unificatrice nommée *PACT (Perception-for-Action-Control Theory)*, développée par Schwartz et collègues (Schwartz, Abry, Boë & Cathiard, 2002 ; Schwartz, Boë, & Abry, 2007 ; Schwartz *et al.*, 2012), permet de rendre compte au mieux de l'implication de processus perceptivo-moteurs dans la perception de la parole. Dans la PACT, les unités d'échange entre auditeurs et locuteurs ne sont ni purement gestuelles (comme dans la Théorie Motrice de Liberman et collègues), ni purement auditives ou perceptives (comme dans les théories auditives), elles ont des propriétés à la fois perceptives et motrices. Ce sont des gestes façonnés par la perception ou autrement dit des percepts structurés par l'articulation. Percevoir consisterait donc à récupérer les contrôles moteurs mis en forme par l'audition et la vision. Dans le cas de la réduction vocalique, cité ci-dessus, bien que la cible acoustique visée par le·la locuteur·rice ne soit pas atteinte (du fait du contexte d'élocution), l'auditeur·rice utilise ses connaissances motrices et se sert de la trajectoire acoustique pour inférer la dynamique du geste articulatoire et la retrouver. La PACT propose aussi que les représentations perceptives et motrices sont co-structurées au cours du développement de l'enfant et restent associées dans le comportement adulte.

En ce qui concerne les niveaux supérieurs de la perception du langage (plus larges que le phonème ou la syllabe), notamment la reconnaissance des mots dans l'énoncé et la compréhension de phrases, différents modèles psycholinguistiques ont été développés.

Concernant la reconnaissance de mots, le modèle princeps est le *Modèle Logogen* de Morton (1969), qui s'applique à la fois à la lecture et à l'écoute de parole et qui introduit la notion clé d'*activation*, utilisée dans de nombreux modèles ultérieurs. Ainsi, dans ce modèle, lors de l'écoute d'un signal de parole ou lors de la lecture d'un texte écrit, il est

supposé que plusieurs mots du lexique mental puissent être activés, disponibles, en parallèle. Plus précisément, chaque entrée lexicale est associée à une unité de reconnaissance, appelée logogen, qui définit les propriétés visuelles (orthographiques), auditives et sémantiques du mot. Lors du processus de reconnaissance de mots, différents logogens sont activés par les informations visuelles et auditives liées aux stimuli entrants ainsi que par les informations sémantiques liées au contexte. Lorsque le niveau d'activation d'un logogen atteint un seuil suffisant, l'identification du mot correspondant a lieu. Le modèle permet de rendre compte de l'effet de contexte, les données expérimentales montrant que la reconnaissance d'un mot peut être facilitée par son contexte (par le début de la phrase par exemple). De plus, dans le modèle, le seuil d'activation dépend de la fréquence lexicale : il est d'autant plus bas que le mot apparaît avec une fréquence élevée dans la langue. Les mots les plus fréquents nécessitent moins d'information en entrée pour atteindre leur seuil. Ceci rend compte des données expérimentales qui indiquent que les mots fréquents sont identifiés plus vite que les mots moins fréquents. Enfin, le niveau d'activation décline après la reconnaissance d'un mot, mais graduellement, ce qui est en accord avec les données suggérant que les mots répétés sont reconnus plus rapidement.

Les modèles ultérieurs de la reconnaissance de mot sont souvent inspirés du Modèle Logogen, en considérant que les activations ou les traitements ont lieu en parallèle plutôt qu'en série (comme dans le modèle *Search* de Forster, 1976), mais s'en démarquent en s'appliquant soit à la lecture de textes écrits, soit à l'écoute de parole. Ces modèles de reconnaissance de mots parlés plus récents prennent ainsi mieux en compte la nature temporelle du signal de parole, en intégrant ses propriétés à la fois transitoires (contrairement à l'écrit, le signal de parole est évanescent) et continues (il n'y a pas systématiquement d'espace entre les mots parlés). Parmi ces modèles, les plus connus sont le modèle *Cohort* de Marslen-Wilson (1987) – à architecture autonome (s'appuyant uniquement sur le signal de parole), distribuée et non-localiste –, le modèle *NAM* ou *Neighborhood Activation Model* de Luce & Pisoni (1998) – dans lequel les mots qui partagent des traits phonologiques sont stockés mentalement dans des voisinages (*neighborhood*) –, le modèle connexionniste *TRACE* de McClelland & Elman (1986) – dans lequel les différentes unités de la parole (les traits phonologiques, les phonèmes, les mots) sont représentés à différents couches, et dans lequel les connexions d'une couche à l'autre sont bi-directionnelles – le modèle *Shortlist* de Norris (1994) – qui suppose deux étapes de traitement, la première consistant en l'activation de listes de candidats potentiels (comme dans le modèle *Cohort*) et la seconde correspondant à une compétition entre ces candidats, avec des inhibitions latérales entre les mots et des boucles de rétroaction (comme dans *TRACE*), le modèle *MERGE* de Norris, McQueen & Cutler (2000) – inscrit dans le cadre général de *Shortlist* et modélisant l'étape de décision phonémique sans rétroaction du lexique sur les niveaux de traitement inférieur – et les versions bayésiennes *Shortlist B* et *MERGE B* (Norris & McQueen, 2008). D'après Weber & Scharenborg (2012), les modèles récents de la reconnaissance de mot s'accordent sur trois points essentiels : premièrement, lorsqu'un mot est entendu, plusieurs candidats sont activés en parallèle, deuxièmement l'activation des différents candidats varie en fonction de la correspondance entre le signal de parole et les représentations lexicales mentales et troisièmement, les différents candidats sont en compétition les uns avec les autres. Mais il n'existe pas encore de consensus sur le flux d'information entre les différents niveaux de traitement (avec ou sans rétroaction, ascendant ou descendant),

ou sur le format des représentations mentales pré-lexicales et lexicales (abstractions ou exemplaires épisodiques).

Ces modèles s'appliquent essentiellement à la reconnaissance de la forme phonologique des mots. Il existe aussi des modèles de la sémantique lexicale, qui traitent spécifiquement de la reconnaissance du sens des mots et non de leur forme, tels que, par exemple, le modèle à réseau hiérarchique de Collins & Quillian (1969), le modèle à traits sémantiques de Smith, Shoben et Rips (1974), le modèle à diffusion de l'activation de Collins & Loftus (1975) ou le modèle ACT de Anderson (1996).

En ce qui concerne les niveaux de traitements à plus large échelle, et en particulier, le niveau de la compréhension de phrases, Selon Mitchell (1994), deux types principaux de modèles ont été proposés pour rendre compte des données comportementales sur la compréhension de phrases. Les premiers ont été des modèles sériels (ou séquentiels), inspirés de la grammaire générative, plaçant le décodage syntaxique comme étape première et autonome (Fodor, Bever & Garrett, 1974 ; Frazier & Rayner, 1982). Dans ce type de modèle, l'auditeur-riche commence par construire une structure syntaxique (un arbre) simple, à partir des informations sur la catégorie syntaxique des mots décodés, indépendamment des informations lexicales sémantiques. Ce n'est que dans une deuxième étape que les informations sémantiques sont prises en compte et que l'assignation des rôles thématiques (qui fait quoi à qui) a lieu. Si les structures syntaxique et sémantique sont incompatibles alors une ré-analyse a lieu. C'est le cas des phrases de type « *garden path*¹⁷ », qui conduisent à une interprétation initiale erronée et à une impasse sémantique qui conduit à la ré-analyse comme dans l'énoncé :

« Le secrétaire a posé le dossier sur les piles sur le bureau vide. »

Ce type de modèle ne permettant pas de rendre compte du fait que la construction de la structure syntaxique initiale dépend par exemple de la plausibilité sémantique associée au verbe principal, et étant incompatible avec la très grande rapidité du traitement langagier, un deuxième type de modèle a été introduit, les modèles à traitement interactif ou parallèle (ou « à satisfaction de contraintes »). Dans ces modèles, de multiples interprétations syntaxiques sont construites en parallèles, avec des poids tenant compte par exemple de la fréquence ou de la plausibilité sémantique de certaines constructions (Marslen-Wilson & Tyler, 1980 ; Altmann & Steedmann, 1988 ; Tanenhaus & Trueswell, 1995). Dans ces modèles, les processus syntaxiques et sémantiques interagissent dès les premières étapes du traitement d'un énoncé, de façon incrémentale. D'autres travaux, comme ceux de Warren *et al.* (1995) ont montré que les processus prosodiques interviennent également de façon très précoce.

Les études en neuroimagerie, notamment les études de potentiels évoqués, ont apporté de nouvelles données qui ont permis de mieux décrire le déroulement du processus de compréhension de la parole. De nouveaux modèles neuro-psycholinguistiques ont vu le jour, décrivant les traitements en jeu aux niveaux computationnel et cérébral (e.g. Friederici, 2002 ; Skeide & Friederici, 2016 ; Hagoort, 2003 ; Hickok et Poeppel 2015). Dans ces modèles, contrairement aux modèles plus classiques dans lesquels le traitement se fait de façon sérielle, des unités les plus petites (phonèmes) vers les unités les plus grandes (mots puis énoncés), le signal de parole est traité de façon partiellement

¹⁷ De l'expression « to lead somebody up the garden path », qui signifie orienter quelqu'un sciemment vers l'allée de jardin, l'équivalent de « mener en bateau ».

parallèle, à différentes échelles temporelles (phonèmes vs. mots et énoncés). Le modèle le plus complet et le plus récent est celui de Skeide & Friederici (2016). Dans ce modèle, la compréhension du langage implique des traitements en série, combinés avec des traitements en parallèle. Elle implique également des processus *bottom-up* précoces et *top-down tardifs*. Elle commence par des processus *bottom-up* précoces (démarrant entre 20 et 120 ms après l'énoncé entendu), se déroulant rapidement (de durée entre 30 et 60ms), inconscients et automatiques. La première étape consiste à déterminer les traits acoustico-phonologiques de chaque segment (phonème) de parole. Si la première syllabe du mot fournit un indice non-ambigu pour reconstruire rapidement le reste du mot, alors la forme phonologique du mot peut être détectée entre 20 et 50ms après l'énonciation. Ensuite, deux traitements ont lieu en parallèle. D'une part, l'information sur la forme phonologique du mot est soumise à la catégorisation morphosyntaxique (i.e. l'assignation à un groupe syntaxique, tel que nom, verbe, préposition), entre 40 et 90 ms après l'énonciation. D'autre part, en parallèle, a lieu la catégorisation lexico-sémantique (qui permet de déterminer si une forme phonologique peut être interprétée sémantiquement), entre 50 et 80 ms après l'énoncé. Cette catégorisation permet d'identifier les pseudo-mots comme non-interprétables. Une fois qu'un mot interprétable est identifié, il est possible d'accéder à l'entrée lexicale correspondante – une représentation mentale associée à la forme phonologique du mot et porteuse d'information sémantique et syntaxique – environ 110 à 170 ms après l'énonciation. En parallèle, dès que les catégories syntaxiques sont identifiées, les structures syntaxiques (syntagmes et phrases) peuvent être construites, environ 120 à 150 ms après l'énoncé. Le traitement prosodique au niveau du syntagme (l'intonation, les informations sur les tons, les accents, le rythme) se fait entre 200 et 400ms après l'énoncé. Les niveaux de traitement plus élevés impliqués dans la compréhension du langage sont descendants (*top-down*), plus tardifs (démarrant entre 200 et 600 ms après l'énoncé), relativement lents (avec des durées d'au moins 150ms), non-automatiques et sous le contrôle de la conscience. Les relations sémantiques entre les items lexicaux (et l'ambiguïté sémantique éventuelle) sont déterminées entre 200 et 400 ms. La construction de syntagmes et de phrases, guidée par des règles, à partir de l'information morphosyntaxique, se fait entre 300 et 500ms. Les syntagmes sont ordonnés et la structure argumentale (ou thématique, qui fait quoi à qui) est construite. Le traitement prosodique au niveau de la phrase (i.e. la focalisation prosodique) se fait en parallèle, entre 300 et 500ms après l'énoncé. Une fois que toutes ces étapes ont été effectuées, toute l'information peut être intégrée en une représentation conceptuelle interprétable.

I.2.1.2.2. Le *monitoring verbal* allocentré : suivi des verbalisations d'autrui

Selon Pickering & Garrod (2014), les psycholinguistes se sont essentiellement penchés sur le *monitoring* de la production du langage, ce qu'ils nomment « self-monitoring » ou « verbal self-monitoring », et que j'ai appelé le *monitoring verbal égocentré* ou $M_S(V_S)$, mais la compréhension du langage implique elle-même des mécanismes de *monitoring verbal*. Pickering & Garrod (2014) notent que, de la même façon que nous sommes capables de reformuler une phrase que nous venons d'énoncer, pour corriger une erreur ou une maladresse, dans une situation de dialogue où nous écoutons autrui, il arrive que nous proposons une correction à notre interlocuteur·rice, ou que nous l'interrogeons sur sa formulation. Ces reformulations ou interrogations sont très fréquentes dans les dialogues entre parents et enfants (Chouinard & Clark, 2003). Ceci suggère que lors d'une conversation, nous utilisons le *monitoring* des productions d'autrui, pour y détecter des

erreurs. J'ajoute que ce *monitoring* ne permet pas uniquement de corriger les erreurs d'autrui, il nous permet de nous assurer que nous interprétons bien ses énoncés. Pickering & Garrod proposent que le *monitoring* des productions d'autrui s'appuie sur des mécanismes prédictifs, similaires à ceux qui s'opèrent dans la production du langage. La compréhension engage des processus de prédiction. Comme je l'ai mentionné ci-dessus, il a été montré que l'analyse des énoncés se fait de façon incrémentale : les auditeurs n'attendent pas la fin de l'énoncé pour commencer à l'interpréter, mais analysent chaque son perçu et l'intègrent avec le contexte précédent. De plus, au cours de situations de dialogue, les auditeurs font des prédictions sur les énoncés qu'ils perçoivent et sont capables même de terminer les phrases de leurs interlocuteurs avant ceux-ci (Gregoromichelaki *et al.*, 2011). Le versant perceptif de la fonction de communication nécessite donc, tout comme le versant productif, un dispositif de *monitoring* spécifique. Il s'agit du *monitoring verbal allocentré*, c'est-à-dire d'un mécanisme permettant de suivre et anticiper les productions langagières d'autrui. Ce *monitoring verbal allocentré* nous permet de percevoir, décoder, analyser, interpréter et comprendre les messages verbaux d'autrui, en utilisant nos connaissances sur le monde, sur nos interlocuteurs, sur le langage, et en tenant compte des signaux de l'environnement. Et de façon miroir au versant production, dont le *monitoring* engage le système de perception de la parole, le *monitoring verbal allocentré* engage lui-même le système de production de la parole, pour faire des prédictions à différents niveaux linguistiques. On a ainsi, comme le notent Skipper *et al.* (2017, p. 97), une « usurpation », par le système de perception du langage, des processus inhérents à la production, ce qui fournit au cerveau un mécanisme général commun, permettant une parité entre les représentations (unités) utilisées en production et celles en perception : les messages reçus par les auditeurs sont similaires à ceux qu'ils transmettraient en tant que locuteurs.

L'hypothèse que le *monitoring* de la production du langage engage le système de perception est assez consensuelle et a été prise en compte dans les modèles de production dès les années 1980 (cf. I.2.1.1.2 ci-dessus). Pour le versant perceptif et la réception du langage, l'hypothèse que le système moteur est lui-même impliqué est plus récente, et les premiers à l'avoir explicitée et prise en compte dans une description générale de la compréhension de la parole sont Pickering & Garrod (2007). Toutefois, l'implication du système moteur dans la perception des unités bas niveau de la parole (phonèmes) a été entrevue très tôt. Les travaux de Liberman à la fin des années 1950 aux *Haskins Laboratories* l'ont conduit à développer, avec ses collègues, la Théorie Motrice de la perception, introduite au paragraphe I.2.1.2.1 ci-dessus. Cette théorie, et les théories qui en découlent ou s'en inspirent, notamment la PACT de Schwartz et collègues citée plus haut, soutient que le système moteur est impliqué dans la perception de la parole, l'auditeur-riche cherchant à recouvrer les gestes moteurs à l'origine du signal de parole. Selon les théories motrices ou perceptuo-motrices, l'unité d'échange entre les interlocuteurs est de type moteur : percevoir la parole c'est percevoir les gestes moteurs initiaux (dans la PACT, l'unité d'échange est plus précisément perceptivo-motrice, voir Schwartz *et al.*, 2012). Corroborant cette hypothèse, de nombreuses études de neuroimagerie (IRMf ou EEG) ou de neurostimulation (TMS) tendent à montrer que les régions cérébrales motrices et prémotrices typiquement impliquées dans la production de parole sont également activées lors de la perception (pour une revue cf. Schwartz *et al.*, 2008). En lien avec la découverte des neurones miroirs, il a ainsi été proposé qu'un phénomène de « résonance motrice » soit à l'œuvre lors de la perception de la parole, l'auditeur-riche entrant en résonance avec le-la locuteur-riche et activant ses régions

motrices (Fadiga *et al.*, 2002). Cette résonance motrice serait même très précise et spécifique : les sons dont la production requiert l'activation des muscles de la langue (comme le son 'r' en italien) activent chez l'auditeur·rice, lorsqu'ils sont entendus, les centres moteurs contrôlant spécifiquement les muscles de la langue (Fadiga *et al.*, 2002, Roy *et al.*, 2008). Ce mécanisme de résonance motrice a été rapproché du système de correspondance entre observation et exécution qui est supposé être impliqué dans la reconnaissance de l'action (neurones miroirs, Rizzolatti *et al.*, 1996). Selon Rizzolatti et collègues, les neurones miroir sont un mécanisme permettant de comprendre les actions produites par autrui. Les actions sont ainsi reconnues car l'agent de l'action et l'observateur partagent tous les deux le même répertoire moteur. Ce mécanisme général contribuerait également à la perception de la parole, l'implication du système moteur permettant de reconnaître les actions de parole d'autrui, en retrouvant les gestes moteurs sous-jacents. Ainsi, selon Fadiga et collègues, la perception de la parole engage un processus automatique de mise en correspondance entre les gestes articulatoires du·de la locuteur·rice et les programmes moteurs de l'auditeur·rice.

D'autres équipes se sont intéressées à l'implication du système moteur dans la perception de la parole et ont proposé leur propre interprétation du rôle fonctionnel du système moteur dans la perception. Selon Skipper et collègues (2007), supposer que le système moteur soit impliqué dans la perception de la parole par un mécanisme de résonance motrice n'est pas suffisamment explicatif. Car simplement reconnaître une action (par système miroir) ne permet pas de l'interpréter, i.e. de comprendre la signification de ce geste, en termes langagiers. Skipper et collègues proposent un modèle de la perception audiovisuelle de la parole qui vise justement à expliquer le mécanisme par lequel l'activité du système moteur est comprise, interprétée comme du langage (et non pas simplement reconnue). Dans ce modèle, inspiré du modèle de « l'analyse-par-synthèse » de Stevens & Halle (1967), la perception de la parole consiste à faire une hypothèse multisensorielle à propos du stimulus audiovisuel présenté. Cette hypothèse, extraite d'informations sensorielles, est associée à des commandes motrices (sur la base des propres expériences de production de parole antérieures). Ces commandes motrices permettent, par un mécanisme de copie d'efférence (cf. I.2.1.1.2 ci-dessus) impliquant un modèle interne direct, de prédire les conséquences acoustiques et somatosensorielles de l'exécution du mouvement de parole hypothétique. Ces prédictions sont comparées à l'entrée sensorielle initiale et permettent d'ajuster la perception de la parole, en contraignant l'interprétation finale.

Hickok & Poeppel ont une version légèrement différente du rôle du système moteur dans la perception de la parole. Ils conçoivent, comme Skipper et collègues, des liens bidirectionnels entre les régions cérébrales sensorielles et motrices mais ils n'envisagent pas que le système moteur soit systématiquement impliqué dans toute tâche de perception du langage. En effet, selon Hickok & Poeppel (2007), le rôle primordial des régions frontales dans la perception de la parole serait de faciliter l'acquisition du langage chez l'enfant. Pour ces auteurs, l'apprentissage de la parole est une tâche essentiellement motrice. L'entrée principale étant sensorielle (auditive), un mécanisme neural permettant de conserver des traces des sons de parole et d'utiliser ces traces sensorielles pour ajuster les gestes articulatoires est nécessaire. Les régions motrices frontales seraient donc particulièrement activées lors de la perception de la parole durant l'enfance afin de maintenir une parité entre les représentations auditives et motrices de la parole. Chez l'adulte, elles seraient cependant moins cruciales lors de conditions

d'écoute de parole normales ou habituelles et redeviendraient fonctionnellement pertinentes lors de la mise en jeu de processus articulatoires, dans des tâches phonologiques, de mémoire de travail verbale ou d'apprentissage de nouveaux sons, syllabes ou mots.

Ces deux modèles sont particulièrement intéressants, car il n'y est pas simplement mentionné que le processus de perception de la parole consiste à retrouver les gestes produits par autrui, mais plus précisément que des processus moteurs sont en jeu chez l'auditeur-riche. Pour Hickok & Poeppel (2007), lors de l'apprentissage ou de la mémorisation, la perception de la parole engage le système moteur en activant les régions qui soutiennent les représentations articulatoires motrices (région de Broca et régions prémotrices articulatoires). On retrouve donc là la notion de résonance motrice présentée par Rizzolatti, Fadiga et collègues, compatible avec la Théorie Motrice de la perception. Mais Skipper *et al.* (2017) vont plus loin. Selon eux, la perception de la parole ne met pas simplement en jeu des correspondances automatiques (plus ou moins directes) unidirectionnelles entre représentations sensorielles et motrices. Les liens entre régions sensorielles et motrices sont bidirectionnels, via une connexion directe (*feedforward*) des régions sensorielles vers les régions motrices puis une connexion retour (*feedback*) des régions motrices vers les régions sensorielles. Ainsi ces auteurs suggèrent explicitement que le mécanisme d'analyse-par-synthèse impliqué dans la perception de la parole consiste effectivement à produire mentalement des gestes de parole, en utilisant une copie d'efférence des commandes motrices pour élaborer (via un modèle interne direct) une prédiction sensorielle qui sera comparée à l'entrée sensorielle initiale. Sur le plan théorique, cette hypothèse de Skipper et collègues est en adéquation avec certaines théories de la perception en général (et notamment de la perception visuelle) dans lesquelles les modèles internes et le système moteur jouent un rôle fondamental (voir aussi Poeppel & Monahan, 2011 ou Tian & Poeppel, 2013). Ainsi, comme l'explique Grush (2004), contrairement aux théories traditionnelles qui considèrent que la perception est un processus principalement ascendant (*bottom-up*), exclusivement issu de l'appareil sensoriel périphérique, les théories plus récentes, inspirées des travaux de Kosslyn (2005) sur la perception visuelle, soutiennent que des processus descendants sont également impliqués. Ainsi, selon Kosslyn (et en accord avec la notion « d'effort de volonté » et de copie d'efférence entrevue par Von Helmholtz dès 1866) des processus d'imagerie visuelle (incluant de la motricité) sont à l'œuvre dans la perception visuelle, pour guider la perception et compléter les éventuelles informations manquantes par des attentes (des prédictions) visuelles. Grush a lui-même élaboré une théorie des représentations fondée sur l'émulation (*Emulation Theory of Representation*) qui permet de rendre compte de façon unifiée de la façon dont le cerveau manipule des représentations, à la fois lors du contrôle moteur, de l'imagerie motrice et de la perception. Dans ce cadre théorique, des circuits neuronaux sont construits, fonctionnant comme des modèles internes du corps et de l'environnement. Grush considère ces modèles internes comme des « émulateurs ». Dans le cadre du contrôle moteur (comme proposé notamment par Kawato *et al.*, 1987 ou Jordan & Rumelhart, 1992), ces modèles internes utilisent des copies d'efférence des commandes motrices envoyées effectivement à l'appareil moteur (au corps) pour faire des prédictions sensorielles, en avance des retours sensoriels effectifs, et ajuster les commandes avant même que l'action soit produite (cf. le mécanisme de copie d'efférence expliqué dans la section I.2.1.1.2). Dans le cadre de l'imagerie motrice, ces émulateurs fonctionnent hors ligne, sans l'implication effective de l'appareil moteur (du corps), simplement avec les

copies d'efférence. Les sorties générées par les émulateurs sont les sensations imaginées. Dans le cadre de la perception, et en particulier de la perception du langage, les prédictions sensorielles générées par les émulateurs sont utilisées pour interpréter les entrées sensorielles effectives. En ce qui concerne l'imagerie motrice, Grush (2004) fait la distinction entre la (ou les) théorie de la « simulation » motrice et celle de « l'émulation ». Selon lui, dans la théorie de la simulation motrice, telle que défendue par Jeannerod (2001) par exemple, l'imagerie motrice correspond simplement à l'opération hors-ligne, court-circuitée en quelque sorte, des systèmes d'efférences moteurs, ne transmettant plus l'information aux effecteurs (ou l'information transmise étant inhibée). Dans la théorie de l'émulation, il est également supposé que le contrôleur, i.e. les aires motrices efférentes, est actif durant l'imagerie motrice, mais pas uniquement. Grush ajoute qu'un émulateur de l'appareil musculo-squelettique, un modèle interne, est à l'œuvre et que l'imagerie est produite par cet émulateur, qui reçoit les commandes émises par les aires motrices efférentes. Je pense que l'interprétation de Grush (2004) des travaux de Jeannerod est erronée et que la théorie de la simulation motrice, telle qu'elle est envisagée par Jeannerod et ses collègues, inclut bien la notion de modèle interne et donc d'émulation. On peut lire par exemple dans Jeannerod & Pacherie (2004, p. 132), à propos des trois types d'actions qui recourent toutes au système de copie d'efférence, i.e. l'action effective (et sa préparation), l'action imaginée (l'imagerie motrice) et l'action observée (la perception des actions d'autrui) :

« the motor system with its internal inverse and forward models computes representations of imagined and observed actions as well as representations of actions to be performed by the agent »¹⁸

Ainsi, Skipper et collègues, tout comme Grush (et comme Jeannerod et Pacherie pour la perception des actions en général), considèrent que la perception du langage engage le système moteur en mettant en jeu un modèle interne, un émulateur du système moteur réel. Cet émulateur construit des représentations mentales qui permettent, selon Skipper et collègues d'ajuster, de contraindre l'interprétation du message, et selon Grush de fournir, en avance des sensations à venir, des attentes sensorielles qui complètent les sensations présentes (elles-mêmes incomplètes) et guident l'interprétation. Le modèle de Skipper *et al.* (2007) est restreint à l'interprétation des unités phonétiques de la parole, phonèmes et syllabes. Rien n'est spécifié pour les niveaux plus élevés et notamment pour l'interprétation sémantique. Grush (2004) n'explique pas comment sa théorie de l'émulation s'applique précisément au langage mais il défend qu'un système d'émulateurs est à l'œuvre dans la représentation et l'interprétation sémantique. Il rapproche sa théorie des travaux menés dans le cadre de la linguistique cognitive (ou de la linguistique incarnée), dans lequel des phénomènes linguistiques variés (anaphore, quantification, composition nominale, construction passive) et la teneur sémantique des messages sont appréhendés à travers des structures représentationnelles dérivées de l'action (il se réfère notamment à Fauconnier, 1985 ; Langacker, 1999, Talmy, 2000 ; Barsalou, 1999 ; Lakoff & Johnson, 1999). Wilson & Knoblich (2005) ont développé la notion d'émulation, telle que présentée par Grush (2004), en argumentant qu'en plus d'offrir un mécanisme pour la compréhension et l'interprétation des actions (en général, pas uniquement le langage), l'émulation permet d'anticiper et de prédire le déroulement

¹⁸ « Le système moteur, avec ses modèles internes inverse et direct, construit des représentations des actions imaginées et observées, ainsi que des représentations des actions à effectuer par l'agent »

de l'action, à l'avance. L'émulation n'est donc pas uniquement post-dictive (interprétation de l'action *a posteriori*) mais prédictive, elle permet de générer des prédictions descendantes (*top-down*) sur l'action en cours. Elle permet donc de suivre les actions d'autrui (*monitoring*) en temps réel, de compléter les informations éventuellement manquantes ou ambiguës et d'anticiper le déroulement temporel futur de l'action en cours. Selon Wilson & Knoblich (2005), l'émulation motrice permet non pas simplement de retrouver l'intention passée qui a causé l'initiation de l'action, mais plus fondamentalement, d'envisager son déroulement temporel probable dans le futur immédiat. Wilson & Knoblich précisent que ce mécanisme fonctionne pour les comportements moteurs prédictibles. Il me semble que l'on peut donc proposer qu'il s'applique au langage. Dans la même veine, et dans le cadre de la Théorie du Codage Commun (*Common Coding Theory*) de Prinz (1997), Sebanz & Knoblich (2009) décrivent comment les mécanismes de prédiction se déroulent dans l'action conjointe. Selon eux, c'est grâce à un codage commun des actions perçues et effectuées que les agents d'action peuvent prédire des informations sur les actions d'autrui, en les caractérisant par des réponses aux trois questions fondamentales « Quoi, Quand, Où ». Selon Sebanz & Knoblich, le codage commun, en termes d'objectif de l'action, fournit une plate-forme représentationnelle qui intègre les actions d'autrui et de soi-même et permet d'expliquer comment des actions conjointes coordonnées, et notamment la communication verbale, peuvent se dérouler de façon fluide. Un élément important dans la proposition de Sebanz & Knoblich (issue de la théorie de Prinz) est que les actions sont planifiées et interprétées à partir de leur effet final désiré, leur objectif distal, et non pas en termes de leurs trajectoires proximales détaillées. Comme nous l'avons vu (cf. I.2.1.1.2), cette distinction entre état final et trajectoire détaillée est importante car elle permet d'expliquer que les états désirés sont dans un format schématique, supramodal, différent du format de l'expérience effective (proximale), qui prend une forme développée temporellement et sensoriellement. Cette distinction permet aussi de rendre compte de la phénoménologie des variantes de la parole intérieure, la petite voix dans la tête, cf. section I.2.2).

Pickering & Garrod (2007, 2013, 2014) appliquent ces principes concernant les actions en général au cas particulier du langage. En particulier, Pickering & Garrod (2013, 2014) offrent une théorie élégante de la production et de la compréhension du langage qui est une des premières théories à intégrer les deux versants de la communication langagière, de façon hiérarchique, du concept pré-verbal aux détails phonétiques, dans un cadre unique. Cependant cette théorie comporte quelques imprécisions.

Comme nous l'avons vu au paragraphe I.2.1.1.2 ci-dessus, sur le versant production, la notion de modèle interne ne convient pas à tous les niveaux de la hiérarchie (notamment aux niveaux syntaxique et sémantique, puisqu'il n'y a pas d'appareil biophysique à modéliser) et la façon dont le *monitoring* s'exerce d'un niveau à l'autre n'est pas explicitée par Pickering & Garrod (2013, 2014).

Sur le versant perception, le modèle prévoit que l'auditeur A récupère l'intention du locuteur B à partir de son action en utilisant un modèle inverse et en tenant compte, en plus, du contexte, c'est-à-dire de l'éventuelle différence entre l'appareil vocal du locuteur B et de l'auditeur A. Il s'agit donc d'adapter le modèle inverse, habituellement utilisé par l'auditeur A pour construire ses propres commandes de parole, au locuteur B. Il est important de clarifier ce point et d'expliquer comment cette adaptation du modèle inverse se met en place. De plus, Pickering & Garrod expliquent que le modèle inverse fournit une commande d'action, qu'ils considèrent correspondre à l'intention de B. Cette

commande motrice serait celle utilisée par le locuteur A pour effectuer l'action de B. Je pense que le terme est impropre ici et qu'il ne s'agit pas de l'intention de B, mais simplement des commandes motrices, donc cette partie du modèle de Pickering & Garrod est ambiguë. Il reste donc nécessaire de mieux expliquer comment l'intention elle-même est récupérée. Car, ensuite, selon Pickering & Garrod, l'auditeur A utilise l'intention pour simuler la prochaine action de B en utilisant son propre modèle direct. Les auteurs donnent pour exemple la complétion de phrase, dans une situation de dialogue où un enfant B a reçu en cadeau un cerf-volant. L'énoncé en train d'être produit par l'enfant B est : « I want to go out and fly my » (« je veux sortir et faire voler mon »). Les auteurs estiment que l'enfant A peut prédire que la suite de l'énoncé sera « kite » (cerf-volant) et qu'il peut faire cette prédiction en s'appuyant sur l'intention récupérée par son modèle inverse. Or si l'intention est la commande motrice, il me semble que la simple utilisation de la commande motrice en cours ne permet pas de connaître quelle sera la prochaine commande motrice. Il faut donc selon moi un mécanisme d'inversion en cascade, permettant, avec le contexte, d'inférer l'intention communicative (« vouloir sortir pour faire voler le cerf-volant ») à partir du signal phonétique perçu. On a ainsi une inversion du niveau phonétique au niveau syntaxique et du niveau syntaxique au niveau sémantique.

Ensuite, selon Pickering & Garrod (2013, 2014) à partir de ce que l'auditeur A pense être l'intention du locuteur B, l'auditeur construit une expérience sensorielle simulée en utilisant sa propre machinerie, son propre modèle direct. Il me semble que cette proposition engendre un conflit. En effet, si le modèle direct utilisé est celui de A, le signal généré sera donc un énoncé prononcé mentalement avec toutes les caractéristiques de l'auditeur A. Il sera donc différent de l'énoncé entendu (la sensation effective). Il faut donc, selon moi, qu'un modèle interne adapté à autrui soit utilisé.

Je propose donc ici une version adaptée de la proposition originale de Pickering & Garrod (2013, 2014) qui tente de lever certaines ambiguïtés et de remédier à certains des problèmes relevés. Cette version est également compatible avec les propositions de Proust & Pacherie (2008) sur le rôle de la simulation par modèles internes hiérarchisés, dans la compréhension d'autrui et avec celle de Poeppel & Manahan (2011) sur l'implication de mécanismes d'analyse-par-synthèse, en perception. Elle est illustrée sur la Figure I-13 ci-dessous. Elle reprend les prédicteurs et contrôleurs utilisés dans le modèle de production hiérarchique décrit sur la Figure I-12, les prédicteurs des deux niveaux supérieurs (phonologico-syntaxico-prosodique et sémantique) étant en fait plutôt des analyseurs. Elle repose sur la même implémentation neurale des différents processus, puisqu'ils sont en commun dans la production et la perception. Cette implémentation est donc compatible avec la notion d'usurpation de la production par la perception, et réciproquement, proposée par Skipper *et al.* (2017). Les régions cérébrales et cérébelleuses impliquées sont d'ailleurs retrouvées dans les propositions de ces auteurs : cortex auditif, bien sûr, mais aussi lobe temporal moyen, régions frontales inférieures, prémotrices et motrices, lobule pariétal inférieur, cervelet, thalamus.

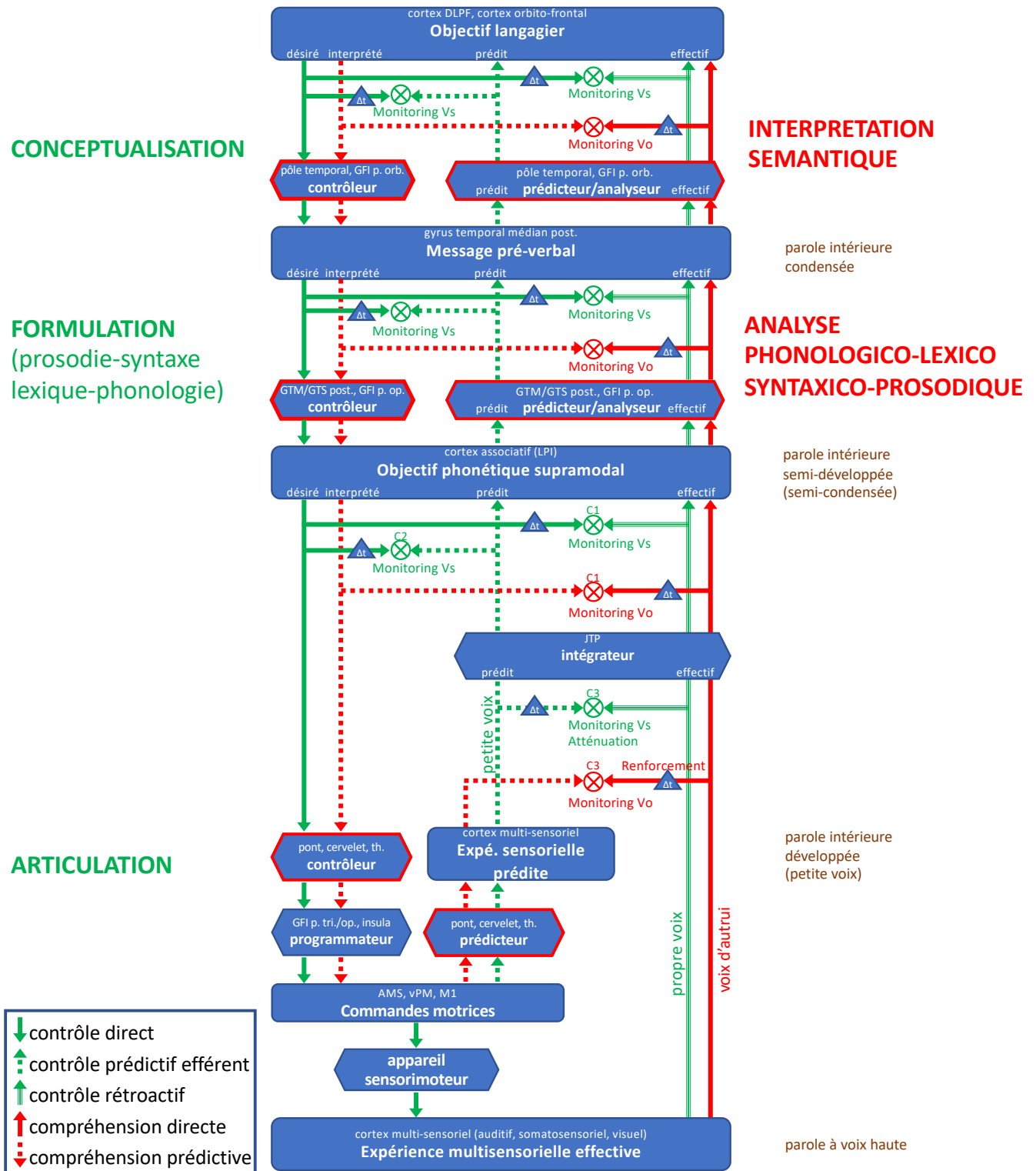


Figure I-13. Modèle hiérarchique du monitoring de la perception du langage (circuit rouge), greffé sur le monitoring de la production de la parole (circuit vert). Ce modèle est inspiré et adapté de Pickering & Garrod (2013, 2014) et de Wolpert et al. (2003). Des mécanismes probabilistes sont en jeu dans la transmission des signaux d'erreur de prédiction à différents niveaux. Les différents contrôleurs et prédicteurs/analyseurs sont des modèles adaptés à autrui (donc entourés de rouge).

Dans le modèle que je propose (Løevenbruck, en préparation), la production de l'interlocuteur B correspond à une expérience sensorielle effective, qui est intégrée pour former un objectif phonétique supramodal. Cet objectif phonétique est converti en un message pré-verbal effectif via le prédicteur/analyseur du niveau phonologico-syntaxico-sémantique. Celui-ci est lui-même transformé en intention communicative effective via le prédicteur/analyseur du niveau sémantique. C'est cette intention communicative qui constituera l'interprétation par l'auditeur A de l'énoncé de B. Mais cette interprétation doit être d'abord validée, par un mécanisme de prédiction. L'intention communicative de B ainsi inférée est d'abord transformée en message pré-verbal à l'aide du contrôleur de message pré-verbal, puis en objectif phonétique à l'aide du contrôleur d'objectif phonétique. L'objectif phonétique est converti en commandes motrices (non transmises à l'appareil moteur) puis en expérience sensorielle simulée. Cette expérience sensorielle prédite peut être comparée à l'expérience sensorielle effective initiale (auquel un délai Δt a été ajouté pour temporiser). En cas d'erreur entre les expériences sensorielles prédite et effective, l'interprétation de l'intention communicative de B est rectifiée.

Dans ce modèle, les différents prédicteurs et contrôleurs sont des modèles allocentriques, adaptés à autrui (B). Ceci signifie que la compréhension d'autrui nécessite de disposer de modèles d'autrui adaptés (cf. von Kriegstein *et al.*, 2008 pour des propositions sur l'utilisation, lors de la perception de la parole, de modèles d'autrui, intégrant des informations auditives et visuelles). La correction appliquée au niveau de l'intention communicative est également apportée aux différents contrôleurs et prédicteurs, par un mécanisme de pondération probabiliste similaire à celui envisagé en production de parole et tenant compte du degré de confiance que l'auditeur·rice a envers ses modèles d'autrui, à différents niveaux (on connaît plus ou moins bien l'accent de son interlocuteur·rice, ses habitudes lexicales, syntaxiques, ses connaissances, ses croyances, ses motivations). Ces modèles d'autrui sont probablement construits à partir des modèles propres du·de la locuteur·rice (utilisés pour sa propre production) et adaptés en fonction des connaissances que le·la locuteur·rice a de l'auditeur·rice. Ces modèles paramétrés pour autrui sont donc entourés en rouge sur la Figure I-13 pour les distinguer des modèles propres du·de la locuteur·rice.

On voit ainsi, comme le résumait élégamment Proust & Pacherie (2008, p. 17) que « *perception et action sont en relation de codépendance régulatrice.* ».

Outre l'utilisation de modèles paramétrés pour autrui, une autre différence importante existe entre le circuit de *monitoring* de la production (en vert sur la Figure I-13) et celui de la compréhension du langage : celle des délais de temporisation. Dans le cas du *monitoring* égocentré, les signaux à temporiser, en attendant les signaux effectifs, sont les signaux prédits. Dans le cas du *monitoring* allocentré, le signal à temporiser, en attente du signal prédit, est le signal effectif. Il est donc concevable que des mécanismes de contrôle exécutif de plus haut niveau enclenchent ces temporisations différenciées pour que les comparaisons aient lieu à bon escient.

Atténuation sensorielle de la parole auto-produite et renforcement de la parole d'autrui

Il a été montré que l'activation du cortex auditif est atténuée lorsque nous parlons nous-mêmes (alors que le signal auditif en sortie de notre bouche est relativement fort, puisque très proche de nos organes auditifs) par rapport à l'activation mesurée lorsque nous écoutons autrui parler, ou lorsque nous écoutons un enregistrement de nous-mêmes (cf. e.g. Ford & Mathalon, 2004 ; Christoffels *et al.*, 2011). Agnew *et al.* (2013)

expliquent que cette atténuation permet de renforcer la saillance des productions vocales d'autrui par rapport à nos propres productions, de les rendre plus détectables et interprétables. Selon moi, l'atténuation concerne probablement aussi le signal prédit et empêche la sensation de deux stimuli auditifs (signal prédit puis effectif). Il a été proposé (cf. section 1.2.3) que ce soit la comparaison entre les signaux prédits et effectifs qui donne lieu à une atténuation sensorielle, si les signaux sont semblables. Cette proposition ne paraît adaptée que si les signaux comparés sont bien ceux qui sont alignés temporellement, car sinon la comparaison devrait aboutir au même résultat dans le cas de la perception de parole (et devrait donc atténuer la parole perçue).

Importance de la temporisation et de la synchronisation des signaux comparés

Il me paraît donc crucial que des mécanismes de temporisation entrent en jeu pour que les comparaisons permettent dans un cas le *monitoring* des auto-productions et dans l'autre celui des productions d'autrui. Dans le cas du *monitoring* verbal égocentré, la temporisation est faite sur le signal prédit, probablement moins saillant perceptivement que le signal effectif. Dans le cas du *monitoring* verbal allocentré, la temporisation est faite sur le signal effectif, ce qui est susceptible de sur-activer le cortex auditif et de rendre la parole d'autrui subjectivement plus forte que la parole que nous produisons nous-mêmes. Ainsi, il me paraît judicieux de supposer, comme l'ont fait Blakemore *et al.* (2003) que la comparaison entre signaux prédits et effectifs dans le cas de la production de parole permette le *monitoring* de nos verbalisations (et la correction en ligne si nécessaire). L'atténuation des signaux auto-produits pourrait provenir d'un mécanisme *ad-hoc* à ce niveau. Il se peut aussi qu'il n'y ait pas d'atténuation des signaux auto-produits, mais simplement un renforcement des signaux extérieurs : la temporisation des signaux sensoriels effectifs pourrait suffire en elle-même à renforcer les signaux extérieurs, les signaux effectifs auto-produits (ne subissant pas de temporisation) paraissant alors relativement atténués. Je reviens sur ce point dans la section 1.2.3. Cette hypothèse sur l'importance de la temporisation est d'ailleurs déjà formulée dans le cadre des actions générales (pas dans le cas de la parole) par Decéty & Jackson (2004), qui proposent qu'une différence de temporisation permette de distinguer les actions d'autrui de nos propres actions.

Encadré 1.24. Rôle crucial de l'alignement temporel des signaux à comparer

Il s'avère que la gestion de l'alignement temporel des signaux à comparer est capitale. Dans le cas du *monitoring* verbal égocentré (cf. 1.2.1.1.2), la temporisation doit se faire sur le signal prédit, qui est mis en attente le temps que le signal effectif soit généré.

Dans le cas du *monitoring* verbal allocentré, la temporisation est faite sur le signal effectif, qui est mis en attente le temps que la prédiction soit effectuée.

Cette temporisation pourrait contribuer à atténuer sensoriellement la parole auto-produite et à renforcer les stimuli externes, ce qui nous permettrait de mieux détecter les paroles d'autrui.

Une mauvaise gestion de ces temporisations peut faire échouer le *monitoring* verbal et pourrait être à l'origine des phénomènes hallucinatoires verbaux (cf.1.2.3).

Je ne détaille pas ici comment sont traités les différents sous-niveaux phonétiques, quelles unités sont gérées à différents sous-niveaux en fonction de la langue (phonème,

syllabe, accent, more) ni comment les informations prosodiques et segmentales s'organisent, mais des pistes compatibles avec le modèle que je propose ont été avancées, notamment par Norris, McQueen & Cutler (2016), dans un cadre prédictif bayésien. Ce modèle est également en accord avec les données expérimentales récentes de Blank & Davis (2016) qui démontrent l'existence d'un mécanisme prédictif dans la perception de parole. À l'aide de simulations computationnelles sur des données d'IRMF, Blank & Davis expliquent comment l'erreur de prédiction interagit avec les signaux sensoriels dans le processus de perception de la parole. Leurs données soutiennent l'hypothèse du *Predictive Coding*, i.e. que les signaux qui correspondent aux prédictions sont atténués (et donc non traités) et que c'est l'erreur de prédiction qui est l'objet de traitements perceptifs (dans le sillon temporel supérieur postérieur).

Encadré 1.25. $M_S(v_0)$: monitoring verbal allocentré

Pour distinguer ce sous-système de *monitoring* verbal des autres sous-systèmes impliqués dans le langage, j'utiliserai par la suite (cf. 1.2.4) le terme de *monitoring par soi-même des verbalisations d'autrui* ou $M_S(v_0)$, *monitoring* verbal allocentré.

1.2.2. Le *monitoring* verbal dans la pensée

Pour que le processus de pensée verbale, impliquant la production de langage intérieur, ou endophasie (Bergounioux, 2001b), se déroule avec succès, un mécanisme de *monitoring* s'appliquant aux verbalisations intérieures est également nécessaire. Je propose que, tout comme le mécanisme de *monitoring verbal égocentré* s'appliquant à la parole extériorisée, décrit au paragraphe 1.2.1.1.2 ci-dessus, ce mécanisme peut engager des processus de perception. C'est également ce que défend Bergounioux (2001b, 2004) :

« *l'endophasie paraît indissociable d'un postulat qui attribue aux éléments sensibles de la langue (liés au signal) une forme et une nature identiques à celles des séquences mentalisées. Entre les deux, les différences seraient d'ordre acoustique (l'une explicite, l'autre implicite) mais non fonctionnelle. La façon dont les messages sont reçus et analysés est en tous points comparable à celle qui s'applique au dévidement intérieur de la parole, les processus de compréhension étant du même ordre qu'ils concernent ce que l'auditeur entend en lui ou ce qu'il perçoit au dehors, l'endophasie ou l'exophasie.* » Bergounioux (2001b, p. 19).

1.2.2.1. Trois axes du langage intérieur : condensation, dialogalité et intentionnalité

Le langage intérieur peut varier sur plusieurs axes, dont le degré de condensation, de dialogalité et d'intentionnalité : il peut prendre des formes plus ou moins condensées (ou développées), varier entre monologue et dialogue (ou conversation) et être plus ou moins délibéré (Alderson-Day & Fernyhough, 2015 ; Løevenbruck *et al.*, 2018). Le *monitoring* verbal n'est probablement pas le même pour toutes les variantes de langage intérieur, il est donc important de les distinguer.

En ce qui concerne le degré de condensation, il semble que certaines variantes du langage intérieur soient condensées par rapport à d'autres, plus formées, développées (e.g. Fernyhough, 2004 ; McCarthy-Jones & Fernyhough, 2011 ; Alderson-Day &

Fernyhough, 2015 ou Geva, Jones, Crinion, Price, Baron & Warburton, 2011b). La raison de la condensation a été expliquée par Egger dès 1881 :

« Dans la parole intérieure, il suffit que nous soyons compris de nous-mêmes ; nous pouvons donc parler très bas, très vite, peu distinctement, abrégé les phrases, remplacer les tournures et les expressions usuelles par d'autres plus simples ou plus expressives à notre goût, modifier la syntaxe, enrichir le vocabulaire par des néologismes ou des emprunts aux langues étrangères ; nous pouvons nous exprimer à nous-mêmes la nuance toute personnelle de nos sentiments par des termes dont nous créons le sens à notre usage ».

Selon Vygotski (1934/1985), cette forme linguistique condensée du langage intérieur le rend difficile à comprendre : « Le discours intérieur est abrégé, décousu, fragmenté, incompréhensible pour un auditeur extérieur. ». Vygotski affirme que le langage intérieur est condensé sur le plan de la syntaxe, et même elliptique. Certains mots ou affixes grammaticaux peuvent être supprimés et des formes prédicatives sont souvent observées, contenant uniquement l'information nécessaire. Le sujet et même parfois le verbe peuvent être omis. On retrouve cette notion de condensation du langage intérieur chez Jakobson :

« Quant au discours non extériorisé, non prononcé, ce qu'on appelle le langage intérieur, ce n'est qu'un substitut elliptique et allusif du discours explicite et extériorisé. » (Jakobson, 1963, p.32).

Pour Bergounioux (2001a, p.120), le langage intérieur reste toutefois proche du langage à voix haute :

« L'endophasie ne semble différer de la parole explicite ni par sa grammaire, ni par son lexique, à la réserve d'un emploi généralisé de l'asyndète et de l'anaphore, et d'une surreprésentation de la prédictivité. ».

Ces hypothèses sur la nature condensée du langage intérieur, aux niveaux lexical et syntaxique, sont corroborées par des données de psycholinguistiques sur le débit d'élocution. Dans une étude astucieuse, Korba (1990) a demandé à des participants de résoudre mentalement une série de courts problèmes verbaux. Les participants devaient rapporter les fragments de parole intérieure produits pour résoudre chaque problème, ce qui a donné une estimation du nombre de mots utilisés dans ce type de raisonnement mental. On leur a demandé ensuite de développer ces énoncés pour fournir des énoncés ostensifs complets décrivant leurs stratégies de résolution, ce qui a fourni un deuxième nombre de mots. Ce nombre de mots a été rapporté à la durée de résolution de chaque problème. Dans la version développée, la vitesse d'élocution obtenue correspondait à plus de 4000 mots par minute, une vitesse inatteignable à voix haute. D'autres études à partir de monologues extériorisés chez les enfants ou issus de corpus de parole intérieure restituée *a posteriori* chez des adultes, confirment que la parole intérieure est souvent fragmentée, et condensée sur le plan lexical et syntaxique (cf. Wiley, 2014 pour des exemples).

On trouve des illustrations de ce phénomène de condensation dans de nombreux romans, notamment ceux inscrits dans le courant du « monologue intérieur » initié par Edouard Dujardin (Smadja, 2018). L'asyndète, repérée par Bergounioux, est une figure de style fondée sur la suppression des liens et des conjonctions, elle est évidente dans

« la maison ; le vestibule », dont la forme à voix haute pourrait être « Voici la maison puis le vestibule » un exemple tiré de « Les Lauriers sont coupés », roman d'Édouard Dujardin publié en 1887. L'anaphore est également une figure de style qui permet la reprise d'un élément en le remplaçant par un démonstratif ou un pronom par exemple. Elle est mise en jeu dans cet autre exemple des *Lauriers sont coupés* : « Ce garçon sera encore chez soi », qu'il faut lire « Mon ami Lucien Chavainne, qui est clerc, sera encore à l'étude notariale ». Enfin, la prédicativité est le procédé par lequel on dit quelque chose à propos d'un thème, sans le rappeler. Dans *Les lauriers sont coupés*, on trouve par exemple « Et c'est l'heure », qui pourrait être verbalisée en « Et l'heure de mon rendez-vous chez mon ami est enfin arrivée. ». Les monologues intérieurs décrits par Dujardin sont cependant constitués de phrases syntaxiquement complètes, développées. Les interprétations littéraires ultérieures – telles que le monologue de Molly Bloom dans *Ulysses* de Joyce, ou le monologue disjoint de *L'Innommable* de Beckett, ou encore les monologues intérieurs humains captés par les anges dans *Les Ailes du Désir* de Wim Wenders – sont plus proches des descriptions de Egger, Vygotski, Bergounioux ou Korba, car elles sont plus fragmentées, abrégées, condensées, à la fois aux niveaux syntaxique et lexical.

Dans la lignée de l'hypothèse de Egger (1881) selon laquelle la parole intérieure est peu distinctement articulée, il a été proposé que la forme phonologique des mots produits en parole intérieure soit elle-même abrégée. Les psychologues soviétiques dans la lignée de Vygotski considéraient que la parole intérieure est réduite sur le plan phonologique, de nombreux phonèmes pouvant être omis, en particulier les voyelles, et seuls les sons à l'initiale des mots étant clairement produits (e.g. Vygotski, 1934/1986, p. 237, 244 ; Anan'ev citée par Sokolov, 1972, p. 50). Quelques études psycholinguistiques récentes sur les débits d'élocution comparés en voix haute et en parole intérieure, ainsi que sur la comparaison des types d'erreur produites dans ces deux modes, semblent confirmer cette hypothèse que le langage intérieur est probablement aussi réduit sur le plan phonologique (pour une revue, cf. Løevenbruck *et al.*, 2018 et Løevenbruck, 2018). Il a en particulier été défendu par certains auteurs que la production de langage intérieur implique bien des unités au niveau de l'énoncé ainsi que des unités de plus bas niveaux : unités lexicales, morphologiques et phonologiques, mais que les représentations articulatoires et auditives en sont absentes. Pour certains chercheurs, la parole intérieure serait en effet de nature abstraite, i.e. ni articulatoire, ni auditive. MacKay (1992b), notamment, affirme que la parole intérieure n'est spécifiée ni au niveau articulatoire, ni au niveau auditif. La génération de parole intérieure, semble phénoménologiquement inclure des représentations articulatoires, mais ces représentations sont en fait de nature phonologique abstraite :

*« the generative component of internal speech, which seems on the surface to be articulatory in nature, in fact involves a phonological representation. Finally, activity of the speech musculature is unnecessary for the normal functioning and acquisition of the processes underlying inner speech and rehearsal. »*¹⁹ (MacKay, 1992b, p. 126).

¹⁹ « La composante générative (productive) de la parole interne, qui semble, à la surface, être de nature articulatoire, implique en fait une représentation phonologique. Enfin, l'activité des muscles de la parole n'est pas nécessaire au fonctionnement normal et à l'acquisition des processus sous-tendant la parole intérieure et la répétition mentale. » (ma traduction)

MacKay précise également que l'expérience phénoménologique d'une voix intérieure ne repose pas sur une expérience auditive mais sur des représentations phonologiques abstraites :

« [t]he seemingly auditory quality of our internal speech cannot be automatically attributed to events within an auditory or acoustic system, or even, as we will see, to any strictly sensory system. »²⁰ (MacKay, 1992b, p. 127)

Se fondant sur des études psycholinguistiques mesurant le début d'élocution ou les examinant les types d'erreurs produites en parole intérieure, MacKay (1992b) défend ainsi l'hypothèse que le langage intérieur n'est pas dépendant d'une expérience corporelle, et inclut des représentations phonologiques abstraites, ni articulatoires, ni auditives.

Oppenheim & Dell (2008) fournissent eux-mêmes des arguments en ce sens. Ils ont étudié en détail les erreurs commises lors de la répétition de virelangues et ont relevé différents biais systématiques. Ils ont ainsi montré que lors de la production répétée de séquences de mots, à voix haute, les erreurs commises comportent un biais lexical : les erreurs de production sont plus souvent des substitutions vers des mots du lexique que vers des non-mots (e.g. "balle dame" donne souvent "dale") plutôt que des non-mots (e.g. "balle nappe" donne rarement "nalle"). Ce biais existe aussi lors de la production de virelangues en parole intérieure. Il a été interprété par les auteurs comme un argument en faveur du modèle à activation interactive de Dell (1986), qui postule des flux d'activation réciproques entre les niveaux phonologique et lexical. Ce biais lexical suggère que la parole intérieure active non seulement des représentations conceptuelles mais aussi lexicales et phonologiques.

Un second biais est observé dans les lapsus, nommé par Dell et collègues « biais de similarité phonémique » et correspondant à un « *biais articulatoire* » : les erreurs vont plus souvent vers des phonèmes qui partagent des traits articulatoires avec le phonème de départ. On trouve ainsi plus fréquemment des erreurs de « balle » vers « malle » (/b/ et /m/ sont deux consonnes occlusives bilabiales voisées, ne se distinguant que par le trait de nasalité) que de « malle » vers « salle » (pour passer de l'occlusive bilabiale voisée /b/ à la fricative alvéolaire non voisée /s/, il faut modifier plusieurs traits articulatoires). Ce biais articulatoire est également expliqué par Dell et ses collègues par des activations réciproques, cette fois entre les niveaux articulatoire et phonologique. Ce biais articulatoire n'a pas été observé en parole intérieure, dans l'étude d'Oppenheim & Dell (2008). Dans une étude ultérieure, Oppenheim & Dell (2010) observent ce biais, mais seulement lorsque la production mentale est accompagnée d'articulation silencieuse. Ils en concluent que la parole intérieure active des représentations abstraites, conceptuelles et lexicales (mots), ainsi que des représentations concrètes, phonologiques (cf. biais lexical) mais que les représentations articulatoires sont appauvries (absence de biais articulatoire lorsque les séquences sont produites mentalement, sans être articulées).

Toutefois, le langage intérieur n'est pas toujours condensé et certaines variantes du langage intérieur sont même complètement développées, incluant des représentations

²⁰ « La qualité apparemment auditive de notre parole interne ne peut pas être attribuée automatiquement à des événements en lien avec un système auditif ou acoustique, ni même, comme nous le verrons, à un quelconque système sensoriel. » (ma traduction).

élaborées, jusqu'aux détails articulatoires et auditifs. En effet, certaines études psycholinguistiques ne répliquent pas les résultats de Oppenheim & Dell et montrent au contraire que la parole intérieure inclut bien une spécification des gestes articulatoires associés aux phonèmes produits mentalement (Corley, Brocklehurst and Moat, 2011 ; Smith, Hillenbrand, Wasowicz, & Preston, 1986 ; Scott, Yeung, Gick and Werker, 2013, voir Løevenbruck *et al.*, 2018 et Løevenbruck, 2018 pour des revues). L'étude de Sun & Peperkamp (2016) apporte également des arguments en faveur d'une implication du système de production de la parole lors de la lecture silencieuse. Les auteurs ont montré que les lecteurs français ont tendance à transformer les séquences de lettres qui ne respectent pas une contrainte phonotactique de leur langue en séquences légales dans leur langue. Ainsi la séquence illicite « tlavier » (/la séquence /tl/à l'initiale d'un mot n'existe pas en français) amorce plus facilement « clavier » que la séquence « plavier ». Cependant cet effet de transformation diminue lorsque les participants effectuent une tâche de suppression articulatoire. Ceci suggère que la transformation de « tlavier » en « clavier » n'a lieu que lorsque les participants peuvent articuler mentalement le mot lu (lorsqu'il n'existe pas de tâche concurrente recrutant le système articulatoire). Cette étude suggère donc que la lecture silencieuse de séquences de lettres engage le système de production de parole intérieure jusqu'aux niveaux phonologique et articulatoire.

En ce qui concerne les représentations auditives, l'effet de transformation verbale indique que la parole intérieure met bien en jeu le système auditif. Lorsque l'on demande à des participants de répéter mentalement le mot "life" en boucle, ceux-ci finissent par entendre "life". Cet effet est réduit en présence d'interférence auditive, ce qui indique que la détection de transformation engage le système auditif (Smith, Wilson & Reisberg, 1995). Nous avons nous-mêmes répliqué cet effet en parole intérieure dans une étude en IRMf (Sato, Baciú, Løevenbruck, Schwartz, Cathiard, Segebarth & Abry, 2004). La recherche de transformation verbale impliquait bien les régions auditives temporales, de façon plus importante que la répétition mentale passive (sans recherche active de transformation). Une étude de neuroimagerie récente suggère même que la parole intérieure durant la lecture inclut des détails auditifs qui vont jusqu'au voisement (Kell *et al.*, 2017). Dans cette étude IRMf, le nombre de phonèmes voisés et non-voisés des phrases lues silencieusement était contrôlé de façon systématique. Les résultats indiquent que l'activation temporelle est modulée par le nombre de segments voisés contenus dans les phrases lues. La lecture intérieure de phrases comportant un nombre plus important de segments voisés correspond à une activation bilatérale plus importante de la partie inférieure du sillon temporal supérieur (région du cortex auditif sélective à la voix) alors que celle de phrases présentant plus de segments non-voisés donne lieu à une activation plus élevée du gyrus temporal moyen postérieur gauche. Ces résultats suggèrent donc que la lecture en parole intérieure inclut même des représentations auditives du voisement, un trait phonétique concret, de bas niveau.

Je présente dans la section II.2.1 des données, issues de travaux sur l'introspection, et des résultats d'études en psycholinguistique ou en neuroimagerie qui permettent de mieux cerner la nature du langage intérieur. Il ressort de ces études qu'il semble tout à fait possible de répéter mentalement des mots ou des phrases, de générer mentalement des phrases à partir de mots (tâche de génération de définitions), de lire mentalement et ceci de façon non condensée : ces opérations mentales sont d'ailleurs accompagnées des corrélats cérébraux typiques de la parole non abrégée, avec des activations des aires auditives, somatosensorielles et motrices, notamment. Il semble donc que certaines

variantes du langage intérieur dépendent de processus perceptivo-moteurs, incluant des détails opérationnels de bas niveau (non abstraits). Ces variantes peuvent être considérées comme incarnées, impliquant des processus physiques qui se déroulent dans le temps et conduisant à la création de percepts articulatoires et auditifs. Elles intègrent probablement, comme la parole à voix haute, une variété de représentations, depuis les concepts sémantiques de haut niveau, jusqu'aux traits articulatoires de bas niveau, en passant par des items lexicaux et des représentations phonologiques.

Encadré I.26. Dimension de condensation

Ainsi, comme Fernyhough (2004) ou Vicente & Martínez-Manrique (2016), on peut considérer que la parole intérieure varie entre des formes condensées et des formes développées. La forme la plus condensée serait dénuée des qualités acoustiques, phonologiques et même syntaxiques de la parole à voix haute et peut donc être décrite comme abstraite, dans sa forme. La parole intérieure développée, à l'opposé, conserve les propriétés phonologiques et phonétiques de la parole à voix haute et est donc de nature concrète (avec des corrélats sensoriels et probablement moteurs). Dans le cadre du modèle de production présenté au I.2.1.1.2, la forme la plus condensée peut être associée au message conceptuel dans sa forme linguistique préliminaire et la forme la plus développée peut être considérée comme issue de la forme condensée, incluant la spécification phonologique et phonétique complète, prête à être articulée. Entre ces deux formes extrêmes, il est probable que des formes intermédiaires, semi-développées (ou semi-condensées) existent, en fonction du niveau auquel le processus de production de la parole est interrompu.

En ce qui concerne le degré de dialogalité, il a été défendu que la parole intérieure peut prendre la forme d'un monologue, mais peut s'exprimer sous forme de dialogue ou de conversation avec de multiples interlocuteurs, reflétant les multiples formes de communication à voix haute (Philippe, 2001 ; Alderson-Day *et al.*, 2016).

Parole intérieure dialogale ou dialogique ?

Il existe en pragmatique française une nuance entre les termes de « dialogal » et « dialogique » qui fait débat (Bres, 2005 ; Bres & Dufour, 2014 ; voir aussi Maingueneau, 2016) et qui semble moins utilisée en anglais (mais voir Roulet & Green, 2006). Je me réfère à Bres (2005) pour la distinction entre les deux, et, à sa suite, je pose que « dialogal » (« dialogical » ou « dialogal » en anglais) suppose deux interlocuteurs en présence, avec une alternance de tours de parole (« monologal » concernant un·e unique locuteur·rice) alors que « dialogique » (« dialogic » en anglais) désigne un énoncé en interaction avec d'autres énoncés, d'autres voix, d'autres discours. Ainsi, un discours peut être monologal mais dialogique, s'il est tenu par un·e seul·e locuteur·rice qui se positionne par rapport à des discours antérieurs d'autrui (comme : « Oui, le climat change », titre d'un éditorial du *Monde*, 2000, analysé par Bres & Dufour, 2014). Dans le cas de la parole intérieure, il me semble que l'hypothèse défendue par Fernyhough (2004), en référence aux travaux de Vygotski (1934), est que celle-ci est de nature « dialogique ». Le terme anglais de « dialogic » est ambigu, mais la description de Fernyhough correspond à une conception où la parole intérieure contient des perspectives différentes (donc est dialogique), mais n'est pas systématiquement dialogale dans sa forme, surtout lorsqu'elle est condensée, abrégée.

« *Mature inner speech is an ongoing dialogue between [...] internalised, simultaneously held perspectives. [...] inner speech is considerably abbreviated relative to external dialogue, and only occasionally manifests the dialogic 'give-and-take' structure of conversation.* »²¹ Fernyhough (2004, p. 53).

Par conséquent, Fernyhough (2004), dans la lignée de Vygotski, défend l'idée que la parole intérieure est de nature dialogique et Alderson-Day *et al.* (2016) montrent que sa forme peut être parfois dialogale (lorsque l'on imagine une conversation) et parfois monologale (lorsque l'on imagine faire une conférence).²²

De plus, lorsqu'elle est monologale, elle peut se faire avec notre voix propre, mais aussi avec une autre voix que la nôtre. Nous sommes ainsi capables d'imiter la voix d'autrui mentalement (cf. section II.2.1 ou Grandchamp *et al.*, soumis). Lorsqu'elle est dialogale, dans le cas par exemple de dialogues ou de conversations anticipés ou revécus mentalement, nous pouvons alors non seulement modifier notre voix pour prendre celle d'autrui, mais aussi changer de perspective et ainsi imaginer qu'autrui est en train de s'adresser à nous (Shergill *et al.*, 2002 ; Fernyhough, 2004 ; Alderson-Day *et al.*, 2016).

Encadré I.27. Dimension de dialogalité

La parole intérieure, de nature probablement essentiellement dialogique, peut être de forme monologale, i.e. prendre la forme d'un monologue intérieur (avec notre voix ou celle d'autrui), ou dialogale, i.e. se dérouler sous la forme d'une conversation imaginée, avec une ou plusieurs personnes, nécessitant la capacité de gérer plusieurs voix et plusieurs perspectives.

Enfin, en ce qui concerne le degré d'intentionnalité, comme nous l'avons déjà mentionné au paragraphe I.1.3.2.6, nous pouvons volontairement parler dans notre tête (compter, faire une liste, planifier nos objectifs, par exemple). Nous pouvons aussi former délibérément des fragments de langage intérieur plus longs, comportant des phrases entières, ou des discours, lorsque nous préparons mentalement un exposé, un cours, lorsque nous imaginons une conversation future ou répétons une conversation passée, lorsque nous cherchons à résoudre un problème. Cette variante de langage intérieur est intentionnelle. Mais il arrive que notre langage intérieur soit plus spontané, moins actif, plus passif, lors du *vagabondage mental verbal* (Perrone-Bertolotti *et al.*, 2014). Cette autre variante semble faire irruption et ne pas être délibérée, être moins sous contrôle cognitif. Il est souvent rapporté, notamment par les écrivains et les poètes, que le langage intérieur nous vient parfois accompagné d'un sentiment d'imprévisibilité, de non-intentionnalité, voire d'inagentivité (cf. la lettre de Rimbaud (I.1.3.2.1 ci-dessus), les témoignages d'écrivains consignés par Fernyhough (2016) et nos propres questionnaires non publiés, issus de l'impromptu scientifique « Des voix dans la tête », avec le groupe n+1 des Ateliers du Spectacle, ainsi que de travaux en cours avec Stéphanie Smadja dans le cadre du projet *Monologuer* qu'elle dirige à l'Université Paris Diderot).

²¹ « La parole intérieure mature est un dialogue en cours entre différentes perspectives internalisées et prises simultanément. [...] la parole intérieure est considérablement abrégée par rapport au dialogue externe, et ne se montre qu'occasionnellement la structure dialogale (constituée de tours de parole) typique de la conversation. » (ma traduction).

²² Alderson-Day *et al.* (2018) dans un article sur le nouveau questionnaire VISQ (*Varieties of Inner Speech Questionnaire*) notent d'ailleurs que le terme de « dialogic » est ambigu en anglais, mais que l'acceptation vygotkienne est celle d'une dialogicalité de la structure de la parole intérieure (pas de la forme).

Est-ce notre héritage culturel qui nous conduit à considérer parfois nos pensées verbales comme non-délibérées, qui leur imprime ce caractère récitatif, comme si nous étions les réceptacles des muses ? Laissons une grande poétesse nous éclairer :

La Muse

*Quand je l'attends au sein de la nuit noire,
La vie, alors, ne compte plus pour rien ;
Qu'importent liberté, jeunesse, gloire,
Lorsque, sa flûte aux doigts, elle **survient**.
Oui, la voilà **brusquement dévoilée**...
Fixé sur moi, son regard m'éblouit
Et je demande : « Est-ce sous ta **dictée**
Que Dante fit l'Enfer ? » Elle dit : « Oui. »*

Anna Akhmatova (1924/1993, traduction J. Malaplate, c'est moi qui souligne).

Ainsi, Calliope ou Euterpe semblent parfois guider nos déambulations verbales intérieures ou, en termes plus contemporains, le langage lui-même paraît susciter la création verbale, ce que joue Bergounioux dans *Dominos* (2014, p.7) :

« Chaque phrase, à sa façon, prolonge la précédente. L'enchaînement conduit à un nouveau thème, pas tellement prévisible (même pour moi) en sorte qu'à la fin, ça parle de quelque chose d'autre d'où part la phrase suivante qui le déporte (...) Un maître conduit le jeu : le langage. C'est en lui, pour lui et avec lui qu'on écrit ; sous sa dictée. »

Nous touchons ici aux limites de l'agentivité, ce mécanisme subtil qui nous donne le sentiment d'être l'auteur·e de nos actions et en particulier de nos verbalisations intérieures, discuté dans la section I.2.3.

Encadré I.28. Dimension d'intentionnalité

Le langage intérieur peut être intentionnel, c'est-à-dire délibéré et contrôlé, ou bien non-intentionnel. Il semble alors survenir de façon vagabonde, involontaire, imprévisible, sans que nous ayons l'impression de l'avoir déclenché ni de le contrôler et avec la sensation que les mots s'immiscent en nous, se jouent de nous : le jeu des mots, les jeux de mots.

Ainsi, le langage intérieur présente plusieurs variantes qui se distinguent par leurs degrés de condensation, de dialogalité et d'intentionnalité et qui sont illustrées sur la Figure I-14.

Sur l'axe vertical représentant la dimension de condensation, les variantes les plus condensées (pavés du haut) engagent uniquement les processus de *monitoring* verbal de plus haut niveau (sémantique, voire lexical), alors que les variantes les plus développées (pavé du bas) requièrent des processus de *monitoring* verbal jusqu'au niveau de la planification articulatoire, impliquant la perception et le contrôle d'une petite voix.

L'axe horizontal représente la dimension d'intentionnalité. Les variantes les plus vagabondes et « irruptives » sont représentées sur la gauche du schéma. Selon moi, elles engagent moins les processus de *monitoring* verbal que les variantes les plus délibérées (les flèches de rétroaction ne sont donc pas indiquées), alors que les variantes délibérées mettent en jeu un suivi et des corrections d'erreur à tous les niveaux de la hiérarchie (partie droite du schéma avec flèches de rétroaction). Il est probable que les formes

développées surviennent surtout lors de la parole intérieure délibérée. Le vagabondage mental verbal est souvent rapporté comme peu audible, évanescent. Il se peut donc que les versions les plus vagabondes n'atteignent pas le dernier stade de développement (la petite voix audible) et stagnent au niveau condensé (représentation conceptuelle pré-verbale) ou semi-condensé (représentation lexico-syntaxico-phonologique). Il semble toutefois que, dans la rêverie et le vagabondage, l'on puisse entendre parfois la petite voix (ce que rapporte Hurlburt, 2011). Il reste donc une incertitude sur l'existence (ou les conditions d'occurrence) de la variante développée de la parole intérieure non délibérée.

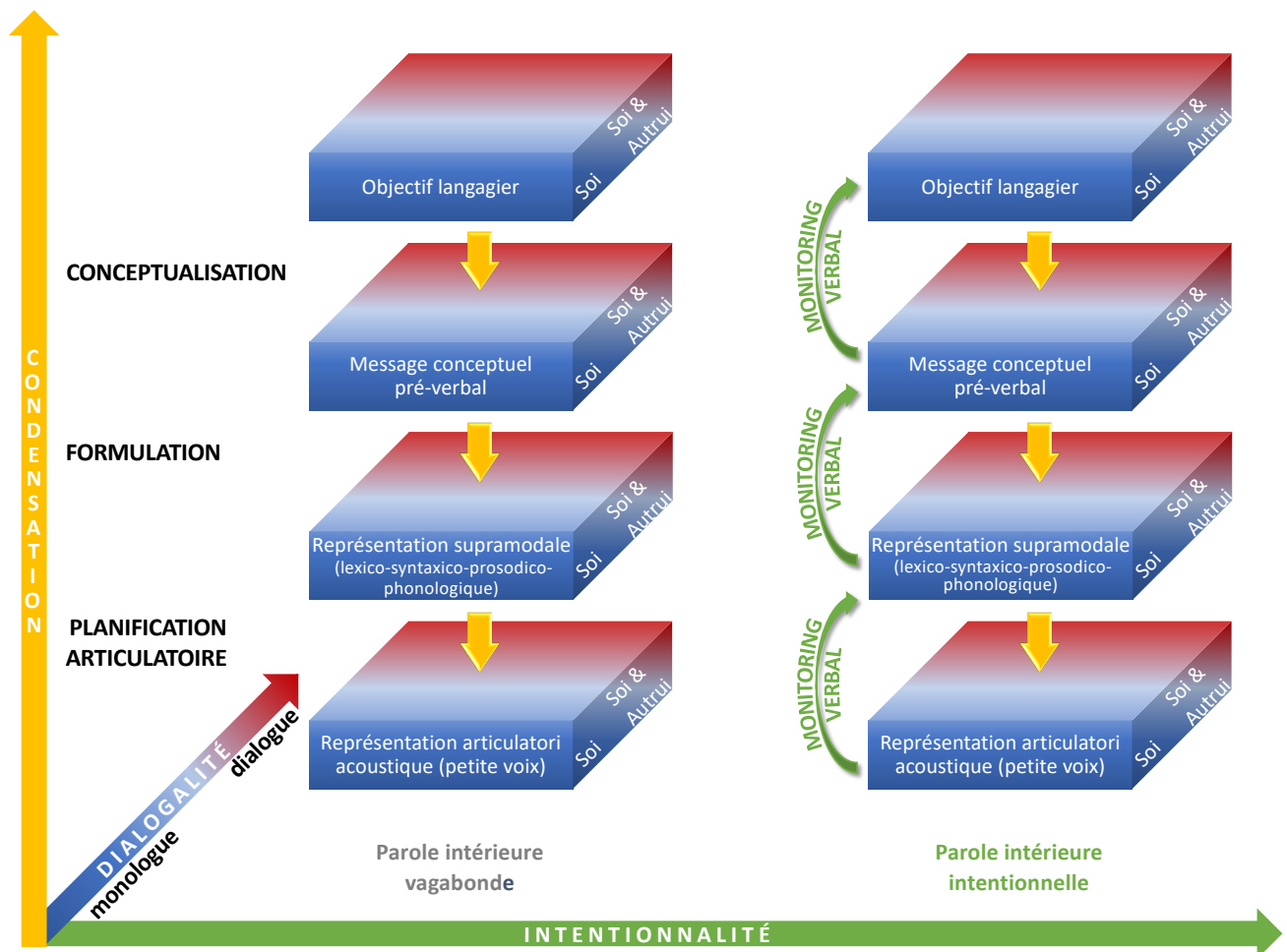


Figure I-14. Trois dimensions de la parole intérieure : condensation, dialogalité et intentionnalité

Sur l'axe de l'intentionnalité on passe de la parole intérieure vagabonde (à gauche) à la parole intérieure délibérée (à droite), sur l'axe de la condensation, on passe des versions condensées (en haut) aux versions développées jusqu'à la petite voix (en bas), et sur l'axe de la dialogalité, on passe du monologue (au premier plan, en bleu) au dialogue (à l'arrière-plan, en rouge).

La troisième dimension représente la variation sur l'axe de la dialogalité, avec, à un extrême, le monologue utilisant sa propre voix (noté « Soi ») et à l'autre, le dialogue qui inclut l'imagination d'autrui parlant (noté « Autrui »). Entre les deux, se trouve le monologue avec d'autres voix que la voix propre. La dialogalité nécessite un autre type de contrôle que le *monitoring* verbal. Elle requiert de savoir qui est en train de verbaliser intérieurement, i.e. qui est l'agent de l'action de parole intérieure. Je traite de cette attribution de l'agentivité dans la section 1.2.3. La façon dont le *monitoring* verbal peut s'opérer en parole intérieure, en fonction de ces trois dimensions, est détaillée ci-après.

I.2.2.2. Le *monitoring* verbal intérieur : suivi du contenu de nos productions intérieures

De façon introspective, nous pouvons nous rendre compte que, comme la parole à voix haute, l'endophasie nécessite un mécanisme de *monitoring* verbal, nous permettant de suivre le déroulement de la construction de nos productions verbales mentales et de les ajuster ou corriger le cas échéant, par exemple lorsque nous raisonnons pour nous-mêmes, lorsque nous préparons une intervention orale future destinée à autrui ou lorsque nous imaginons librement une conversation passée ou future.

Données empiriques sur le *monitoring* verbal intérieur

Ces impressions subjectives sont corroborées par des données empiriques. Les études psycholinguistiques sur la détection de lapsus ou de transformations verbales montrent en effet que des mécanismes de *monitoring* sont en jeu, même en parole intérieure.

Les erreurs de production de parole observées à voix haute (cf. I.2.1.1) peuvent se retrouver en parole intérieure et plusieurs travaux montrent qu'elles peuvent bien être détectées par les locuteurs. En particulier, les études de Dell & Repka (1992), Oppenheim et Dell (2008, 2010) ainsi que Corley *et al.* (2011) montrent en effet qu'il est possible de demander à des participants de répéter mentalement des virelangues et que ceux-ci sont capables de détecter et rapporter les erreurs commises (qui sont du même type que celles qu'ils rapportent à voix haute). L'effet de transformation verbale, décrit au I.2.2.1 ci-dessus, indique, lui aussi, qu'il est possible de détecter mentalement des transformations lorsque l'on imagine qu'un stimulus (syllabes ou mot) est répété en boucle par un ami (Reisberg *et al.*, 1989 ; Smith *et al.*, 1995) ou lorsque l'on répète soi-même mentalement le stimulus (Sato *et al.*, 2004). Ceci indique donc que les participants peuvent détecter des erreurs de production interne, en situation de monologue ou de dialogue intérieur, en tout cas, lorsqu'on leur demande explicitement de le faire.

Tout ceci suggère qu'un mécanisme de suivi de nos productions verbales opère même pour nos productions mentales (monogales et dialogales), en tout cas lorsqu'elles sont délibérées. Le niveau auquel ces erreurs sont détectées varie selon les auteurs. Pour Oppenheim et Dell (2008, 2010 ; cf. aussi Dell & Oppenheim, 2015), les erreurs sont détectées au niveau phonologique, puisqu'on observe bien un biais lexical mais pas un biais articulatoire. Pour Corley *et al.* (2011), les erreurs peuvent également être détectées plus bas dans la hiérarchie, i.e. au niveau de la représentation articulatoire-acoustique. Les recherches sur l'effet de transformation verbale ayant montré que les erreurs sont « entendues », i.e. que des processus auditifs sont engagés (cf. I.2.2.1 ci-dessus), corroborent l'hypothèse que le *monitoring* peut s'effectuer à ce niveau. On a donc des arguments pour faire l'hypothèse que, lors de la production de parole intérieure, le *monitoring* s'exerce au moins aux deux niveaux inférieurs de la hiérarchie. Il n'existe pas à ma connaissance, d'études montrant que le *monitoring* s'applique aussi au niveau de la représentation conceptuelle pré-verbale. Toutefois, l'introspection suggère qu'il est possible de s'auto-corriger lors de la génération d'un message pré-verbal, lorsque l'on cherche à exprimer mentalement une idée, lorsque l'on réfléchit. Ainsi, je propose que, dans le cas de la parole intérieure délibérée, comme indiqué par les flèches de rétroaction sur la Figure I-14, la production intérieure puisse être contrôlée et corrigée aux différents niveaux hiérarchiques, de la représentation articulatoire-acoustique à la représentation conceptuelle.

Toutefois, il se peut que ce *monitoring* soit moins sévère que dans le cas de la parole à voix haute. En effet, comme l'ont noté Egger (1881), Vygotski (1934) ou Bergounioux (2001a), la parole intérieure n'a besoin d'être comprise que de nous-mêmes, et nous pouvons donc accepter des formes phonologiquement, lexicalement, syntaxiquement abrégées, voire erronées, du moment que le sens est conservé (cf. I.2.2.1). Wiley (2014) propose même que les processus de contrôle soient différents entre parole à voix haute et intériorisée. Selon lui, la contrainte la plus impérieuse en parole intérieure est celle de l'efficacité, visant à la rapidité et l'économie. Cette contrainte a probablement pour conséquence de relâcher le *monitoring*, au moins aux niveaux inférieurs de la hiérarchie. Un deuxième relâchement qui s'opère en parole intérieure est celui du contrôle des émotions. Selon Wiley, la parole intérieure est beaucoup plus émotive que la parole à voix haute. Lorsque nous répétons mentalement un énoncé que nous comptons prononcer à voix haute, nous avons tendance à modérer le ton, à réprimer nos émotions, alors qu'en parole intérieure, nous pouvons laisser libre cours à nos émotions, personne n'étant témoin. Le *monitoring* en parole intérieure est donc probablement moins strict, plus approximatif, qu'en parole à voix haute (d'où les symboles d'approximation au niveau des comparaisons entre signaux désirés et prédits sur la Figure I-15).

Dans le cas du vagabondage mental verbal, il n'existe pas, à ma connaissance, d'études montrant que des mécanismes de *monitoring* verbal sont en jeu. Il me semble que, par définition, le vagabondage mental opère sans contrôle exécutif, ou avec un contrôle intermittent. Le *monitoring* verbal doit donc être réduit, d'où l'absence de flèches d'auto-correction sur le schéma de gauche de la Figure I-14. Je reviens sur la question du *monitoring* de la parole intérieure vagabonde dans la section I.2.3.

Encadré I.29. *Monitoring verbal intérieur*

Un mécanisme de suivi du contenu de nos verbalisations opère même pour nos productions intérieures (qu'elles soient monologiques ou dialogales), en tout cas lorsqu'elles sont délibérées. Ce *monitoring* verbal intérieur est probablement moins strict que dans le cas de la parole à voix haute. Dans le cas de la parole intérieure vagabonde, le *monitoring* verbal est probablement inexistant.

Modélisation du *monitoring* verbal intérieur : modèle Condialint

Dans Grandchamp *et al.* (soumis), nous proposons un modèle fonctionnel neuroanatomique de la production de parole intérieure, qui vise à rendre compte du *monitoring* des différentes variantes listées ci-dessus. Le modèle Condialint s'appuie sur la modélisation neurocognitive de la dernière étape de la production de parole intérieure délibérée (la planification articulatoire : de l'objectif phonétique supramodal à l'expérience sensorielle d'une petite voix) que nous avons proposée dans Løevenbruck *et al.* (2018). Il reprend également les principes et hypothèses présentés ci-dessus sur le *monitoring* hiérarchique de la production de parole à voix haute (cf. I.2.1.1.2, Figure I-9, Figure I-10 et Figure I-12) intégrant la notion de contrôle prédictif. Il est provisoirement restreint à la production de *parole* intérieure et ne décrit pas la production de *signe* intérieur, par manque de données sur la production mentale en langue des signes. Je fais simplement la spéculation que les processus auditifs invoqués dans ce modèle provisoire pourraient être remplacés par des processus visuels, pour rendre compte de la production de signes intérieurs.

Le modèle Condialint, illustré sur la Figure I-15, rend compte des trois dimensions de parole intérieure.

En ce qui concerne la dimension de condensation, le modèle suppose que la production délibérée de parole intérieure engage les processus de la production de langage à voix haute, mais que des processus inhibiteurs ont lieu, qui peuvent interrompre la production à plusieurs stades, en fonction du degré de condensation. Des mécanismes de contrôle cognitif inhibiteur (flèche orangée sur la Figure I-15) opèrent pour interrompre la production aux différents niveaux. L'inhibition appliquée après la conceptualisation donne la variante la plus condensée de parole intérieure. Après la formulation, elle donne la variante semi-développée et, après la planification articulatoire, la variante la plus développée, soit la petite voix dans la tête. Plus précisément, les processus de contrôle prédictif décrits dans les modélisations de la Figure I-9 et de la Figure I-12 (cf. section I.2.1.1.2 ci-dessus) sont accompagnés de processus inhibiteurs. L'objectif langagier initial est d'abord transformé en message pré-verbal puis, le cas échéant, en objectif phonétique supramodal : les deux premières variantes de la parole intérieure, sur l'axe de la condensation. Dans le cas de la variante la plus développée, l'objectif phonétique est ensuite converti en commandes motrices par un contrôleur couplé à un programmateur. Ces commandes motrices sont inhibées et ne sont pas envoyées à l'appareil sensorimoteur (ou bien sont envoyées en parallèle de l'envoi de commandes inhibitrices qui les annulent). Par un mécanisme prédictif, des actes moteurs (phonation, articulation) sont simulés, à partir d'une copie des commandes motrices (la copie d'efférence). Ces actes multimodaux simulés donnent lieu à des percepts multisensoriels mentaux : des voix, des sensations proprioceptives, tactiles, la parole intérieure développée.

En ce qui concerne la dimension d'intentionnalité, dans le cas de la parole intérieure délibérée (notée <v_s> : imagination de *self verbalising*, cf. I.2.4), le *monitoring* verbal peut s'opérer à chacun des niveaux de condensation. Les signaux générés par les contrôleurs à chaque niveau (à partir de signaux désirés) sont convertis en signaux prédits ascendants qui sont transmis au niveau hiérarchique supérieur pour être comparés aux signaux désirés. La comparaison est probablement moins stricte qu'en parole à voix haute (cf. ci-dessus), mais elle nécessite de maintenir et temporiser les signaux désirés et prédits pour effectuer les comparaisons adaptées. A chaque niveau, en cas de différence entre les signaux désirés et prédits, un signal d'erreur est émis. Comme pour la production de parole à voix haute, et en fonction des contraintes et buts de la parole intérieure en cours, la paire de contrôleur/prédicteur peut alors être ajustée.

Dans le cas de la parole intérieure vagabonde, le *monitoring* verbal est probablement inexistant : l'éventuel signal d'erreur issu de la comparaison entre signaux désirés et prédits ne serait pas utilisé pour ajuster la production en cours. Comme je l'explique dans la section I.2.3, le mécanisme prédictif ne sert pas uniquement à contrôler que les objectifs désirés sont atteints en termes de contenu verbal. Il joue également un rôle dans l'agentivité, i.e. le sentiment d'être l'agent, l'auteur-e de nos actions. Dans Rapin *et al.* (2013, 2016) et Lœvenbruck *et al.* (2018), nous avons proposé que, dans le cas de la parole intérieure délibérée, l'agentivité résulte de la comparaison entre signaux prédits et désirés (cf. I.2.3). Ainsi, durant le vagabondage mental, si l'on fait l'hypothèse qu'il n'y a pas de *monitoring* verbal, et donc pas de comparaison, il n'y aurait alors pas de sensation d'agentivité. C'est ce qui est parfois rapporté sur la parole vagabonde et les pensées spontanées (*unbidden thoughts*, cf. Gallagher, 2004).

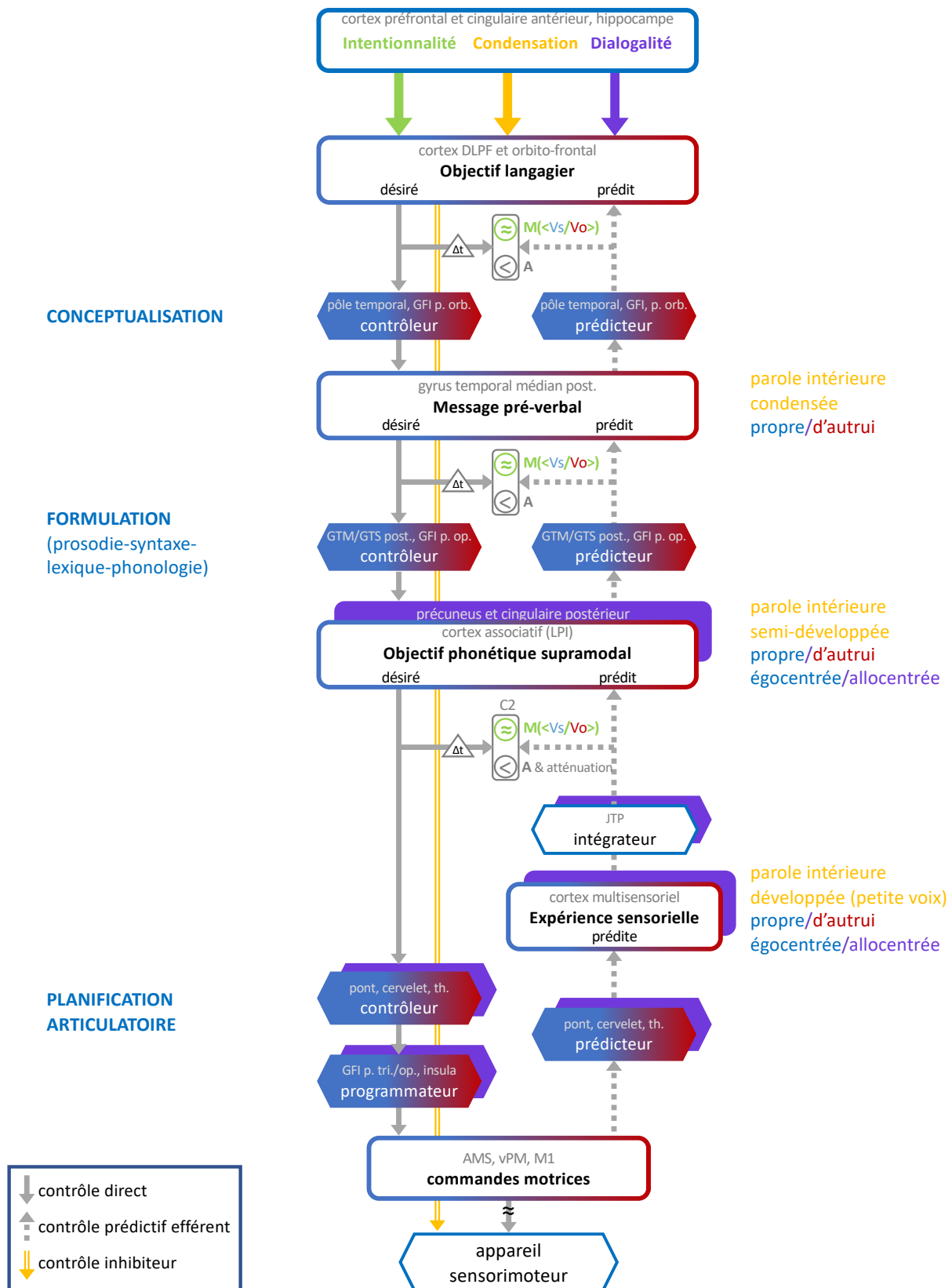


Figure I-15. ConDialInt : Modèle hiérarchique de la production de parole intérieure
Les variantes sur l'axe de condensation émanent du contrôle inhibiteur (flèche orange).
L'agentivité (A) opère dans tous les cas, le monitoring (M), moins strict qu'à voix haute,
uniquement dans le cas délibéré. La voix propre vs. d'autrui est figurée par le dégradé bleu-
rouge sur les prédicteurs et contrôleurs. La perspective allocentrique dialogale correspond aux
boîtes violettes de l'étape de planification articulatoire.

Je reviendrai sur ce point au I.2.3, car une hypothèse alternative peut être formulée, séparant le *monitoring* des verbalisations (noté $M(<v_s>)$ et $M(<v_o>)$, pour la parole intérieure de soi et d'autrui) de l'attribution d'agentivité (notée A). Il est donc concevable qu'une comparaison ait bien lieu entre signaux désirés et prédits, mais qu'elle soit uniquement utilisée à des fins d'attribution d'agentivité (et non de *monitoring* verbal). Cette hypothèse est plus compatible avec les explications actuelles des hallucinations auditives verbales (cf. section I.2.3). C'est celle qui est indiquée sur la Figure I-15 : le *monitoring* du contenu du flux verbal, $M(<v_s/v_o>)$, n'aurait lieu que pour la parole intérieure délibérée (comparaison du contenu des signaux désirés et prédits, symbolisée de couleur verte), alors que l'attribution d'agentivité, A , aurait lieu pour les variantes délibérées comme vagabondes (comparaison de l'agencement temporel des signaux désiré et prédit, couleur grise).

Encadré I.30. Monitoring et agentivité lors de la parole intérieure non délibérée ?

Par définition, le vagabondage mental verbal est involontaire et les verbalisations intérieures semblent surgir de façon incontrôlée. Il semble donc inapproprié de concevoir que ces verbalisations vagabondes fassent l'objet d'un *monitoring* verbal visant à vérifier qu'elles sont conformes aux objectifs désirés. Toutefois, il se peut qu'une attribution de l'agentivité soit à l'œuvre, car même si la parole intérieure apparaît souvent comme spontanée et imprévisible, dans le cas le plus fréquent, nous nous reconnaissons finalement comme les auteurs de ces productions verbales. Je reviens sur ce point dans la section I.2.3. Cette question du *monitoring* verbal et de l'attribution d'agentivité dans le cas de la parole intérieure non délibérée reste encore à éclaircir.

En ce qui concerne la dimension de dialogalité, le modèle rend compte de l'utilisation de multiples voix par l'adaptation des modèles internes prédicteurs et contrôleurs. Comme nous l'avons mentionné ci-dessus, nous pouvons avoir des conversations intérieures, au cours desquelles nous pouvons imaginer la voix d'autrui. Cette production mentale s'appuie probablement sur les mêmes mécanismes prédictifs qui nous permettent de simuler les productions d'autrui dans le cas de la compréhension de la parole. Nous utilisons dans ce cas des modèles d'autrui paramétrés à partir de nos modèles propres (utilisés pour la production de notre voix) et adaptés en fonction des connaissances que nous avons d'autrui, son timbre, son accent, ses habitudes lexicales, syntaxiques, etc. Cette possibilité d'utiliser des modèles internes adaptés à autrui est représentée sur la Figure I-15 par un dégradé de couleurs (de bleu pour les modèles propres à rouge pour les modèles d'autrui) dans les modèles contrôleurs et prédicteurs. Ce dégradé vise à rendre compte de la possibilité de génération mentale de la voix d'autrui (notée $<v_o>$: imagination de « *other verbalising* »). Les variantes de parole intérieure qui résultent de l'utilisation de modèles adaptés sont schématisées par des boîtes entourées d'un dégradé de bleu et rouge. Il apparaît donc qu'au *monitoring* du contenu de la parole intérieure, il faut ajouter des dispositifs de contrôle exécutif de plus haut niveau permettant de passer d'une voix imaginée à une autre (pour paramétrer les modèles prédicteurs et contrôleurs). Par ailleurs, la dialogalité suppose plus que des dispositifs nous permettant d'imiter mentalement la voix d'autrui. Pour l'imagination de conversations, il faut pouvoir prendre une perspective égocentrée ou allocentrée : s'imaginer parlant, avec une perspective à la 1^{ère} personne (égocentrée) vs. imaginer autrui parlant, avec une perspective à la 3^{ème} personne (allocentrée). Il a été montré que

des mécanismes cognitifs dédiés sont en jeu lorsque l'on demande à des participants d'imaginer être l'agent d'une action, comparé à lorsqu'ils imaginent qu'autrui est l'agent (Ruby & Decety, 2001). Il se peut donc que le fait d'imaginer autrui parler recrute des mécanismes de changement de perspectives similaires à ceux observés pour les actions en général, ainsi que des processus de contrôle exécutif spécifiques. Il me semble que la prise de perspective allocentrée, le fait d'imaginer qu'autrui parle, joue essentiellement un rôle dans l'étape ultime de la production de parole intérieure, celle qui met en jeu des représentations sensorielles, i.e. durant la planification articulatoire. L'imitation de la voix, elle, affecte tous les niveaux de la production. La conceptualisation et la formulation peuvent en effet être initiées en ayant pour objectif de produire de la parole avec les spécificités linguistiques de soi-même ou d'autrui. Elles requièrent donc des mécanismes qui permettent d'utiliser les caractéristiques langagières de soi-même ou d'autrui. Mais le changement de perspective (substituer l'agent de l'acte de parole imaginé) ne nécessite pas de modifier les deux premiers niveaux (conceptualisation et formulation). Seule la planification articulatoire est affectée par un changement de perspective (de la 1^{ère} à la 3^{ème} personne), car un système de coordonnées corporelles allocentré doit alors être utilisé. Des boîtes de couleur violette ont été ajoutées sur la Figure I-15 au niveau de la planification articulatoire pour rendre compte du mécanisme de changement de perspective requis pour le dialogue intérieur.

Les corrélats cérébraux supposés de la production de parole sont également indiqués sur Figure I-15. En plus de ceux impliqués dans la production de parole à voix haute, détaillés ci-dessus (cf. I.2.1.1.2 et Figure I-12), la parole intérieure met en œuvre des réseaux liés au contrôle cognitif, pour inhiber l'exécution motrice, pour la variante la plus développée ou pour arrêter (ou ne pas déclencher) le processus de production avant l'étape d'articulation pour les variantes condensées (dimension de condensation). Des processus de contrôle cognitif sont aussi à l'œuvre pour déclencher l'adaptation des modèles internes prédicteurs et contrôleurs à chaque niveau, lorsque différentes voix sont imaginées (dimension de dialogalité) et pour lancer les comparaisons entre signaux désirés et prédits liés au *monitoring* verbal dans le cas de la parole intérieure délibérée (dimension d'intentionnalité). Un niveau de contrôle cognitif a donc été ajouté par rapport au modèle sur la production de parole à voix haute, avec le recrutement des cortex préfrontal et cingulaire antérieur. L'utilisation de modèles internes adaptés à autrui requièrent probablement en outre des processus de récupération mnésique de la qualité de la voix et des caractéristiques linguistiques de l'autrui imaginé. L'hippocampe a donc également été ajouté. En ce qui concerne le changement de perspective sur l'action, il a été montré qu'imaginer qu'autrui est l'agent d'une action, comparé à imaginer être soi-même l'agent induit des réponses dans le lobule pariétal inférieur droit, le précuneus, le cortex cingulaire postérieur et le cortex fronto-polaire (Ruby & Decety, 2001). Ces régions sont censées jouer un rôle important dans la prise de perspective soi/autrui (Decety, 2005). Elles sont donc ajoutées au niveau de l'entrée dans la planification articulatoire, qui serait le lieu où le changement de perspective pourrait s'enclencher. Decety & Grèzes (2006) ont également montré que le LPI droit est impliqué dans l'attribution des actions, émotions et pensées à leurs agents respectifs lorsque l'on simule mentalement des actions exercées par soi-même ou par autrui. Leur revue de la littérature montre qu'il est difficile de préciser si la région cruciale dans ce processus est la partie rostrale du LPI droit ou la jonction temporo-pariétale droite. Je reste donc imprécise sur ce point. Les boîtes violettes de la Figure I-15, qui représentent les opérations de construction de l'objectif phonétique, de traitement de l'expérience sensorielle et d'intégration multisensorielle

rendent compte du changement de latéralisation qui semble s'opérer (du LPI et de la JTP gauche vers leurs homologues droits) lorsque l'on imagine autrui parler.

Encadré I.31. $M_S(<v_S>)$, $M_S(<v_S>)$, et $M_S(<v_O>)$: monitorings verbaux intérieurs égocentrés et allocentré

En résumé, la pensée verbale met en jeu des mécanismes de *monitoring* égocentré de nos verbalisations intérieures, nous permettant de suivre et contrôler le contenu de notre pensée. Ces mécanismes s'appuient sur le *monitoring* verbal égocentré déployé habituellement pour la parole à voix haute, mais sont exercés de façon moins rigoureuse et plus souple. Ils nécessitent aussi un contrôle exécutif de haut niveau permettant de générer mentalement différentes voix et de prendre des perspectives égocentrée ou allocentrée. Pour distinguer ce sous-système de *monitoring* des autres, j'utiliserai par la suite (cf. I.2.4) le terme de *monitoring* verbal intérieur.

Celui-ci se décline en deux *monitoring verbaux intérieurs égocentrés*, avec notre propre voix ou $M_S(<v_S>)$ ou une voix modifiée $M_S(<v_S>)$, et un *monitoring verbal intérieur allocentré* ou *monitoring* de l'imagination d'autrui verbalisant, noté $M_S(<v_O>)$. Il serait surtout utilisé dans le cas de la parole intérieure délibérée et serait probablement inopérant (ne donnant pas lieu à un signal d'erreur) dans le cas de la parole vagabonde.

Ce *monitoring* verbal intérieur, dans ses différentes déclinaisons, serait accompagné d'une attribution de l'agentivité (décrite au I.2.3), qui serait systématiquement en œuvre dans les variantes délibérées et parfois opérante dans les variantes vagabondes.

I.2.3. L'attribution de l'agentivité dans l'auto-évaluation

La dernière fonction du langage considérée ici, l'auto-évaluation, engage elle-même des mécanismes de contrôle et notamment un suivi de soi ou *self-monitoring*²³. Elle nécessite un dispositif nous permettant de reconnaître que nos actions et en particulier les productions verbales que nous générons à voix haute et surtout celles que nous imaginons (délibérément ou non) et que nous percevons, ont bien été produites par nous-mêmes, que nous en sommes les auteur·e·s.

Encadré I.32. Agentivité : subjectivité vs. altérité

Je nomme ce mécanisme « l'attribution de l'agentivité » et je le note A. Ce dispositif imprime aux actions perçues une sensation de *subjectivité* quand nous en sommes les auteurs et *altérité* lorsqu'autrui en est auteur. Le terme de « sens de l'agentivité » est utilisé pour définir le sentiment d'être l'auteur ou l'agent d'une action, donc de s'auto-attribuer l'agentivité. Il porte cependant à confusion (l'agentivité peut être la nôtre ou celle d'autrui, comme la propriété). Dans mes modélisations et descriptions, j'utilise donc le plus possible « subjectivité » (vs. « altérité »), mais la plupart des auteurs utilisent « agentivité » pour « sens de l'agentivité ».

²³ Il s'agit bien ici d'un suivi de soi, d'un contrôle de qui est l'agent de l'action, et non d'un « verbal self-monitoring », qui est un suivi verbal égocentré, i.e. un suivi des verbalisations auto-produites, un contrôle du contenu verbal et des éventuelles erreurs d'énonciation.

La conscience de soi s'appuie sur deux composantes essentielles : le sens de l'agentivité et le sens de la propriété (Gallagher, 2000). Le sens de l'agentivité selon Gallagher (2000) est la conscience de causer l'action, d'en être l'auteur-e, d'en être à l'initiative, de la contrôler. Le sens de la propriété est la conscience d'être le sujet de l'action en cours, de posséder le corps qui est en mouvement. Dans le cas de mouvements involontaires (par exemple, lorsque notre menton est tiré vers le bas par un dentiste, sans notre volonté), nous pouvons sentir que nous sommes déplacés involontairement et nous savons que c'est notre propre corps qui est manipulé. Nous avons donc le sens de la propriété du corps en mouvement. Mais nous n'avons pas le sens de l'agentivité du mouvement car nous ne l'avons pas initié nous-mêmes. Dans le cas d'un mouvement volontaire, par contre, nous avons à la fois le sens de l'agentivité et de la propriété. Il en serait de même pour le langage : un mécanisme d'attribution d'agentivité nous donne un sens de l'agentivité des sons (ou gestes) produits et permet à notre système auditif (ou visuel ou somatosensoriel) de distinguer les énoncés d'autrui de ceux que nous avons nous-mêmes produits, nous permettant ainsi d'avoir des conversations fluides avec nos congénères.

Pertinence de l'attribution d'agentivité en parole intérieure

L'attribution de l'agentivité semble également à l'œuvre dans le cas de l'endophasie. Certains chercheurs ont questionné la pertinence du système d'attribution d'agentivité pour la parole intérieure, celle-ci n'étant pas, contrairement à la parole à voix haute ou à nos autres actions motrices, en présence d'autres stimulations similaires venant de l'extérieur : *"we are never confronted with the problem of having to sort out our own thoughts from other people's thoughts"* (« Nous ne sommes jamais confrontés au problème de devoir faire le tri entre nos pensées et celles d'autrui » ; Stephens & Graham, 2000, p. 138 ; cf. aussi Gallagher, 2004).

Cette question est résolue par Jones & Fernyhough (2007a) qui inscrivent la nécessité d'un système d'attribution d'agentivité en parole intérieure dans une perspective développementale. Selon eux, et en accord avec les hypothèses de Vygotski, les très jeunes enfants commencent par penser à voix haute (*parole privée, ou discours privé extériorisé*) en simulant des dialogues avec autrui et la pensée verbale ne deviendrait parole intérieure qu'après quelques années, au cours d'un processus graduel d'internalisation. Lors de ce processus, il reste crucial pour l'enfant de savoir attribuer à soi ou autrui les stimuli qui parviennent à son système auditif.

J'ajoute que cette remise en question n'a pas réellement de fondement si nous considérons notre capacité à évoquer mentalement la parole d'autrui. Comme mentionné ci-dessus (cf. I.2.2), nous entendons notre voix intérieure, son timbre, sa hauteur, nous pouvons même y détecter des erreurs. Nous pouvons aussi imaginer quelqu'un en train de parler et nous entendons alors sa voix. Nous pouvons avoir des dialogues imaginaires dans notre tête avec des personnes de notre entourage et nous percevons alors diverses voix. C'est grâce à l'attribution d'agentivité que nous ne confondons pas ces différentes voix intérieures avec des voix externes, car nous savons que nous les avons créées, imaginées nous-mêmes. L'attribution d'agentivité, qui semble bien être en œuvre lorsque nous parlons dans notre tête, contribue à construire ce sentiment de l'unité du soi, de contrôle de nos actions et pensées, notre sens de la subjectivité (cf. Frith, 2010). Outre son rôle dans la conscience de soi, le système d'attribution d'agentivité pourrait aussi jouer un rôle dans la théorie de l'esprit, notre

capacité d'attribuer à autrui des intentions et des pensées différentes des nôtres (Gallagher & Frith, 2003). Il joue ainsi un rôle crucial dans nos interactions sociales.

Hallucinations auditives verbales : défaillance de l'attribution de l'agentivité en parole intérieure

Feinberg (1978) a proposé que les hallucinations auditives verbales (HAV) dans la schizophrénie résultent d'un dysfonctionnement du mécanisme d'attribution d'agentivité lors de la production de parole intérieure. Les propres actions du patient, y compris ses pensées verbales, ne seraient pas perçues comme émanant de sa volonté, ce qui donnerait lieu à des troubles de l'attribution des actions. En particulier, les propres pensées verbales du patient seraient perçues comme non volontaires et attribuées à des voix venant de l'extérieur. Cette hypothèse d'un défaut d'attribution de l'agentivité a été développée en détail par Chris Frith pour expliquer le défaut d'étiquetage interne des actions en général chez les patients schizophrènes (Frith, 1987, 1992 ; Frith & Done, 1989). Selon Frith, une défaillance dans le mécanisme d'attribution de l'agentivité fait que les informations concernant une action intentionnelle ne sont pas transmises au système sensoriel et entraîne donc l'occurrence d'une action sans conscience ni de l'intention volontaire relative à celle-ci, ni de ses conséquences. Cette action serait alors attribuée à un agent externe et donnerait les symptômes de délire d'influence observés dans la schizophrénie. Comme le proposait déjà Feinberg, Frith a appliqué cette hypothèse au cas particulier de la pensée verbale : lors d'un dysfonctionnement de l'attribution de l'agentivité appliqué à la parole, il existerait un défaut d'étiquetage de la pensée verbale. Les hallucinations auditives verbales seraient donc des voix intérieures auto-générées par le patient mais interprétées comme venant de l'extérieur.

Hallucinations auditives verbales non psychiatriques : les entendeurs de voix

Les HAV sont des perceptions langagières en l'absence de stimuli externes appropriés (Franck & Thibaut, 2003). L'étymologie du mot hallucination vient du latin *hallucinatio* (méprise), de *hallucinatus*, *hallucinari* (errer, divaguer), qui représente bien le non contrôle du phénomène.

Les HAV sont souvent associées à la schizophrénie, car elles en sont le symptôme le plus fréquent. Comme nous l'avons rappelé dans Rapin *et al.* (2016), elles peuvent cependant être vécues par plusieurs types de populations. Elles peuvent par exemple survenir chez tout un chacun à l'endormissement (hallucinations hypnagogiques), au réveil (hallucinations hypnopompiques) ou à l'éveil en pleine journée (Jones *et al.*, 2010). Il semble que les HAV touchent entre 4 et 15% de la population saine, sans trouble psychiatrique (Linden *et al.*, 2010 ; Beavan *et al.*, 2011). Charles Fernyhough et ses collègues de l'université Durham au Royaume Uni ont créé un projet et un site internet dédié à cette population (<http://hearingthevoice.org>, cf. aussi Fernyhough, 2016). En France, comme dans de nombreux pays, il existe de nombreux regroupements d'entendeurs de voix et un réseau, le REV (Réseau français sur l'Entente de Voix : <http://revfrance.org>) dont l'objectif est de « promouvoir une approche des voix et des autres perceptions, expériences ou vécus inhabituels, respectueuse des personnes et de leur expertise ». Le REV considère, à juste titre, « que le fait d'entendre des voix n'est pas, en soi, un symptôme de maladie mentale mais qu'il s'agit d'un phénomène porteur de sens pour les personnes concernées et que, pour ces raisons, il convient de prendre les voix en considération ».

Toutefois ce phénomène est parfois plus préoccupant, notamment quand il est accompagné d'autres symptômes tels que la dépersonnalisation ou l'impression d'étrangeté face à soi-même. L'hallucination peut alors devenir un symptôme de maladie mentale, notamment dans la schizophrénie, un trouble psychiatrique qui touche environ 1 % de la population. Les HAV sont très fréquentes chez les patients atteints de schizophrénie et affectent de 50 à 80% d'entre eux (Nayani & David, 1996). Elles sont la plupart du temps très dérangeantes et invalidantes pour les patients (Franck, 2006). Il est donc crucial que le phénomène soit mieux compris, et que les mécanismes qui le sous-tendent soient mieux décrits, pour que les personnes qui en souffrent soient mieux prises en charge.

Hallucinations auditives verbales : explication dans le cadre du *monitoring* par contrôle prédictif

L'hypothèse d'une défaillance de l'attribution de l'agentivité pour expliquer les HAV a été explicitée dans le cadre du MCP, présenté au I.2.1.1.2 (cf. Figure I-9). Dans ce modèle, les signaux ascendants de *feedback* sensoriel effectif (ainsi que les signaux descendants d'état désiré) nous fournissent le sens de la propriété corporelle. Ils permettent de savoir que c'est bien notre corps qui est en mouvement (Gallagher, 2000 ; Franck & Thibaut, 2003 ; Tsakiris *et al.*, 2007). Le sens de l'agentivité, lui, est un sous-produit du contrôle prédictif. Le mécanisme de copie d'efférence permet en effet d'avertir nos systèmes sensoriels que l'action en cours est auto-produite (Wolpert *et al.*, 1995 ; Wolpert, 1997). En effet, si les deux *feedbacks* sensoriels prédits et vécus sont similaires (comparaison C3 sur la Figure I-9), alors les aires sensorielles sont averties que les stimuli perçus sont auto-générés, ce qui fournit le sens de l'agentivité.

Selon certains auteurs, la simple présence d'un état prédit (avant même la comparaison avec le *feedback* effectif) pourrait fournir le sens de l'agentivité (Blakemore *et al.*, 2002 ; Frith, 2002). En effet, les mesures temporelles de Libet *et al.* (1983) ou Haggard *et al.* (1999) indiquent que nous sommes conscients d'avoir initié un mouvement bien avant de recevoir le *feedback* sensoriel lié à ce mouvement (mais voir ci-dessous des arguments contre cette hypothèse).

Le mécanisme d'atténuation sensorielle lors d'actions auto-produites, dont la parole

Un autre aspect du MCP est qu'il est censé atténuer les expériences sensorielles lorsqu'elles ont été auto-produites. Comme expliqué ci-dessus (cf. I.2.1.2.2), d'après Frith, Blakemore et collègues (cf. Blakemore *et al.*, 2002, par exemple), lorsque la comparaison C3 est nulle, une atténuation perceptive a lieu. En effet, il a été montré que les aires sensorielles sont moins activées pour les actions auto-produites que pour une même stimulation sensorielle produite par un agent externe (e.g. Blakemore *et al.*, 1998 ; Farrer & Frith, 2002). Dans d'autres domaines que la parole, ce mécanisme d'atténuation perceptive par la comparaison avec la copie d'efférence permet d'expliquer l'impossibilité de se chatouiller soi-même ou bien la stabilité des images visuelles perçues en dépit de nos mouvements oculaires (cf. Holst & Mittelstaedt, 1950 ; Sperry, 1950 ; Mack & Bachant, 1969). Si une stimulation est auto-produite, il n'est pas nécessaire de recruter notre système sensoriel pour l'analyser. De même, si la production d'un son a été auto-initiée, il est inutile de recruter notre système perceptif de façon aussi intense que lorsque le son entendu provient de l'extérieur. Et surtout, il est important de réduire les conséquences des signaux auto-produits pour renforcer la détection des signaux de

parole produits par autrui (Agnew *et al.*, 2013). Enfin, cette atténuation sensorielle, combinée à une gestion de la temporalité des signaux perçus, contribue probablement à réduire l'impression de double percept lors de la parole auto-produite. Ceci expliquerait pourquoi, lorsque nous parlons à voix haute, notre propre voix n'est pas perçue deux fois (la prédiction auditive et l'onde sonore effectivement émise).

Probable atténuation sensorielle lors de la parole intérieure

Cette atténuation sensorielle n'est pas observée que dans le cas de la parole à voix haute. Certaines équipes défendent l'hypothèse qu'elle a aussi lieu lors de la parole intérieure. L'étude EEG de Ford et Mathalon (2004) citée ci-dessus (I.2.1.1.2) comprenait une condition de parole intérieure. La diminution de la composante N1 (réponse du cortex auditif) observée pour l'écoute passive de syllabes lors de la production de parole a lieu même en parole intérieure. Notons cependant que les phrases prononcées dans cette condition étaient présentées auditivement avant d'être produites mentalement. Comme le notent Ford & Mathalon, l'atténuation de la composante N1 pourrait ainsi être simplement liée à l'interférence acoustique causée par ces stimuli langagiers. Dans une étude MEG, Numminen & Curio (1999) observent une atténuation de la composante M100 lors de la production de voyelles en parole intérieure accompagnée d'une écoute passive de voyelle, par rapport à une condition d'écoute sans production. La diminution observée est toutefois bien moindre que celles qu'ils observent lors de la production de parole à voix haute. D'ailleurs, dans une autre étude, Numminen *et al.* (1999) rapportent que la réponse M100 à des tons purs de 1000Hz est significativement atténuée lors de la production de parole comparée à la parole intérieure. Ainsi les preuves de l'atténuation de la composante N1 ou M100 lors de la parole intérieure semblent encore minces et il semble que cette atténuation soit assez faible.

Quoi qu'il en soit, d'après Blakemore et collègues, l'atténuation sensorielle peut avoir lieu lorsque la comparaison C3 donne un signal nul (par similitude entre la prédiction et le retour sensoriel effectif). J'ai expliqué au I.2.1.2.2 que cette hypothèse n'est pertinente que si l'on accepte que le *monitoring* des actions d'autrui repose sur une temporisation entre signaux prédits et effectifs différenciée de celle qui a lieu dans les actions auto-produites (cf. aussi Encadré I.24).

L'atténuation sensorielle censée s'opérer au niveau de la comparaison C3, dans le cas des stimulations auto-produites, pourrait résulter d'un mécanisme *ad-hoc*, ou simplement d'une absence de stockage renforteur de l'expérience sensorielle auto-produite effective. En effet, dans le cas des stimulations externes (non auto-produites), la temporisation nécessaire pour que ces signaux puissent être comparés aux prédictions à venir induirait au contraire un renforcement sensoriel. Ainsi les signaux effectifs auto-produits (non-renforcés) paraissent atténués par rapport aux signaux effectifs externes (renforcés par temporisation).

Le point crucial pour l'explication classique des HAV est cette comparaison entre état prédit et *feedback* sensoriel vécu (C3), qui nous permettrait de distinguer les stimulations externes des actions auto-produites. La comparaison C3 donnerait une similitude entre ces états dans le cas des actions auto-produites, alors qu'elle conduirait à une différence dans le cas des actions produites par autrui (Frith *et al.*, 2000 ; Franck & Thibaut, 2003).

Nous serions ainsi équipés d'un mécanisme d'attribution de l'agentivité des actions à soi ou à autrui, via la comparaison C3 entre expériences sensorielles effectives et prédites.

En cas de défaillance du mécanisme de contrôle prédictif, soit par détérioration de la copie d'efférence soit par un dysfonctionnement du modèle direct ou du système de comparaison lui-même, la personne reste capable de formuler des intentions et de produire des actions, mais n'est plus apte à prédire correctement les conséquences de ses propres actions. Cette défaillance la conduit à une attribution d'agentivité erronée : la personne est privée de la sensation de subjectivité et attribue les stimulations perçues à un agent externe. De plus, dans l'explication classique (mais voir plus bas pour une hypothèse alternative ne nécessitant pas d'atténuation *ad-hoc*) la non-atténuation sensorielle, qui résulte d'une comparaison défectueuse, augmente la saillance des expériences sensorielles vécues et renforce encore leur qualité externe.

Ainsi, en élaborant l'hypothèse originale de Feinberg et en la replaçant dans le cadre du MCP, Frith et collègues (Frith, 1992 ; Frith *et al.*, 1995), puis Seal *et al.* (2004) et Jones & Fernyhough (2007b) ont proposé des explications du phénomène hallucinatoire qui suggèrent toutes que les HAV sont liées à une défaillance dans l'estimation de l'état prédit lors de la production de parole intérieure. Ces explications diffèrent sur la nature de la défaillance (absence d'état prédit ou mauvaise estimation de l'état prédit), mais toutes soutiennent que la comparaison C3 entre état prédit et feedback sensoriel vécu est alors non nulle et que l'atténuation sensorielle n'a pas lieu. Ainsi, soit à cause d'un biais d'attribution (Seal *et al.*, 2004), soit parce que le sens de l'agentivité n'est pas présent (Jones & Fernyhough, 2007b), la parole intérieure est alors vécue comme une voix externe, produite par autrui. Selon ces auteurs, un dysfonctionnement dans le mécanisme prédictif lors de la production de parole intérieure ne nous permettrait alors plus de reconnaître nos propres pensées verbales comme étant auto-produites et serait à l'origine des HAV.

Toutefois, nous avons montré que l'application du MCP à la parole intérieure est problématique (Rapin *et al.*, 2013). Comme le note Frith (2012) lui-même, quel peut-être le feedback sensoriel vécu dans le cas de la parole intérieure, par essence silencieuse et sans mouvement ? Dans le cas de la parole intérieure, quel sens donner à la comparaison C3 entre un *feedback* prédit et un *feedback* inhibé ? Nous avons proposé deux explications possibles.

La première explication s'appuie sur l'hypothèse selon laquelle, lors de la production de parole intérieure, des commandes motrices sont générées et envoyées aux muscles appropriés. Des signaux inhibiteurs seraient envoyés en parallèles pour limiter l'intensité des commandes motrices à un seuil ne déclenchant pas le mouvement des organes de la parole. Ainsi, ces signaux inhibiteurs empêcheraient le mouvement effectif des articulateurs, mais la simple présence de commandes motrices dans les muscles orofaciaux pourrait augmenter légèrement la tension musculaire et pourrait correspondre à un feedback proprioceptif détectable. Lors de la production de parole intérieure, le *feedback* sensoriel vécu pourrait ainsi exister et être de nature sensorielle proprioceptive (plutôt qu'auditive et articulatoire). Ce *feedback* proprioceptif résiduel pourrait donc être comparé à celui prédit par un mécanisme prédictif défaillant (comparaison C3) et pourrait être ré-interprété comme une voix externe.

La deuxième explication est celle sur laquelle repose le modèle que j'ai présenté dans la section 1.2.2.2. Elle suppose, comme la première, que lors de la production de la parole intérieure chez des personnes saines, des commandes motrices sont envoyées au système moteur et sont inhibées pour ne donner lieu à aucun mouvement articulatoire

ni à aucun son. Une copie d'efférence est toutefois envoyée au modèle direct ce qui fournit un état prédit sous forme d'une image auditive (et/ou visuelle, articulatoire), qui serait justement la petite voix que l'on entend dans sa tête (cf. I.2.2). Pour le sens de l'agentivité (subjectivité) en parole intérieure, la comparaison pertinente ne serait pas celle entre *feedbacks* prédit et vécu (C3), puisque le *feedback* vécu ne correspond à aucun son ni mouvement articulatoire. Le sens de l'agentivité ne peut pas non plus provenir de la simple présence d'un état prédit, car si l'HAV est une production de parole intérieure, alors l'état prédit existe bien lors d'une HAV, c'est même lui qui donne la sensation d'entendre une voix. Cet état prédit ne peut donc lui-même être à l'origine du sens de l'agentivité, qui est justement supposé absent lors d'une HAV. Le sens de l'agentivité proviendrait donc de la comparaison entre états désiré et prédit (C2). Ainsi lorsque l'état prédit ne correspond pas à l'objectif initial (à cause par exemple d'une défaillance du modèle prédicteur), la comparaison donne lieu à un défaut d'agentivité.

Ces deux explications sont en fait compatibles. La première, qui suppose qu'il pourrait subsister une proprioception dans les muscles orofaciaux lors de la parole intérieure est étayée par l'expérience introspective de Stricker (1885, voir Løevenbruck *et al.*, 2017 et 2018). Elle est également soutenue par des travaux récents montrant que les représentations de parole incluent non seulement des représentations articulatoires et auditives, mais également des informations proprioceptives ainsi que des informations tactiles de pression (Gick, 2015). La deuxième explication a été formulée en parallèle par Tian & Poeppel (2012) ainsi que Swiney & Sousa (2014), qui rejoignent nos questionnements sur l'applicabilité du modèle prédictif à la parole intérieure et ont fait la même remarque que la comparaison C3 n'est pas pertinente dans le cas de la parole intérieure. Ces auteurs, ainsi que Scott (2013) et Scott *et al.* (2013), proposent également que la voix et les sensations perçues lors de la production de parole intérieure consistent précisément en l'état prédit par la copie d'efférence.

Selon moi, cette petite voix est d'ailleurs présente aussi lorsque nous parlons à voix haute, même si elle est alors moins audible. Mais il existe des situations où nous pouvons l'entendre. L'expression française « il faut tourner sept fois sa langue dans sa bouche avant de parler » est remarquablement juste. Si on la prend au sens littéral, elle signifie que lorsque l'on prépare un acte de parole, il est prudent de simuler cet acte, de le répéter, et donc d'engager notre système moteur, notamment notre langue. La simulation mentale engage le modèle prédicteur qui fournit une représentation auditive de cette réponse, une petite voix intérieure, non encore énoncée, simplement préparée. Ces moments où la réponse simulée est répétée avant la réponse orale effective sont simplement une exagération du processus plus courant, où nous parlons à voix haute, après avoir prédit et simulé notre réponse, en voix intérieure. L'introspection suggère que lorsque nous cherchons nos mots, lorsque nous hésitons, nous pouvons mieux entendre notre voix intérieure, car elle n'est pas encore recouverte par l'énoncé définitif à voix haute (et elle n'est pas atténuée par la comparaison C3, cf. p. 119).

De plus, la suggestion que la comparaison pertinente pour l'agentivité soit C2 plutôt que C3 est reprise par Swiney & Sousa (2014) et avait même été proposée par Frith (2005) lui-même. On la retrouve aussi, de façon incidente dans Gallagher (2000). La proposition que l'agentivité résulte de la comparaison C2 (et non de C3) reste de plus valable dans le cas de la parole à voix haute et des actions en général, des mesures temporelles ayant révélé que le sens de l'agentivité précède la réalisation de l'action (cf. ci-dessus). Dans le

cas de la parole à voix haute, C3 pourrait cependant bien jouer un rôle dans l'attribution de l'agentivité, celle-ci pouvant être renforcée par le résultat de plusieurs comparaisons.

Notons qu'une alternative au MCP propose la notion de « filtre » plutôt que de comparaison (Langland-Hassan, 2008). Langland-Hassan, ainsi que d'autres chercheurs, notamment Gallagher (2004) (cf. I.2.1.1.2), argumentent qu'il n'est pas nécessaire de générer une prédiction en parole intérieure, puisque l'état prédit suffirait à la parole intérieure. La simple existence de la copie d'efférence, sans application du modèle direct et sans calcul d'un état prédit, agirait comme un filtre lors de la production de parole intérieure. Ce filtre doterait nos actions du sens de l'agentivité. Toutefois, ce modèle alternatif admet que la comparaison a bien lieu pour les autres modalités somatosensorielles que celles impliquées dans la parole intérieure et requiert donc des traitements différents pour la modalité auditive, ce qui semble peu parcimonieux. De plus, comme je l'ai expliqué plus haut, l'état prédit est dans un format supramodal qui ne suffit pas à générer une petite voix. C'est le contrôle prédictif qui permet, en utilisant un simulateur de l'appareil vocal, de générer une petite voix dans la tête.

Encadré I.33. Agentivité : comparaison entre signaux désiré et prédit

Je m'en tiens donc à cette hypothèse, désormais partagée par d'autres groupes de recherche, que le sens de l'agentivité proviendrait non pas de la comparaison entre signaux effectif et prédit, mais de celle entre signaux désiré et prédit (C2).

Ainsi lorsque l'état prédit ne correspond pas à l'objectif initial (à cause par exemple d'une défaillance du modèle prédicteur), la comparaison donnerait lieu à un défaut d'agentivité et pourrait conduire à percevoir la petite voix auto-générée (la prédiction) comme externe. Cependant, cette révision du MCP, considérant l'agentivité comme mettant en œuvre signaux désiré et prédit, reste, elle aussi, problématique. Car, comme nous l'avons noté dans Rapin *et al.* (2013, 2016), les personnes qui souffrent d'HAV sont capables de produire volontairement de la parole intérieure sans entendre systématiquement des hallucinations. Et surtout les voix entendues sont décrites comme possédant des timbres particuliers et elles sont parfois identifiées comme celles de personnes connues.

Il faut donc accepter que des mécanismes supplémentaires entrent en jeu pour que la parole intérieure ne soit défaillante que dans certains cas et que la voix entendue ne soit pas la propre voix de la personne qui hallucine.

Rôle de la dialogalité dans les hallucinations auditives verbales

C'est probablement la dimension dialogale (décrite au I.2.2.1) de la parole intérieure qui sous-tend le phénomène hallucinatoire. C'est aussi ce que notent Fernyhough (2016) et Alderson-Day *et al.* (2016) : « *Recognizing this complexity of inner speech, particularly its conversational and social features, is important both for ecological validity (Fernyhough, 2013) and for understanding atypical cognition (Fernyhough, 2004). Auditory verbal hallucinations (AVH) have been proposed to reflect misattributed instances of inner speech (Bentall, 1990; Frith, 1992), but studies inspired by this view have arguably relied on a relatively impoverished, 'monologic' view of inner speech.* »²⁴

²⁴ « Reconnaître la complexité de la parole intérieure, et particulièrement ses caractéristiques conversationnelles et sociales, est fondamental, à la fois pour la validité écologique et pour la compréhension de la cognition atypique. Il a été proposé que les hallucinations auditives verbales reflètent

L'étude IRMf de Shergill *et al.* (2000) sur la comparaison entre parole intérieure et imagerie auditive verbale (imagination d'une personne parlant avec une voix préalablement entendue) chez huit patients schizophrènes en rémission, mais ayant souffert d'HAV, suggère une troisième explication alternative, toujours dans le cadre du MCP. Les données de ces patients ont été comparées à celles de six participants sains. Dans la tâche de parole intérieure de Shergill et collègues, les participants devaient générer une phrase mentalement (sur le modèle « j'aime... »). Plusieurs tâches d'imagerie auditive verbale ont été examinées. La tâche d'imagerie à la première personne était la même que celle de parole intérieure, avec l'instruction supplémentaire d'utiliser explicitement sa propre voix. Dans la tâche d'imagerie à la deuxième personne, les participants devaient imaginer une personne (entendue au préalable) s'adressant à eux et leur disant « tu aimes... ». Enfin dans la tâche d'imagerie à la troisième personne, les participants devaient imaginer une personne (entendue au préalable) parlant d'eux-mêmes et disant « il-elle aime... ». Bien que les données ne concernent qu'un nombre peu élevé de participants, elles me semblent éclairantes. Shergill et collègues ont en effet montré que lors de la tâche de génération de parole intérieure les activations cérébrales des patients n'étaient pas différentes de celles des participants sains, autrement dit que leur parole intérieure était générée correctement, sans dysfonctionnement. Toutefois, dans les trois conditions d'imagerie auditive verbales combinées, les activations des patients différaient, avec notamment, par rapport aux participants sains, une atténuation des activations bilatérales du cortex cérébelleux postérieur et des hippocampes, ainsi que, dans l'hémisphère droit, des cortex temporaux moyen et supérieur et du thalamus.

Il apparaît donc que lorsque les patients (qui ont souffert d'HAV dans le passé) doivent générer mentalement la voix d'autrui, ce qui nécessite de paramétrer les modèles internes pour qu'ils produisent les signaux d'autrui (cf. I.2.2.2), alors le schéma des activations cérébrale et cérébelleuse est différent de celui des participants sains. La diminution de l'activation dans le cervelet, le thalamus et le cortex temporal droit est compatible avec une mauvaise paramétrisation des modèles internes (cf. I.2.1.1.2) de la dernière étape de la production de parole intérieure, i.e. l'étape de planification articulatoire. Ces résultats sont cependant à prendre avec précaution, car il ne s'agit que d'un petit nombre de patients, d'une part, et parce que les différentes conditions d'imagerie ont été combinées, et intègrent donc la condition d'imagerie auditive à la première personne (avec sa propre voix).

L'étude IRMf de Sommer *et al.* (2008) apporte de nouveaux éléments intéressants. Ces auteurs ont comparé les activations cérébrales de 24 patients diagnostiqués avec une schizophrénie, dans deux conditions : pendant qu'ils faisaient l'expérience d'une hallucination auditive verbale et pendant qu'ils produisaient de la parole intérieure (génération mentale de mots). Ils ont observé que la différence principale entre les deux conditions était la latéralisation des activations, avec un engagement prédominant de la région frontale inférieure droite pendant l'HAV.

Je propose donc l'hypothèse suivante. Les HAV résultent bien d'une défaillance dans le *monitoring* de l'étape de planification articulatoire lors de la production parole intérieure, mais uniquement lorsque celle-ci vise à générer la voix d'autrui, i.e. utilise les modèles internes paramétrés pour autrui. Dans ce cas, comme illustré sur la Figure I-16, la

des cas de parole intérieure attribuées par erreur à autrui, mais les études menées dans ce cadre théorique ont une vision de la parole intérieure relativement simpliste et monologique. » (ma traduction).

défaillance pourrait par exemple provenir de modèles internes mal construits : les contrôleur et prédicteur du dernier niveau ne se correspondent plus : le contrôleur, de couleur rouge (pour continuer à utiliser les conventions de couleur employées ci-dessus, puisqu'il s'agit d'un modèle adapté à autrui), n'est plus la réciproque du prédicteur, de couleur rose saumon. Le signal prédit par le prédicteur sera donc différent du signal désiré. La comparaison donne alors lieu à la sensation d'une stimulation auditive non auto-générée. Ceci fait que cette stimulation va être traitée comme provenant d'une source externe, donc remontée aux niveaux supérieurs, comme elle l'est dans le cas de la perception de parole d'autrui (cf. I.2.1.2.2). On a donc une flèche rose saumon qui remonte vers les niveaux supérieurs pour être interprétées (comme toute parole extérieure), voire fournir une nouvelle prédiction. Ceci fait alors réverbérer le signal initial, qui renforce la stimulation auditive et donne à la voix perçue le caractère d'une voix externe nettement audible, telle que les personnes qui hallucinent la rapportent (cf. Rapin *et al.*, 2016).

Une hypothèse explicative complémentaire tient non pas dans une défaillance dans la nature des signaux générés par les contrôleurs/prédicteurs (égaux/différents), mais dans la gestion de leur alignement temporel. J'ai montré plus haut que cette gestion est capitale. Dans le cas du *monitoring* verbal intérieur égocentré (cf. I.2.2.2), une temporisation est à faire sur le signal désiré pour attendre le signal prédit (même si elle est minime car le signal prédit est généré sans délai). Le signal désiré est donc généré avant le signal prédit, ce qui est représenté par le symbole « < » sur la Figure I-15. Dans le cas du *monitoring* verbal allocentré (cf. I.2.1.2.2), la temporisation est faite sur le signal effectif, qui est mis en attente le temps que la prédiction soit effectuée, ce qui renforce son côté saillant. Le signal désiré (interprété) arrive dans ce cas après le signal effectif. Une mauvaise gestion de ces temporisations peut faire échouer l'attribution de l'agentivité, car les signaux apparaissent comme survenant dans un ordre inversé.

Cette deuxième hypothèse est compatible avec l'étude de Linden *et al.* (2011) qui montre que la séquence des activations cérébrales est différente entre les HAV et l'imagerie auditive verbale (imagination d'autrui parlant). Ils ont observé chez des personnes saines présentant des HAV que, lors des hallucinations, l'activité des aires temporales (notamment la TVA, l'aire temporelle de la voix) est concomitante à celle des régions prémotrices (notamment AMS), alors que chez les personnes non hallucinées à qui l'on demande de délibérément imaginer la voix d'autrui, la séquence d'activation suit le déroulement attendu : AMS puis TVA. Il se pourrait donc qu'un mécanisme de temporisation inadéquat ait lieu, ce qui activerait les aires temporales plus tôt que prévu et créerait un dysfonctionnement de l'attribution de l'agentivité.

Rôle de la non-intentionnalité dans les hallucinations auditives verbales

S'il est nécessaire de prendre en compte la dimension de dialogalité pour expliquer le phénomène hallucinatoire, car c'est lors de la génération de la voix d'autrui que le phénomène se produit, cela ne suffit pas. Il faut aussi, selon moi, tenir compte de la dimension d'intentionnalité. Le caractère externe et intrusif des hallucinations provient en effet aussi du fait que celles-ci correspondent probablement à des variantes non délibérées de parole intérieure. On a vu ci-dessus (cf. I.2.2.2) que la parole intérieure vagabonde ne met probablement pas en jeu de mécanisme de *monitoring* de la verbalisation, mais qu'une petite voix (la voix propre ou celle d'autrui) pourrait cependant être générée dans certains cas. Ceci m'amène donc à dissocier le mécanisme

d'attribution de l'agentivité du mécanisme de *monitoring* verbal (ce qui n'est pas fait habituellement dans le MCP). Dans le cas du vagabondage verbal, par définition, le contrôle des verbalisations est limité : les paroles intérieures surviennent et l'on ne cherche pas à s'assurer qu'elles sont bien conformes aux objectifs initiaux (contrairement au cas de la parole intérieure délibérée). Toutefois, le sentiment d'intégrité du soi requiert que, même lorsque nos verbalisations semblent imprévisibles, on ne les attribue pas systématiquement à une source extérieure. Je nuance donc le propos de Gallagher (2004) cité ci-dessus (cf. I.2.2.2), qui considère que les pensées vagabondes ne sont pas accompagnées d'un sentiment d'agentivité.

Encadré I.34. Dissociation entre *monitoring* verbal et attribution de l'agentivité

Ainsi, il semble qu'il faille séparer le dispositif de *monitoring* du contenu de la verbalisation, qui ne serait à l'œuvre que dans la parole à voix haute et la parole intérieure délibérée, du dispositif d'attribution de l'agentivité, qui pourrait opérer dans tous les cas.

Je défends finalement l'hypothèse que, même dans le cas de la parole intérieure vagabonde, les verbalisations restent le plus souvent identifiées comme étant auto-générées (cf. aussi Encadré I.30). Dans le modèle explicatif que je propose (cf. Figure I-16), les hallucinations sont considérées comme résultant d'une défaillance spécifique de l'attribution de l'agentivité, lors de la survenue de parole intérieure vagabonde avec la voix d'autrui. Ce dysfonctionnement proviendrait d'une non-équivalence entre les sorties des modèles internes prédicteur et contrôleur paramétrés pour la voix d'autrui (essentiellement dans l'étape d'articulation, mais pas uniquement car l'hallucination est parfois condensée). Cette non-équivalence serait due elle-même soit à des modèles internes inadaptés, soit à un mauvais alignement temporel entre les signaux générés par ces paires de modèles, le signal désiré apparaissant comme survenant après le signal prédit, d'où le symbole « > » au niveau des comparaisons sur la Figure I-16.

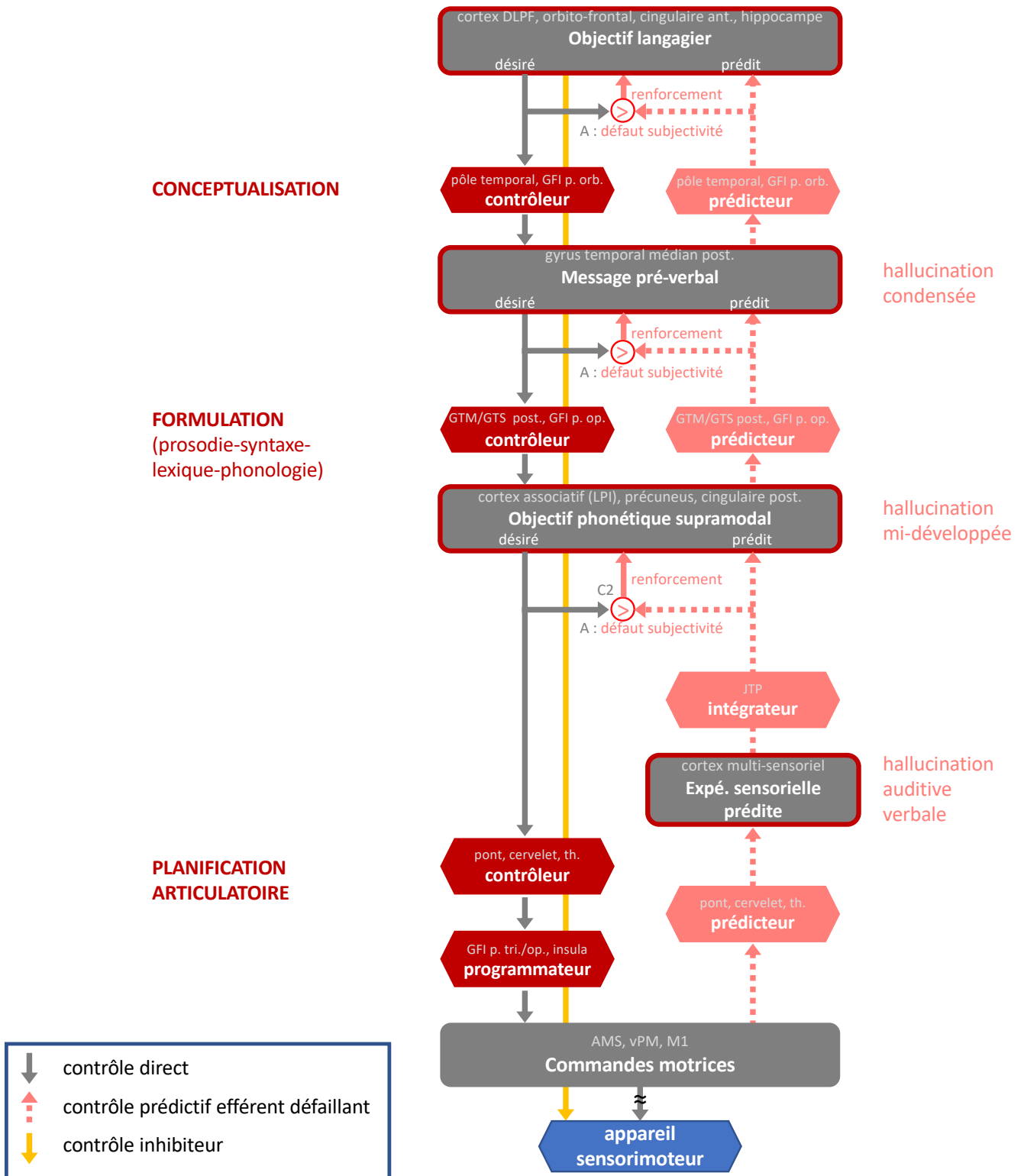


Figure I-16. Modèle explicatif des hallucinations auditives verbales fondé sur une défaillance de la production de parole intérieure vagabonde, avec la voix d'autrui. Des différences entre les sorties des modèles internes contrôleur et prédicteur d'autrui (ou une mauvaise gestion de leur alignement temporel) pourraient résulter en une comparaison non nulle entre signaux prédits et désirés, conduisant à une attribution à autrui de la petite voix ainsi générée et la rendant plus saillante auditivement.

Encadré I.35. Les hallucinations auditives verbales : défaillance spécifique de l'attribution de l'agentivité lors de l'imagination non délibérée d'autrui parlant

Ainsi, ma nouvelle proposition est que les HAV résultent d'un dysfonctionnement de l'imagerie auditive verbale (i.e. de l'imagination d'autrui parlant) non délibérée. Ce dysfonctionnement proviendrait d'une défaillance de l'équivalence entre les sorties des modèles internes prédicteur et contrôleur paramétrés pour la voix d'autrui. Au niveau de la comparaison C2, la petite voix prédite serait ainsi dotée d'une qualité d'altérité. Cette petite voix étant perçue comme externe, elle serait traitée comme telle et remontée aux niveaux linguistiques supérieurs pour être interprétée, entraînant une nouvelle prédiction et une réverbération, lui conférant une intensité sonore typique des voix externes. Le même phénomène pourrait se répliquer aux niveaux de condensation supérieurs : l'objectif phonétique supramodal ainsi que le message pré-verbal pourraient ainsi eux-mêmes subir un défaut de subjectivité et être attribués à autrui.

Avec ces nouveaux éléments, il est possible de proposer un modèle intégré de l'attribution de l'agentivité en général (pour la voix haute et la parole intérieure, pour la production et la perception) qui repose sur les différentes comparaisons entre signaux effectifs, produits et planifiés (désirés dans le cas de la production de parole et interprétés dans le cas de la perception) et qui utilise une même hiérarchie de modèles contrôleurs et prédicteurs/analyseurs pour la parole intérieure (avec voix propre ou voix d'autrui), la parole à voix haute et la perception de la parole. Ce modèle est illustré sur la Figure I-17.

On y voit que la gestion de la temporisation des signaux est cruciale pour l'attribution à soi-même ou à autrui des signaux auditifs perçus. On y voit aussi que la gestion de la correspondance entre modèles contrôleurs et prédicteurs de soi et d'autrui est essentielle. Une défaillance dans l'une ou l'autre de ces gestions pourrait conduire à l'hallucination.

Encadré I.36. Attribution de l'agentivité

En résumé, l'autooïse met en jeu un mécanisme d'attribution de l'agentivité des verbalisations intérieures ou extérieures, nous permettant de savoir qui est à l'origine du son perçu (soi ou autrui) et quelle voix est utilisée (voix propre ou voix d'autrui, entendue ou imaginée). Ce mécanisme se greffe sur les *monitorings* verbaux égocentré et allocentré déployés habituellement pour la production et la perception de la parole à voix haute, et nécessite une gestion fine des temporisations entre les différents signaux sensoriels. Il nous permet de distinguer les paroles qui viennent d'autrui de celles que nous avons nous-mêmes produites à voix haute ou silencieusement, nous rendant ainsi capables d'avoir des conversations fluides avec nos congénères et surtout de réguler le suivi de soi et d'autrui dans nos constructions narratives autobiographiques.

Il contribue ainsi à renforcer la distinction entre soi et autrui, qui contribue à l'ipsité. J'utilise le terme « d'attribution de l'agentivité » (noté A) pour désigner ce mécanisme (cf. aussi I.2.4).

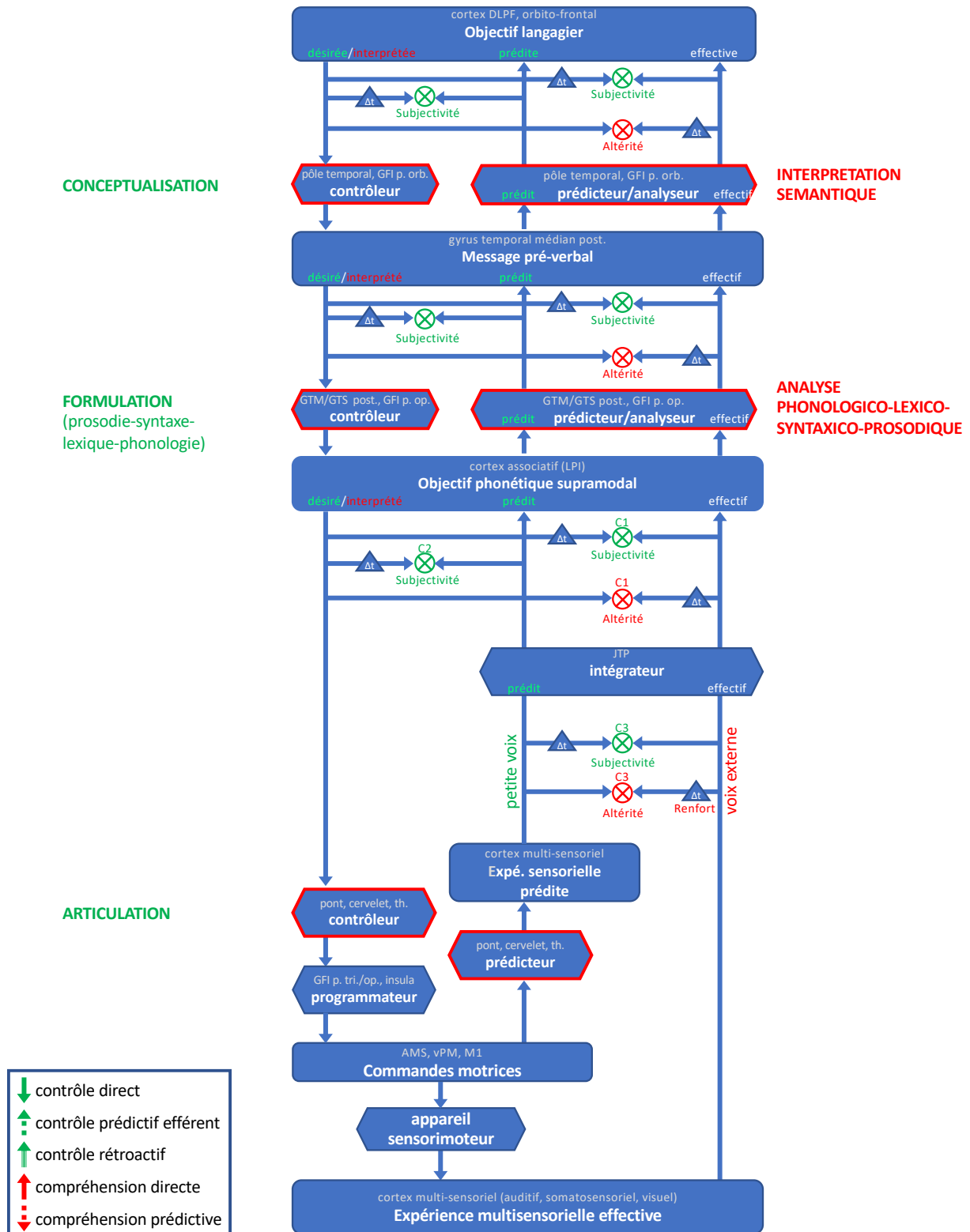


Figure I-17. Modèle intégré de l'attribution de l'agentivité lors de la production ou de la perception de parole, incluant les variantes de parole intérieure monologique et dialogale.

I.2.4. Premières briques d'un formalisme intégré de la cybernétique de la triade communication-cognition-autooèse

Dans ce qui précède, j'ai montré comment des mécanismes de contrôle et de suivi s'exercent dans chacune des trois fonctions majeures du langage, communication, pensée et autooèse, et quels peuvent être leurs corrélats neuroanatomiques. Or, ces trois fonctions du langage sont intriquées et le contrôle de l'une joue sur le contrôle de l'autre. Je cherche donc à construire une modélisation neurocomputationnelle intégrée qui rassemblera en un seul formalisme les modèles de *monitoring* verbal et d'attribution de l'agentivité que j'ai proposés pour chaque fonction. Je pose ici les premières briques de ce formalisme encore en construction.

La triade communication-cognition-autooèse, repose tout d'abord sur une capacité fondamentale, celle de l'attribution de l'agentivité, c'est-à-dire la capacité d'attribuer des actions à soi-même ou à autrui. Dans la communication, il faut en effet savoir qui parle, dans la pensée dialogale, il faut garder le fil de qui de nos interlocuteurs imaginaires, et dans l'autooèse, il faut également pour se distinguer soi-même d'autrui, savoir qui parle, qui pense et qui est en train d'effectuer un *monitoring* de l'autre. Ensuite, une deuxième capacité fondamentale est celle du *monitoring* du contenu des verbalisations, qu'elles soient intérieures ou extériorisées, qu'elles soient les nôtres ou celles d'autrui.

De façon plus détaillée, comme on l'a vu dans la section I.2.1, la fonction communicative requiert un *monitoring* verbal égocentré ($M_S(v_S)$) et allocentré ($M_S(v_O)$). Dans son versant productif, elle nécessite en effet un mécanisme permettant de suivre et réguler nos propres productions langagières, de les vérifier, de détecter des erreurs par rapport à l'objectif communicatif initial et éventuellement les corriger immédiatement, en ligne, parfois avant même qu'elles soient complètement réalisées ou juste après (ce que j'ai nommé $M_S(v_S)$, cf. I.2.1.1.2). Elle nous permet également de tenir compte des signaux de l'environnement, du bruit ambiant par exemple, et d'adapter notre énonciation pour permettre à autrui de nous comprendre. Sur le versant perception, lorsqu'en tant qu'auditeurs, nous percevons le langage d'autrui, nous mettons en œuvre des mécanismes de *monitoring* des productions verbales d'autrui, à partir de nos connaissances sur le monde, sur nos interlocuteurs et sur le langage, afin de suivre et comprendre nos interlocuteurs (ce que j'ai désigné par $M_S(v_O)$, cf. I.2.1.2.2). Outre le *monitoring* du contenu des verbalisations, la fonction communicative requiert également l'attribution de l'agentivité de la verbalisation, pour surveiller qui parle (de soi ou d'autrui), et l'attribution de l'agentivité du *monitoring* verbal, pour identifier qui est en train de suivre l'autre (et vérifier qu'autrui nous suit).

Dans sa fonction cognitive de vecteur et architecte de la pensée, le langage requiert le suivi de soi verbalisant mentalement, i.e. le contrôle de ce que l'on verbalise intérieurement par rapport à ce que l'on avait l'intention de penser. Plus exactement, il faut contrôler le contenu de nos productions mentales avec notre point de vue et notre voix ($M_S(<v_S>)$) ou celle d'autrui ($M_S(<v_S>)$), i.e. le *monitoring* verbal intérieur égocentré. Il faut aussi suivre le contenu des productions imaginées avec le point de vue d'autrui ($M_S(<v_O>)$), le *monitoring* verbal intérieur allocentré. Il faut également savoir quel est le point de vue imaginé en cours, l'attribution de l'agentivité de la verbalisation intérieure.

Dans sa fonction d'autooèse, le langage, et en particulier le langage intérieur, requiert d'attribuer à soi ou à autrui les verbalisations extérieures et intérieures, ainsi que l'action

de *monitoring* en cours. Elle implique donc l'attribution de l'agentivité, d'une part, des verbalisations (intérieures et extérieures) et, d'autre part, du *monitoring* verbal.

Ainsi, chacune des trois fonctions du langage (communication, pensée, auto-évaluation) requiert des dispositifs spécifiques. Or nous avons vu que ces trois fonctions sont étroitement liées, intriquées. Il en découle que les dispositifs de contrôle se recoupent et se chevauchent dans une cybernétique commune, ce que je tente ici de formaliser.

Je propose de distinguer deux processus qui nécessitent un contrôle et peuvent être ajustés, corrigés :

- (1) le processus de production verbale ou verbalisation extériorisée, noté v ,
- (2) le processus de verbalisation imaginée (intérieure), que je note $\langle v \rangle$.

Je distingue ensuite deux dispositifs :

- (1) le dispositif d'attribution d'agentivité des actions que je note A .
- (2) le dispositif de *monitoring* du bon déroulement des actions, que je note M .

Ces deux dispositifs doivent être dissociés, car dans le cas du vagabondage, j'ai montré qu'il peut être pertinent d'attribuer l'agentivité de la production verbale sans avoir besoin de réguler son contenu (cf. Encadré I.34).

Dans cette cybernétique commune, le dispositif primordial est l'attribution de l'agentivité : il faut avant tout savoir qui parle, qui pense, voire qui régule. Le dispositif d'attribution de l'agentivité définit l'agent des actions en cours, qui peut être soi ou autrui (que je note S pour *SELF* et O pour *OTHER*).

Ainsi, pour une verbalisation en cours, le système d'attribution de l'agentivité étiquette la verbalisation comme étant produite par soi : $A(v) = v_S$, ou par autrui : $A(v) = v_O$.

De même, une verbalisation intérieure peut être l'imagination d'une auto-production verbale, $A(\langle v \rangle) = \langle v_S \rangle$, ou être l'imagination d'une production verbale d'autrui, $A(\langle v \rangle) = \langle v_O \rangle$, puisque dans une situation de dialogue imaginaire, nous pouvons imaginer autrui s'adressant à nous-mêmes.

L'attribution d'agentivité s'applique au dispositif de *monitoring* lui-même, qui peut être le sien propre, on a alors $A(M) = M_S$ (nous régulons, par exemple, les mots que nous prononçons nous-mêmes ou les mots qu'autrui prononce) ou être celui d'autrui, $A(M) = M_O$ (autrui régule, par exemple, les mots que nous prononçons ou qu'il-elle prononce lui-elle-même).

On a donc finalement deux dispositifs :

- L'agentivité A qui s'applique au dispositif de *monitoring* et aux deux processus de verbalisation
- Le *monitoring* qui se décline en M_S ou M_O et qui s'applique aux deux processus de verbalisation ainsi que, de façon récursive, au dispositif de *monitoring* lui-même (voir ci-après).

Et on a deux processus de verbalisation qui se déclinent en deux possibilités d'agentivité :

- Verbalisation extériorisée : v_S ou v_O
- Verbalisation imaginée : $\langle v_S \rangle$ ou $\langle v_O \rangle$

I.2.4.1. La cybernétique de la fonction de communication

La fonction de communication a deux versants : « soi parle » (v_s) et « autrui parle » (v_o). Il faut donc tout d'abord un mécanisme d'attribution de l'agentivité à la production langagière en cours de traitement :

$A(v)$ qui peut être soit v_s , soit v_o .

Il faut aussi un mécanisme d'attribution de l'agentivité au *monitoring* verbal en cours :

$A(M)$ qui peut être M_s ou M_o .

Ensuite, concernant le premier versant, lorsque, dans une situation de communication, nous parlons nous-mêmes, nous réglons les verbalisations dont nous sommes les agents :

(a) $M_s(v_s)$: nous contrôlons ce que nous disons nous-mêmes.

Et autrui nous suit :

(b) $M_o(v_s)$.

Pour que la communication soit réussie, nous contrôlons aussi la façon dont autrui reçoit ce que nous disons. Nous estimons donc la façon dont autrui nous suit. Je note \hat{M}_o l'estimation que nous faisons du *monitoring* verbal d'autrui. On a donc ainsi un *monitoring* de second degré :

(c) $M_s(\hat{M}_o(v_s))$: nous contrôlons le *monitoring* par autrui de nos propres verbalisations.

En plus de contrôler ce que nous disons, autrui peut contrôler que nous contrôlons ce que nous disons, donc autrui peut faire une estimation de notre *monitoring* :

(d) $M_o(\hat{M}_s(v_s))$.

Un *monitoring* de troisième degré peut probablement s'appliquer : nous pouvons contrôler qu'autrui contrôle bien que nous contrôlons ce que nous disons, en utilisant une estimation de ce *monitoring* de la part d'autrui :

(e) $M_s(\hat{M}_o(\hat{M}_s(v_s)))$.

Autrui peut réciproquement contrôler que nous contrôlons qu'autrui suit bien ce que nous disons :

(f) $M_o(\hat{M}_s(\hat{M}_o(v_s)))$.

Un *monitoring* de quatrième degré est peut-être envisageable : nous pouvons contrôler qu'autrui contrôle bien que nous contrôlons qu'autrui suit ce que nous disons :

(g) $M_s(\hat{M}_o(\hat{M}_s(\hat{M}_o(v_s))))$.

Autrui peut contrôler que nous contrôlons bien qu'autrui contrôle que nous contrôlons ce que nous disons :

(h) $M_o(\hat{M}_s(\hat{M}_o(\hat{M}_s(v_s))))$.

Un *monitoring* de cinquième degré est concevable mais me paraît peu probable. Il semble toutefois que l'on puisse envisager une récursivité des *monitorings*, au moins jusqu'au deuxième degré, qui assure le succès de la communication.

Concernant le deuxième versant, lorsque, dans une situation de communication, autrui parle, nous suivons ses verbalisations. Il existe donc un *monitoring* des verbalisations d'autrui, au premier degré :

(a') $M_S(v_O)$: nous suivons ce qu'autrui dit.

(b') Et autrui contrôle ses propres productions : $M_O(v_O)$.

Pour que la communication soit réussie, un *monitoring* de second degré s'opère : autrui contrôle aussi la façon dont nous recevons ses productions. Autrui estime donc la façon dont nous le·la suivons. On a donc ainsi :

(c') $M_O(\hat{M}_S(v_O))$: autrui contrôle notre *monitoring* (estimé) de ses verbalisations.

En plus de suivre ce qu'autrui dit, nous pouvons contrôler qu'autrui contrôle ce qu'il·elle dit :

(d') $M_S(\hat{M}_O(v_O))$.

Il existe probablement aussi un *monitoring* des verbalisations d'autrui au troisième degré. Autrui peut contrôler que nous contrôlons bien qu'autrui contrôle ce qu'il·elle dit :

(e') $M_O(\hat{M}_S(\hat{M}_O(v_O)))$.

Nous pouvons nous-mêmes contrôler qu'autrui contrôle que nous suivons bien ce qu'il·elle dit :

(f') $M_S(\hat{M}_O(\hat{M}_S(v_O)))$.

Enfin, un *monitoring* des verbalisations d'autrui au quatrième degré est envisageable : autrui peut contrôler que nous contrôlons bien qu'autrui contrôle que nous suivons bien ce qu'il·elle dit :

(g') $M_O(\hat{M}_S(\hat{M}_O(\hat{M}_S(v_O))))$.

Nous pouvons contrôler qu'autrui contrôle bien que nous contrôlons qu'autrui ce qu'il·elle dit :

(h') $M_S(\hat{M}_O(\hat{M}_S(\hat{M}_O(v_O))))$.

Et ainsi de suite.

Il existe donc, comme pour les verbalisations de soi, une récursivité des contrôles des verbalisations d'autrui, au moins jusqu'au deuxième degré.

Pour se dérouler de façon optimale, un dialogue nécessite donc une récursivité des *monitorings* par les deux agents de la communication, soi et autrui. Une conversation à plusieurs agents nécessite un enchevêtrement de *monitorings* encore plus complexe.

Les mécanismes computationnels et les réseaux neuraux impliqués dans les deux contrôles de premier degré, le *monitoring* verbal égocentré ($M_S(v_S)$, suivi des verbalisations propres) et le *monitoring* verbal allocentré ($M_S(v_O)$: suivi des verbalisations d'autrui) sont décrits dans les sections 1.2.1.1.2 et 1.2.1.2.2. Ces *monitorings* verbaux s'appuient sur des comparaisons entre signaux désirés, prédits et effectifs, dont les substrats neuraux ont été décrits plus hauts. Les *monitorings* verbaux (b) et (b') sont des symétries de (a) et (b), où l'agent du *monitoring* verbal n'est plus soi mais autrui. On peut donc envisager les mêmes modèles computationnels et substrats neuraux.

Les *monitorings* de second degré : (c) $M_S (\hat{M}_O (v_S))$ et (d') $M_S (\hat{M}_O (v_O))$ et leurs symétriques de la part d'autrui (c' et d) nécessitent des mécanismes permettant de détecter qu'autrui effectue un contrôle de ses propres verbalisations ou des nôtres. Ces détectations peuvent se faire sur la base d'indices pragmatiques, linguistiques et paralinguistiques.

Pour (c), nous utilisons, au cours de la conversation, des outils linguistiques d'ordre pragmatique, permettant de vérifier qu'autrui nous suit : « hein ? » « d'accord » « tu vois » « tu me suis », etc. Nous pouvons aussi interpréter des signaux paralinguistiques, gestuels, corporels : regards perplexes, froncement de sourcil, moues dubitatives, bâillements, etc.

Pour (d) Nous savons interrompre autrui et utiliser des constructions linguistiques ou paralinguistiques pour vérifier que son contrôle de ses propres verbalisations est adapté : « tu veux dire... », « tu es sûre ? », « quoi ? », etc., ainsi que signaux gestuels et corporels.

Les *monitorings* de degrés plus élevés reposent probablement sur des indices pragmatiques similaires.

1.2.4.2. La cybernétique de la fonction de pensée

La fonction dialogale de pensée a deux versants : « imaginer soi parlant » (avec sa propre voix $\langle v_S \rangle$ ou une autre voix $\langle v_O \rangle$) et « imaginer autrui parlant » (et s'adressant à nous ou quelqu'un d'autre). Elle implique donc un *monitoring* verbal intérieur égocentré : $M_S (\langle v_S \rangle)$ et un *monitoring* verbal intérieur allocentré $M_S (\langle v_O \rangle)$. Ces deux versants s'appuient sur un *monitoring* de la parole intérieure générique (*monitoring* verbal intérieur égocentré) dont les mécanismes computationnels et les substrats neuraux sont décrits dans la section 1.2.2.2. Comme je l'ai mentionné ci-dessus, à ce *monitoring* générique, il faut ajouter des dispositifs de contrôle de plus haut niveau permettant de passer de notre propre voix à la voix d'autrui lorsque nous imitons mentalement autrui (sans prendre son point de vue), car nous devons alors utiliser des modèles internes paramétrés pour autrui. Nos verbalisations sont donc alors produites par nous-mêmes, mais avec une voix différente, soit $\langle v_S \rangle$ (*self* modifié). Il faut aussi des dispositifs de contrôle permettant de passer d'une perspective à une autre, i.e. d'une agentivité imaginée à l'autre, de la première ($\langle v_S \rangle$) à la troisième personne ($\langle v_O \rangle$).

1.2.4.3. La cybernétique de la fonction d'autoïté

La fonction d'autoïté requiert l'attribution de l'agentivité, c'est-à-dire l'attribution à soi ou autrui des deux processus de verbalisation (extériorisée et intérieure) et des différents *monitorings*. C'est en sachant à tout instant si c'est nous-mêmes (ou autrui) qui verbalisons effectivement, si c'est nous-mêmes ou autrui que nous imaginons en train de verbaliser ou si c'est nous-mêmes (ou autrui) qui réglons l'action en cours que notre conscience auto-réflexive (conscience de soi) se nourrit, que nos différents récits autobiographiques s'élaborent, se complètent, se corrigent. Le soi s'enrichit ainsi des différentes actions de verbalisation effectives et imaginées, l'autoïté s'élabore, au fil du temps²⁵. Le dispositif d'attribution de l'agentivité, qui étiquette l'agent de l'action, est

²⁵ Un autre processus important est à l'œuvre aussi, celui de la conscience du temps, du passé, du présent, du futur. Il se construit lui aussi en lien avec le langage, mais est hors du champ de mes recherches actuelles.

lui-même en retour affecté par cette nouvelle définition du soi. Il y a donc une réciprocité de l'attribution de l'agentivité (étiquetage soi vs. autrui) et de l'autoïse (contenu du soi). J'ai proposé une modélisation de l'attribution de l'agentivité appliquée à nos verbalisations intérieures et aux verbalisations extérieures (propres et d'autrui) dans la section I.2.3 (cf. Figure I-17). Il reste à expliquer comment l'attribution de l'agentivité du *monitoring* lui-même s'effectue. Car c'est sur elle que repose notre capacité à vérifier qu'autrui nous suit. Elle se fonde probablement sur des indices pragmatiques, comme suggéré ci-dessus (I.2.4.1).

I.2.4.4. Vers une cybernétique intégrée des trois fonctions du langage

Compte tenu de ces éléments, la cybernétique engagée dans les deux versants de la communication (parler et écouter), dans la pensée et dans l'autoïse peut être résumée comme proposée dans la Figure I-18 avec un tableau à trois colonnes, dont la première ligne concerne le dispositif d'attribution de l'agentivité et la deuxième, le dispositif de *monitoring*. Elle est divisée elle-même en sous-lignes pour les différents degrés de *monitoring* décrits ci-dessus. Les sections dans lesquelles ces différents mécanismes et leur implémentation neurocognitive sont traités sont indiquées le cas échéant. Les lignes grisées correspondent à des *monitorings* de 3^{ème} et 4^{ème} degrés qui ne sont peut-être pas pertinents. Les lignes en rouge sont celles pour lesquels des travaux sont nécessaires et constituent donc des pistes de recherche que j'envisage d'explorer dans les années à venir. Un de mes objectifs théoriques futurs sera de rassembler les différents modèles neurocognitifs de *monitorings* verbaux et d'attribution de l'agentivité proposés indépendamment pour chacune des fonctions dans mes travaux antérieurs, en un modèle neurocomputationnel intégré de la cybernétique verbale qui explique comment l'attribution de l'agentivité et les *monitorings* verbaux sont imbriqués.

	Communiquer : Parler	Communiquer : Ecouter autrui	Pensée	Auto-noëse
A G E N T I V I T É	$A(v) : v_{s,0}$ $A(M) : M_{s,0}$		$A(<v>) : <v_{s,0}>$	$A(v) : v_{s,0}$ cf. I.2.3 $A(<v>) : <v_{s,0}>$ cf. I.2.3 $A(M) : M_{s,0}$ Indices pragmatiques
M O N I T O R I N G	(a) $M_s(v_s)$: égocentré cf. I.2.1.1.2	(a') $M_s(v_o)$: allocentré cf. I.2.1.2.2	$M_s(<v_{s,s'}>)$ cf. I.2.2.2.2	
	(b) $M_o(v_s)$ symétrie de (a')	(b') $M_o(v_o)$ symétrie de (a)	$M_s(<v_o>)$ cf. I.2.2.2.2	
	(c) $M_s(\hat{M}_o(v_s))$ Indices pragmatiques	(c') $M_o(\hat{M}_s(v_o))$ Indices pragmatiques	+ indices de haut niveau	
	(d) $M_o(\hat{M}_s(v_s))$ Indices pragmatiques	(d') $M_s(\hat{M}_o(v_o))$ Indices pragmatiques		
	(e) $M_s(\hat{M}_o(\hat{M}_s(v_s)))$ Indices pragmatiques	(e') $M_o(\hat{M}_s(\hat{M}_o(v_o)))$ Indices pragmatiques		
	(f) $M_o(\hat{M}_s(\hat{M}_o(v_s)))$ Indices pragmatiques	(f') $M_s(\hat{M}_o(\hat{M}_s(v_o)))$ Indices pragmatiques		
	(g) $M_s(\hat{M}_o(\hat{M}_s(\hat{M}_o(v_s))))$ Indices pragmatiques	(g') $M_o(\hat{M}_s(\hat{M}_o(\hat{M}_s(v_o))))$ Indices pragmatiques		
	(h) $M_o(\hat{M}_s(\hat{M}_o(\hat{M}_s(v_s))))$ Indices pragmatiques	(h') $M_s(\hat{M}_o(\hat{M}_s(\hat{M}_o(v_o))))$ Indices pragmatiques		

Figure I-18. Proposition provisoire de formalisme intégré de la cybernétique des trois fonctions du langage : communication-pensée-auto-noëse

Ainsi, en m'appuyant sur ces premières ébauches de modélisation, je vise à construire un formalisme neurocomputationnel visant à mieux comprendre les fondements neurophysiologiques de la cybernétique verbale, pour la triade communication-cognition-auto-noëse. En parallèle, je recueille des données expérimentales, me permettant d'affiner les modèles computationnels échafaudés petit à petit. Ces données sont présentées dans le chapitre suivant.

Publications associées (articles de revues ou chapitres d'ouvrage, sites web)

- Grandchamp R., Rapin L., Perrone-Bertolotti M., Pichat C., Haldin, C., Lachaux J.P., Dohen M., Perrier P., Garnier M., Baciú M., Løevenbruck H. (soumis). The Condialint model: a neurofunctional account of varieties of inner speech along condensation, dialogicality and intentionality dimensions. *Frontiers in Psychology – Cognitive Science*, collection on Exploring the Nature, Content, and Frequency of Intrapersonal Communication
- Løevenbruck, H. (1996). *Pistes pour le contrôle d'un robot parlant capable de réduction vocalique*. Thèse de doctorat de l'Institut National Polytechnique de Grenoble.
- Løevenbruck, H. (2018). What the neurocognitive study of inner language reveals about our inner space. *Épistémocritique*, n° 18 : *Langage intérieur - Espaces intérieurs / Inner Speech - Inner Space*, Stéphanie Smadja, Pierre-Louis Patoine (eds.) [<http://epistemocritique.org/what-the-neurocognitive-study-of-inner-language-reveals-about-our-inner-space/>] - hal-02039667
- Løevenbruck H., Baciú M., Segebarth C. & Abry C. (2005). The left inferior frontal gyrus under focus: an fMRI study of the production of deixis via syntactic extraction and prosodic focus. *Journal of Neurolinguistics*, 18, 237-258. doi:10.1016/j.jneuroling.2004.12.002 - hal-00371865
- Løevenbruck H., Dohen M. & Vilain C. (2009). Pointing is 'special'. In Fuchs S., Løevenbruck H. Pape D. & Perrier P. (eds.), *Some Aspects of Speech and the Brain*, Peter Lang, 211-258, ISBN 978-3-631-57630-4. - hal-00360758
- Løevenbruck, H., Nalborczyk, L., Perrone-Bertolotti, M., Baeyens, C., Grandchamp, R. (2017). Un cas d'incarnation linguistique : implication du système moteur orofacial dans les ruminations mentales. *La cognition incarnée*, J. Dokic et D. Perrin (dir.), Recherches sur la philosophie et le langage, N° 33, Vrin, Paris, 297-324, ISBN : 978-2-7116-8423-6. - hal-01887636
- Løevenbruck, H., Grandchamp, R., Rapin, L., Nalborczyk, L., Dohen, M., Perrier, P., Baciú M. & Perrone-Bertolotti, M. (2018). A cognitive neuroscience view of inner language: to predict, to hear, to see and to feel. In *Inner Speech: New Voices*. Peter Langland-Hassan & Agustín Vicente (eds.), Oxford University Press, 131-167. - hal-01898992
- Perrier, P., Løevenbruck, H., & Payan, Y. (1996). Control of tongue movements in speech: The equilibrium point hypothesis perspective. *Journal of Phonetics*, 24(1), 53–75.
- Perrone-Bertolotti, M., Rapin L., Lachaux J.-P., Baciú M. & Løevenbruck H. (2014). What is that little voice inside my head? Inner speech phenomenology, its role in cognitive performance and its relation to self-monitoring. *Behavioural Brain Research*, 261, 220-239. Epub: 10.1016/j.bbr.2013.12.034. - hal-00967730
- Perrone-Bertolotti M., Grandchamp R., Rapin L., Baciú M., Lachaux J.P. & Løevenbruck H. (2016). Langage intérieur. In Pinto, S. & Sato, M. (eds.). *Traité de Neurolinguistique : du cerveau au langage*. Collection Neuropsychologie, De Boeck Supérieur, 109-123. - hal-01467107
- Rapin L., Dohen M., Polosan M., Perrier P., Løevenbruck H. (2013). *Journal of Speech, Language, and Hearing Research*. PMID : 24687444 - hal-00926630
- Rapin L., Dohen M. & Løevenbruck H. (2016). Les hallucinations auditives verbales. In Pinto, S. & Sato, M. (eds.). *Traité de Neurolinguistique Neurolinguistique : du cerveau au langage*. Collection Neuropsychologie, De Boeck Supérieur, 347-370. - hal-01442463
- Sato M., Baciú M., Løevenbruck H., Schwartz J.-L., Segebarth C., Cathiard M.-A. & Abry C. (2004). Multistable representation of speech forms: a functional MRI study of verbal transformations. *NeuroImage*, 23, 1143-1151. PMID: 15528113
- Schwartz J.-L. & Løevenbruck H. (2012). La parole, entre objets vocaux et objets du monde. In *La psychologie cognitive*. Collection Cogniprisme. M. Denis (ed.). Editions de la Maison des Sciences de l'Homme, ISBN 978-2-7351-1509-9, 165-199. - hal-00782274

Thèses co-encadrées : Marc Sato, soutenue en 2004, Lucile Rapin, soutenue en 2011

Sites web :

- Projet ANR InnerSpeech : <https://lpnc.univ-grenoble-alpes.fr/InnerSpeech-2014-2018>
- Projet MONOLOGUER, porté par Stéphanie Smadja, Université Paris Diderot : <https://cerilac.univ-paris-diderot.fr/monologuer>
- Impromptu Scientifique « des voix dans la tête » : <http://www.ateliers-du-spectacle.org/spectacle/les-impromptus-scientifiques/#voix>

Collaborations :

Grenoble : GIPSA-lab, LPNC, IRMaGe, GIN

Lyon : CRNL

Paris : CERILAC

Canada : CNoS, University British Columbia

Chapitre II.

TRAVAUX EXPERIMENTAUX

En parallèle de ces travaux théoriques, je suis engagée dans une démarche empirique, qui m'a permis de tester les premières ébauches du formalisme présenté ci-dessus, et de leur apporter, au fil du temps, des corrections ou de formuler de nouvelles hypothèses. Mon objectif est de tester expérimentalement l'implémentation des mécanismes de la cybernétique verbale requise par la triade communication, pensée, autoconscience.

J'ai recueilli des données comportementales et neurophysiologiques sur la production ou la perception du langage, chez l'adulte, au cours du développement chez l'enfant et dans des populations souffrant de troubles du langage, de l'audition ou de pathologies neurologiques ou psychiatriques (aphasie, schizophrénie). Je présente un résumé succinct de ces travaux, publiés dans des revues ou congrès internationaux.

II.1. MONITORING VERBAL POUR LA FONCTION DE COMMUNICATION

La fonction de communication nécessite, on l'a vu ci-dessus, des outils de suivi verbal de soi et d'autrui. Le pointage, ou la deixis, est particulièrement adapté pour ce type de *monitoring*. Des manœuvres prosodiques diverses sont également employées. Je présente ci-dessous quelques études que j'ai menées ou auxquelles j'ai participé.

II.1.1. Les capacités de pointer et de comprendre le pointage

Le pointage – ou deixis – vise à attirer l'attention d'autrui sur une personne, un objet ou un événement, centre de notre propre attention. Il permet à autrui de suivre notre propre état mental. Il peut prendre des formes multiples, telles que le pointage manuel, naturellement, mais aussi, le pointage oculaire ou même verbal, *via* l'extraction syntaxique (avec l'utilisation d'un présentatif, par exemple : « C'est Chomsky qui soutient la thèse de l'innéité. ») ou la focalisation prosodique, avec un contour intonatif mettant

en relief l'élément désigné (comme dans « CHOMSKY est l'instigateur du nativisme linguistique », voir par exemple Touati, 1989 ; Morel & Danon-Boileau, 1998 ; Rossi, 1999 ; Di Cristo, 2000 ; Touratier, 2000).

Il a été montré que le pointage oculaire (ou regards déictiques, vers 6-9 mois) puis le pointage manuel (ou gestes communicatifs déictiques, avec l'index, vers 9-11 mois) sont deux étapes clés du développement cognitif des nourrissons dont l'apparition est corrélée avec des étapes du développement langagier oral (construction du lexique et émergence de la syntaxe, cf. Bates & Dick, 2002 ; Volterra *et al.* 2005). La capacité de pointage, d'abord avec l'œil, puis avec le doigt et enfin avec l'intonation et la syntaxe, est une composante majeure du mécanisme d'attention partagée dans l'interaction adulte-enfant (cf. Tomasello *et al.*, 2007). Il apparaît donc qu'au cours du développement, les pointages oculaire et manuel sont intrinsèquement liés à la fonction communicative du langage. Il est donc particulièrement intéressant de mieux comprendre comment le pointage se développe chez l'enfant, c'est l'objet des travaux décrits dans la section II.4.3.

II.1.1.1. Production du pointage prosodique chez l'adulte

Les travaux présentés ci-dessous ont été conduits avec Marion Dohen, d'abord au cours de sa thèse à l'ICP (soutenue en novembre 2005), sous la co-direction de Jean-Luc Schwartz, et depuis son entrée comme Maître de Conférence à l'ICP/GIPSA-lab-DPC.

Les corrélats traditionnellement explorés de la prosodie sont des paramètres acoustiques tels que la fréquence fondamentale, l'intensité et la durée. Cependant, quelques équipes de recherche ont suggéré que la prosodie possède également des corrélats articulatoires qui pourraient être visibles. L'objectif des travaux de thèse de Marion Dohen était précisément d'examiner la cinématique de différents articulateurs visibles, dans diverses conditions prosodiques et de vérifier que ces variations sont utilisées par les spectateurs. Nous nous sommes penchées en particulier sur la focalisation prosodique contrastive, celle-ci jouant un rôle essentiel dans la communication et le *monitoring* verbal d'autrui.

Sur le plan acoustique, nous avons montré que la focalisation prosodique implique une augmentation de la fréquence fondamentale (F_0) et de la durée du constituant focalisé. Elle est aussi associée à une diminution de la F_0 sur le constituant pré-focal et une désaccentuation des constituants post-focaux (Dohen & Lœvenbruck, 2004). Le pointage prosodique requiert donc un contrôle laryngé très fin de la part du locuteur-riche.

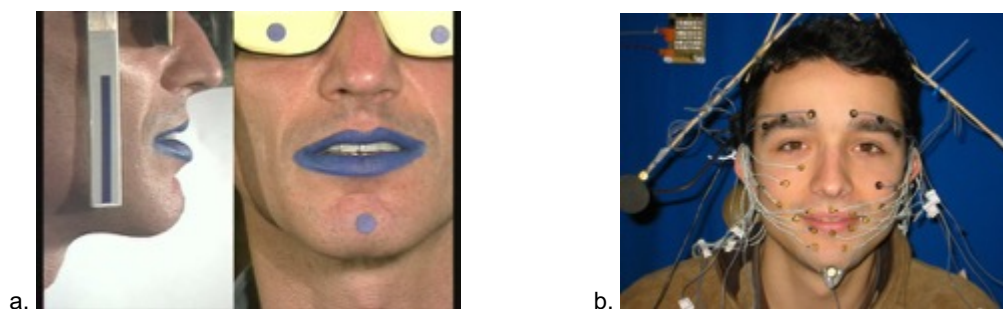


Figure II-1. Recueil de données articulatoires : a. suivi vidéo, b. suivi optoélectrique

Sur le plan articulatoire, nos mesures par suivis vidéo et optoélectrique (cf. Figure II-1) indiquent que la focalisation prosodique met en jeu un contraste articulatoire entre la partie focalisée de l'énoncé et le reste, les mouvements linguaux et labiaux étant plus

amples et de plus longues durées sur le constituant focalisé que sur les autres (Dohen *et al.*, 2011). Nos résultats indiquent de plus un lien entre la focalisation prosodique et les mouvements de la tête (hochements) et des sourcils. Ces résultats sont compatibles avec les résultats préliminaires que j'avais obtenus sur les mouvements de la langue par de l'électromagnétométrie articulatoire (Løevenbruck, 2000, cf. aussi II.1.2.1). Un contrôle articulatoire très précis est donc également requis pour le pointage prosodique.

II.1.1.2. Perception audiovisuelle du pointage prosodique chez l'adulte

Si nous commençons à mieux cerner les corrélats acoustiques et articulatoires de la production du pointage vocal, qu'en est-il de sa perception ? Nos tests perceptifs visuels indiquent qu'il est possible de détecter la focalisation prosodique uniquement à partir des données articulatoires visuelles (Dohen & Løevenbruck, 2005) et qu'en parole chuchotée, les scores de détection de la focalisation augmentent quand on passe de la modalité auditive seule à la modalité audiovisuelle (Dohen & Løevenbruck, 2009). Ainsi, les indices articulatoires et phonatoires identifiés semblent bien récupérés par les auditeurs/spectateurs et utilisés pour détecter la focalisation.

II.1.1.3. Corrélats cérébraux de la production de pointage multi-formes

Nos données chez l'adulte indiquant que les locuteurs organisent leur phonation et leur articulation de façon à transmettre la focalisation prosodique, j'ai émis l'hypothèse que des représentations sensorimotrices intégrées (auditives, visuelles et proprioceptives) sont nécessaires pour produire le pointage prosodique. Ces représentations seraient formées via l'activation d'aires cérébrales associatives, telles que les régions temporales et pariétales. Or, les pointages *manuel* et *oculaire* recrutent eux-mêmes le cortex pariétal postérieur associatif. Avec Francesca Carota, en post-doctorat, nous avons comparé les corrélats cérébraux du pointage dans quatre modalités : vocale prosodique, vocale syntaxique, manuelle et oculaire. Les données IRMf recueillies pour 15 sujets révèlent que le lobule pariétal supérieur gauche est activé pour trois modalités de pointage : manuel, oculaire et prosodique, mais pas pour le pointage syntaxique (cf. Figure II-2).

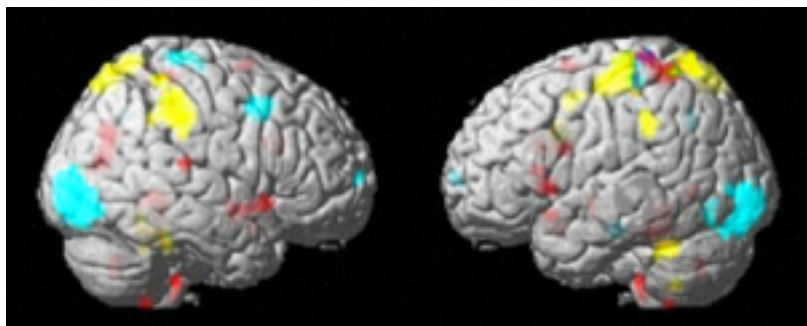


Figure II-2. Activations observées dans chacun des pointages chez les 15 participants. Analyse de groupe à effets aléatoires. En jaune : pointage digital, en bleu : pointage oculaire, en rouge : pointage prosodique. Dans le lobule pariétal supérieur, on observe une superposition des activations pour les pointages digital et oculaire (vert) et prosodique et oculaire (violet).

Comme nous l'avons expliqué plus haut, nos études comportementales 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 manuel ou oculaire. Ces représentations seraient formées via l'activation d'aires associatives 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, qui utilise une construction grammaticalisée, « prête-à-énoncer » (Løevenbruck *et al.*, 2008, 2009). L'ensemble de ces travaux confirme notre hypothèse que le pointage linguistique en ligne (la focalisation prosodique) est inscrit dans le même réseau cérébral que les pointages gestuels (manuel et oculaire). Pointages oculaire, manuel et vocal seraient ainsi bien liés, autant sur le plan développemental (cf. ci-dessus, II.1.1) que sur le plan cérébral.

II.1.1.4. Corrélats cérébraux de la perception du pointage prosodique

Examiner les corrélats cérébraux de la perception de la focalisation prosodique permet de mieux décrire les mécanismes cérébraux qui sous-tendent le suivi d'autrui pointant verbalement. J'ai ainsi participé, avec Marion Dohen et Marc Sato du GIPSA-lab, à un projet mené par Monica Baciú du LPNC, dans le cadre de la thèse de Marcela Perrone-Bertolotti. Nous avons recueilli des données IRMf lors d'une tâche de perception auditive de la focalisation prosodique contrastive. Les résultats révèlent que la détection de la focalisation chez autrui recrute un réseau cérébral bilatéral comportant des régions ventrales (incluant le gyrus temporal postéro-supérieur droit, les gyri temporaux moyens bilatéraux) ainsi que des régions dorsales (incluant les gyri frontal inférieur et pariétal inférieur bilatéraux, le gyrus pariétal supérieur gauche). Nos résultats semblent ainsi en faveur d'un modèle de la perception de la focalisation à deux voies (cf. Hickok & Poeppel, 2007) et sont compatibles avec l'hypothèse « Asymmetric Sampling in Time » (Poeppel, 2003). Ils suggèrent que la détection de la focalisation prosodique implique une interaction des deux hémisphères : les résultats du traitement des indices prosodiques « à variation lente » par l'hémisphère droit seraient envoyés au modèle interne utilisé de façon simultanée par l'hémisphère gauche, impliqué, lui, dans les traitements acoustiques sur des fenêtres temporelles plus étroites. L'observation d'activations dans les régions frontales pourrait refléter l'utilisation d'informations motrices *feed-forward* lors de la détection de la focalisation prosodique (Perrone-Bertolotti *et al.*, 2013).

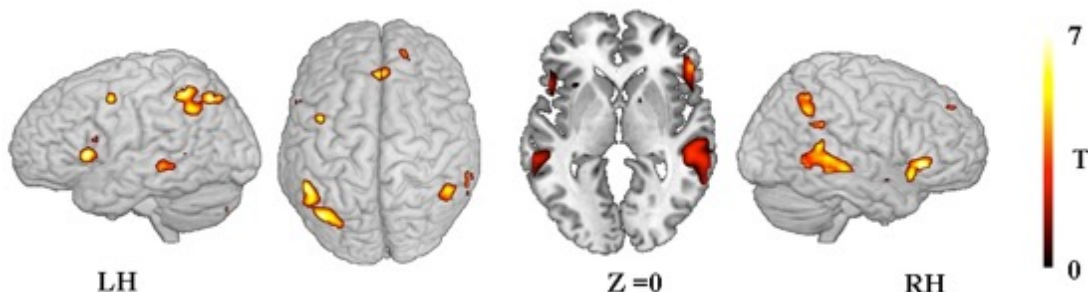


Figure II-3. *Contraste entre perception d'énoncés Focalisés vs. Neutres (analyse de groupe à effet aléatoire, LH : hémisphère gauche, RH : hémisphère droit).*

Publications associées (articles de revues ou chapitres d'ouvrage, thèse)

Carota F., Løevenbruck H., Vilain C., Baciú M., Abry C., Lamalle L., Pichat C. & Segebarth C. (2006). An fMRI study of multimodal deixis: preliminary results on prosodic, syntactic, manual and ocular pointing. *Actes de Speech Prosody 2006*, Dresden, Germany, May 2-5 2006, Vol. II, 915-918. - hal-00367514

- Carota F., Løevenbruck H., Vilain C., Baciú M., Abry C., Lamalle L., Pichat C. & Segebarth C. (2006). Left parietal activation during multimodal deixis: preliminary results on prosodic, syntactic, manual and ocular pointing. *Actes de 12th Annual Meeting of the Organization for Human Brain Mapping*, Florence, Italy, June 11-15, 2006, NeuroImage, Vol.31, Suppl. 1, S36. - hal-00371961
- Dohen, M., & Løevenbruck, H. (2004). Pre-focal rephrasing, focal enhancement and postfocal deaccentuation in French. In *Eighth International Conference on Spoken Language Processing, Jeju, Korea, 4-8 October 2004* (Vol. 1, pp. 785–788).
- Dohen, M., & Løevenbruck, H. (2005). Audiovisual Production and Perception of Contrastive Focus in French: a multispeaker study. In *Ninth International Conference on Speech Communication and Technology, September 4-8, Lisboa, Portugal* (pp. 2413–2416).
- Dohen M. & Løevenbruck H. (2009). Interaction of audition and vision for the perception of prosodic contrastive focus. *Language and Speech*, 52 (3), 177-206. - PMID: 19624029 - hal-00371180
- Dohen M., Løevenbruck H. & Hill H. (2009). Recognizing prosody from the lips: is it possible to extract prosodic focus from lip features? In Alan Wee-Chung Liew & Shilin Wang (Eds.), *Visual Speech Recognition: Lip Segmentation and Mapping*, Medical Information Science Reference, Hershey, New York, ISBN 978-1-60566-186-5, 416-438. - hal-00360765
- Dohen M., Løevenbruck H., Roustan B., Vilain C. (2011). Pointage manuel et vocal : corrélats sensorimoteurs et neurocognitifs, coordination geste manuel / parole. *Faits de Langue*, 37, 97-116. - hal-00784592
- Løevenbruck H., Vilain C., Carota F., Baciú M., Abry C., Lamalle L., Pichat C. & Segebarth C. (2007). Cerebral correlates of multimodal pointing: An fMRI study of prosodic focus, syntactic extraction, digital- and ocular- pointing. *Proceedings of the XVIth International Congress of Phonetic Sciences*, 6-10 August 2007, Saarbruecken, 1861-1864. - hal-00275718
- Løevenbruck H., Vilain C. & Dohen M. (2008). From gestural pointing to vocal pointing in the brain. *Revue Française de Linguistique Appliquée*, Vol. XIII, 2, 23-33. - hal-00360753
- Løevenbruck H., Dohen M. & Vilain C. (2009). Pointing is 'special'. In Fuchs S., Løevenbruck H. Pape D. & Perrier P. (eds.), *Some Aspects of Speech and the Brain*, Peter Lang, 211-258, ISBN 978-3-631-57630-4. - hal-00360758
- Perrone-Bertolotti, M. Dohen, M., Løevenbruck, H., Sato, M., Pichat, C., Baciú, M. (2013). Neural correlates of the perception of contrastive prosodic focus in French. An fMRI study. *Human Brain Mapping*, 34 (10), 2574-2591. - hal-00762363, PMID: 22488985

Thèse co-encadrée : Marion Dohen, soutenue en 2005

Post-doctorat : Francesca Carota, 2004-06

Site web : *Du Pointage au Langage* : une des *Histoires Courtes*, les romans-photos de la recherche, J.F. Dars & A. Papillault (<http://llx.fr/site/du-pointage-au-langage/> et <https://vimeo.com/328843449>), 2012

Financements :

Ce thème a bénéficié d'un financement par la Région Rhône-Alpes, Cluster 11 Handicap, Vieillesse, Neurosciences 2008-2010 (projet ARC 2 porté par Monica Baciú).

Collaborations :

Grenoble : GIPSA-lab, LPNC, IRMaGe

Japon : ATR

II.1.2. Manœuvres et variations prosodiques liées au suivi d'autrui

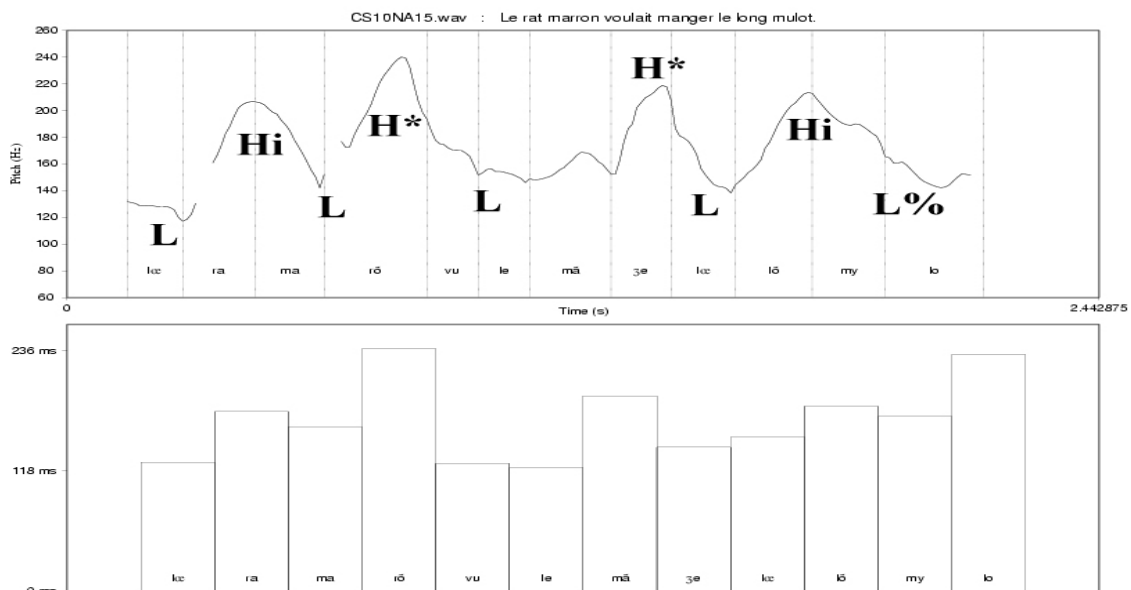
II.1.2.1. Structuration prosodique des énoncés en français : données acoustiques, articulatoires et perceptives

Si l'exploration de la structuration prosodique du message du point de vue des locuteurs (acoustique et articulatoire) est bien engagée, le point de vue des auditeurs est encore souvent ignoré (mais voir II.1.1.4). Or, comme les autres aspects de la phonologie, la structure prosodique est façonnée à la fois par des contraintes physiologiques et des exigences perceptives. L'objectif du travail présenté ci-dessous était de mettre en

relation production et perception d'une des entités de la hiérarchie prosodique du français, le Syntagme Accentuel, tel que décrit dans le modèle de l'intonation de Jun & Fougeron (2000, 2002). Il existe pour le français différents modèles hiérarchiques de la structure prosodique (par exemple, Verluyten, 1982 ; Dell, 1984 ; Rossi, 1985 ; Bailly, 1986 ; Martin, 1987 ; Vaissière, 1992 ; Monnin & Grosjean, 1993 ; Mertens, 1993 ; Post, 2000 ; Jun & Fougeron, 2000 ; Hirst & Di Cristo, 1998), mais ces modèles ne s'accordent pas sur le type de hiérarchie, le nombre de niveaux prosodiques en jeu et les représentations tonales ou intonatives sous-jacentes. En outre, en ce qui concerne les systèmes de transcription prosodique du français, la complexité des objets à représenter ainsi que la grande diversité des objectifs visés par les transcrip-teurs (théorie phonologique, évaluation phonétique, traitement automatique de la parole, apprentissage du français langue étrangère, etc.) font qu'il n'existe pas de consensus sur la notation ni la représentation (tonale, par contour, par niveaux, par traits, etc.). Le modèle de Jun & Fougeron (2000, 2002), qui s'inscrit dans le cadre de la phonologie prosodique développée notamment par Pierrehumbert (1980), Selkirk (1986), et Beckman (1986), présente l'avantage d'être en accord avec la plupart des descriptions de la prosodie du français et d'utiliser un système de transcription prosodique similaire au système ToBI qui, s'il est contesté, a montré son utilité dans la description de nombreuses langues. Ce modèle comporte trois unités prosodiques définies tonalement. L'unité la plus basse de la hiérarchie est le Syntagme Accentuel (SA) et la plus élevée est le Syntagme Intonatif (SI). Le SI, correspond à l'intonème majeur de Rossi, ou l'unité intonative de Hirst & Di Cristo. Le SA, situé au-dessus de l'unité tonale de Hirst & Di Cristo, correspond au « mot prosodique » de Vaissière, à « l'intonation group » de Mertens, « l'intonème mineur » de Rossi. Le Syntagme intermédiaire (Si) constitue un niveau hiérarchiquement intermédiaire entre les 2 précédents. D'après le modèle de Jun & Fougeron, le Syntagme Intonatif est marqué soit par une montée de continuation finale majeure de la fréquence fondamentale (F_0), qu'elles nomment ton de frontière H%, soit par une chute finale majeure de F_0 (ton de frontière L%), ainsi que par un allongement final, éventuellement suivi d'une pause. La représentation tonale sous-jacente proposée pour le Syntagme Accentuel est /LHiLH*/, avec un ton haut initial Hi (aussi nommé sommet de « l'accent secondaire »), un ton haut final H* réalisé sur la dernière syllabe pleine (sommet de « l'accent primaire ») et deux tons bas L réalisés sur les syllabes précédant les syllabes portant un ton H. Enfin, Jun & Fougeron ont proposé un Syntagme intermédiaire qui porte un ton de syntagme (T-) – qui peut correspondre à une montée (H-) ou une descente (L-) – couvrant l'espace entre le dernier ton H* et le ton de frontière et n'apparaissant par conséquent que lorsque le dernier H* est réalisé en position non finale dans le syntagme intonatif. Aux représentations sous-jacentes ainsi définies peuvent correspondre diverses réalisations phonétiques effectives. Au niveau du Syntagme Accentuel, les patrons tonals possibles dépendent par exemple du nombre de syllabes constituant ce syntagme (voir Jun & Fougeron, 2000). Ainsi, on pourra observer, pour un SA comportant trois syllabes (comme « Le chat blanc »), les contours possibles suivants : LH*, LLH*, LHiH*, HiLH*. Il reste à préciser lequel ou lesquels de ces contours peuvent subsister après application d'un certain nombre de contraintes (absence de pic tonal sur les mots-outils (« le »), limitation des collisions accentuelles (HiH*), etc.). Au niveau du Syntagme Intonatif, pour l'exemple précédent, l'interrogation (« Le chat blanc ? ») pourra être marquée par un contour de frontière montant H% alors que le syntagme déclaratif sera marqué par un contour descendant L%. Ainsi la modulation du niveau SA par le niveau SI (ou, en termes phonologiques, la préemption par le ton de

niveau supérieur) donne finalement les contours LH%, LLH%, LHiH% et HiLH% pour le syntagme interrogatif et LL%, LLL%, LHiL% et HiLL% pour le syntagme déclaratif.

Les travaux de recherche que j'ai encadrés sur les corrélats acoustiques de la prosodie du français fournissent des résultats compatibles avec ce modèle. Au cours du stage de DEA de Guillaume Rolland, nous avons recueilli un corpus, en français, mettant en œuvre des Syntagmes Accentuels avec diverses structures syllabiques et dans diverses conditions d'élocution (variation du débit et de l'emphase). Les résultats de l'étude acoustique (Rolland & Lœvenbruck, 2002, 2003) nous ont permis de montrer, pour deux locuteurs et deux locutrices francophones, que les syntagmes nominaux sujets, situés en position initiale d'énoncés assertifs, possédaient bien, quand le nombre de syllabes était compris entre quatre et cinq, un patron tonal de type [LHiLH*], tel que le prévoient Jun & Fougeron (2002) et comme illustré sur la Figure II-4. De plus, les résultats révèlent une différence entre l'accent secondaire et l'accent primaire. Ils suggèrent que si la séquence LH* est perçue comme un accent primaire c'est probablement parce qu'elle atteint des valeurs de F_0 plus élevée et parce que la montée de F_0 est plus dynamique que celle de LHi (amplitude plus large et durée plus courte) et portée par une seule syllabe.



*Figure II-4. Suivi de F_0 et durées syllabiques pour
 {[Le rat marron]_{SA1} [voulait manger]_{SA2} [le long mulot.]_{SA3}SI.
 La transcription prosodique correspondante est {[LHiLH*] [LH*] [LHiL%]}.*

Ces données acoustiques sont compatibles avec des données articulatoires sur les mouvements de la langue que j'avais recueillies antérieurement à l'Institut de la Communication Parlée, en utilisant un électromagnétomètre (EMA, Carstens), comme illustré sur la Figure II-5.

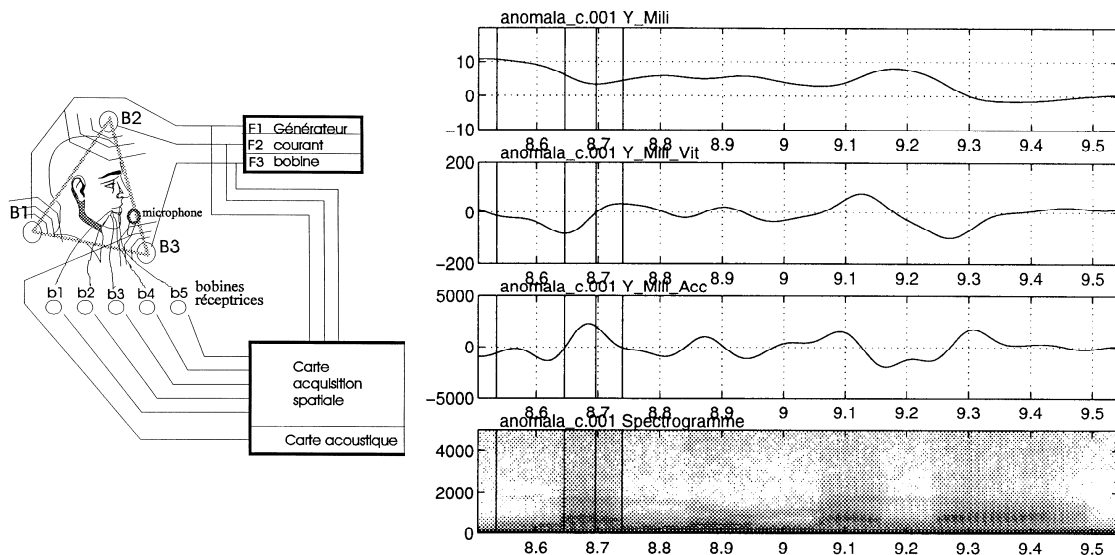


Figure II-5. Recueil de données prosodiques articulatoires
A gauche : Suivi des mouvements de la langue avec l'électromagnétomètre. A droite : repérage d'événements sur les tracés de la position verticale (mm), de la vitesse (mm/s) et de l'accélération (mm/s²) d'une bobine placée sur le milieu de la langue.

Dans cette étude (Lœvenbruck, 1999 ; 2000), j'avais observé les durées et les amplitudes des mouvements de la langue pour les accents initiaux (LHi) et finaux (LH*) chez deux locuteurs, dans deux modes d'élocution (assertion neutre et focalisation prosodique). Dans la condition d'assertion neutre, l'accent initial était, dans la plupart des cas, plus faible que l'accent final, en termes de durée et d'amplitude des mouvements linguaux. Dans la condition de focalisation, pour le locuteur masculin, l'effet était minime et se concentrait sur le renforcement de l'accent final. Pour la locutrice, la focalisation avait plus d'impact et renforçait l'accent initial qui devenait similaire à l'accent final. Ainsi ces données acoustiques et articulatoires suggèrent qu'en français, l'accent final est plus fort que l'accent initial. Le syntagme accentuel peut donc bien être décrit comme associé au patron par défaut /LHiLH*/ avec les corrélats acoustiques et articulatoires suivants : un pic de F_0 , une amplitude du mouvement de la langue et une durée de la syllabe portant le pic intonatif tous plus élevés pour l'accent final que pour l'accent secondaire.

Pour vérifier que ces marques acoustiques et articulatoires permettent bien aux auditeurs de découper l'énoncé en unités, nous avons mené une étude perceptive sur ces données acoustiques. Nous avons présenté aux auditeurs ces énoncés dans leur forme originale ou dans une version délexicalisée, i.e. resynthétisés en réitérant la syllabe /ma/ par synthèse par concaténation (algorithme PSOLA). Dans ces énoncés délexicalisés, la durée et la fréquence fondamentale originales étaient ainsi conservées, mais le lexique et la syntaxe étaient supprimés. Cinq types de découpages des énoncés en unités plus petites étaient proposés aux auditeurs. Les découpages choisis par les auditeurs sont conformes aux résultats attendus dans la majorité des cas : les auditeurs retrouvent bien la structuration en syntagmes accentuels. Ainsi, les auditeurs francophones sont capables de découper correctement des énoncés en unités prosodiques, sur la seule base de l'intonation (durée et fréquence fondamentale). De plus, les scores de découpage sont les meilleurs (les plus conformes aux énoncés planifiés par les locuteurs) pour les énoncés effectivement prononcés avec le patron par défaut [LHiLH*]. Il semble donc que le syntagme accentuel soit bien perçu comme une unité prosodique structurant l'énoncé.

II.1.2.2. Alignement tonal en français : données acoustiques et articulatoires

L'alignement tonal peut être défini comme la synchronisation entre les cibles tonales (pics et creux de F_0) et les unités segmentales (le texte) d'un énoncé (e.g. Silverman & Pierrehumbert, 1990). Il joue un rôle pragmatique important, facilitant le suivi d'autrui. En français, on trouve souvent une montée de F_0 signalant la fin d'un syntagme prosodique. Cette *montée finale* est facultativement précédée d'une *montée initiale* au début du syntagme. Le début de la montée initiale est associé à la frontière entre un mot outil et un mot de contenu. Des « coudes » sont observés à cette frontière, même sans montée initiale (Welby, 2003) et signaleraient le début du syntagme. Au cours du post-doctorat de Pauline Welby, nous nous sommes intéressées à l'alignement du pic de la montée finale, censé être un accent « fixe ». Les résultats vont contre la notion d'ancrage segmental fixe (prôné pour certaines langues) et introduisent la notion d'ancre flottante (Welby & Lœvenbruck, 2006) : le pic (le ton H) de la montée finale n'est pas aligné de manière uniforme selon la structure syllabique. Il est aligné à la fin de la syllabe pour les syllabes ouvertes ou les syllabes fermées par des obstruantes, mais parfois à la fin de la voyelle (Figure II-6.a), parfois dans la coda (Figure II-6.b) pour les syllabes fermées par des sonantes. Ces résultats renseignent sur la façon dont les locuteurs manient la prosodie pour aider autrui à décoder le signal de parole.

Dans le cadre d'une collaboration avec le Laboratoire Parole et Langage, un corpus de données acoustiques et articulatoires a été recueilli. La coordination entre gestes articulatoires et coude de la montée initiale de F_0 a été examinée. Les résultats indiquent que le coude est coordonné avec le pic de vitesse de l'articulateur principal associé à l'attaque de la syllabe. Le pic de vitesse étant lié à l'effort articulatoire, ce résultat suggère qu'une synchronisation temporelle fine existe entre efforts laryngé et supralaryngé, qui faciliterait le découpage de l'énoncé pour les interlocuteurs (D'Imperio *et al.*, 2007).

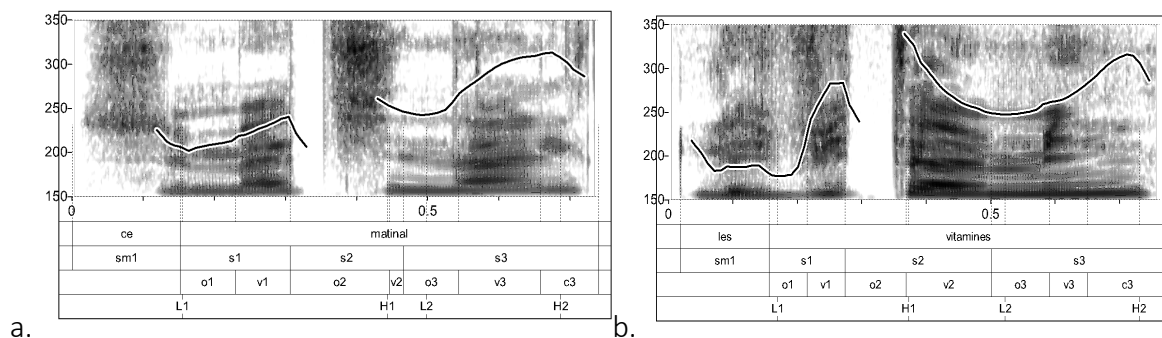


Figure II-6. Variabilité de l'alignement pour une même syllabe CVC : a. le pic de la montée finale est aligné avec la voyelle ; b. le pic est aligné avec la consonne coda

II.1.2.3. Ironie et sarcasme

D'autres formes prosodiques fournissent à autrui des outils pour suivre nos productions verbales. L'ironie est un mode d'expression dans lequel ce qui est dit diffère de ce qui est signifié. Il existe un débat sur le rôle de la prosodie dans l'interprétation d'un énoncé ironique. Avec Mariapaola D'imperio et Maud Champagne-Lavau du Laboratoire Parole et Langage à Aix-en-Provence, nous avons étudié les marqueurs acoustiques et la perception du sarcasme, une forme particulière d'ironie.

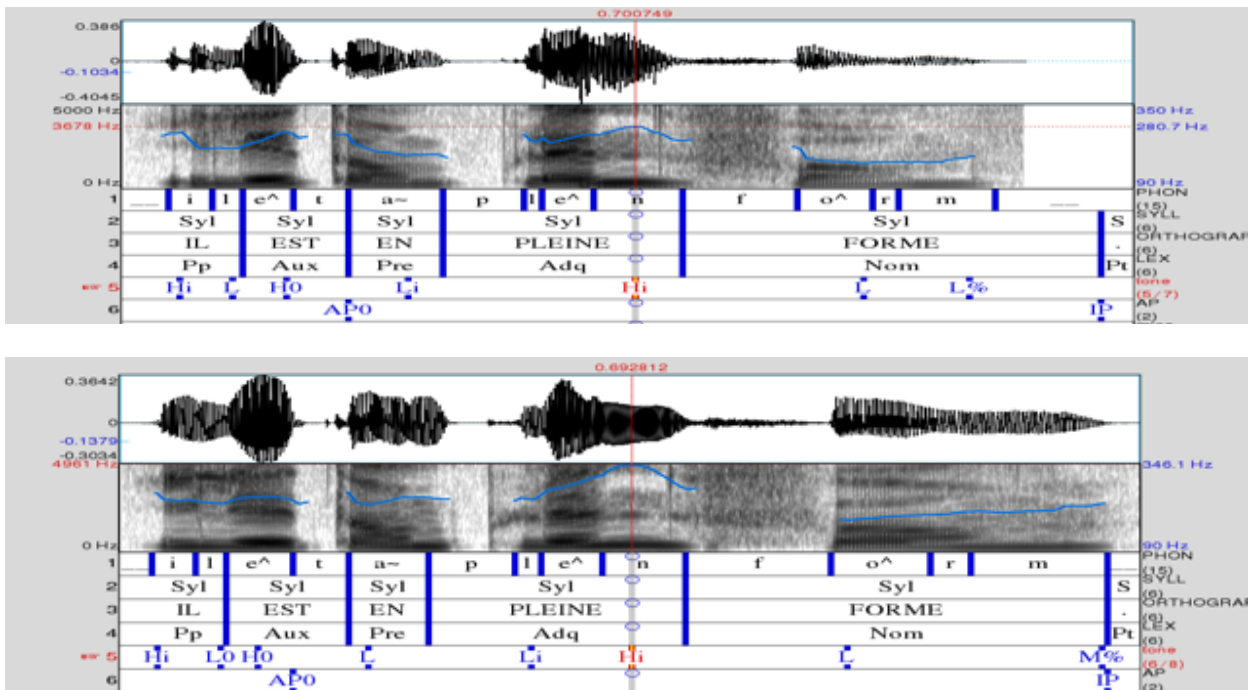


Figure II-7. Signaux acoustiques correspondant aux productions littérale (en haut) et sarcastique (en bas) de la phrase « il est en pleine forme » chez une locutrice française. La F_0 atteint au maximum 281Hz dans la version littérale et 346 Hz dans la version sarcastique.

Les données acoustiques que nous avons recueillies suggèrent qu’il existe en français des marqueurs acoustiques qui véhiculent le sarcasme sans que le contexte soit nécessaire (cf. Figure II-7). Notre étude suggère que la valeur moyenne et l’étendue de F_0 sont des paramètres sur lesquels les locuteurs jouent pour produire le sarcasme. L’augmentation du nombre de pics de F_0 semble aussi être une stratégie. Enfin, l’allongement de l’énoncé, et du constituant syntaxique sémantiquement fort (l’adjectif) en particulier, constitue une dernière ressource pour transmettre le sarcasme (Lœvenbruck *et al.*, 2013).

II.1.2.4. Effet Lombard sur la prosodie : données acoustiques et articulatoires

Il a été montré que dans le bruit, les locuteurs parlent plus fort, plus lentement et plus haut. C’est ce que l’on nomme « l’effet Lombard » (Junqua, 1993). Les manifestations articulatoires de l’effet Lombard sont mal connues. Ses corrélats intonatifs sont également peu décrits. Si l’on sait que l’étendue de la fréquence fondamentale est globalement plus large (les locuteurs parlent plus haut), on ne sait pas si les pics et creux de F_0 sont plus ou moins marqués, ni quelles parties de la courbe F_0 sont rehaussées. J’ai participé à une étude menée par Maëva Garnier, utilisant la plate-forme de suivi labiométrique de l’Institut de la Communication Parlée. Nous avons montré que l’effet Lombard s’accompagne d’une augmentation significative de l’aire intéro-labiale, liée à une hyper-articulation. Nous avons également observé des stratégies visant à renforcer des indices de structuration du discours au niveau de la syllabe, du mot et de l’énoncé. Ainsi, le renforcement dans le bruit de la dynamique d’intensité par syllabe pourrait contribuer à une meilleure délimitation des syllables d’un mot. L’augmentation de la force de pincement des lèvres sur les consonnes bilabiales en début de mot de contenu peut être liée à une stratégie visuelle ou acoustique de renforcement de l’indice de début du mot. Nous avons également observé un allongement plus important dans le bruit des syllabes initiale et finale d’énoncé, de même qu’une augmentation de l’ambitus tonal de

la chute mélodique finale (Garnier *et al.*, 2006). Nous avons aussi observé des différences sur les modifications entraînées par l'effet Lombard en fonction de la nature des mots (déterminants ou mots de contenu) qui suggèrent que les locuteurs sont capables de faire varier leur articulation et leur prosodie pour permettre à autrui de mieux les suivre.

Publications associées (articles de revues, chapitres, congrès, mémoire de master)

- D'Imperio M., Espesser R., Løevenbruck H., Menezes C., Nguyen N. & Welby P. (2007). Are tones aligned to articulatory events? Evidence from Italian and French. Cole, Jennifer and José I. Hualde (eds.), *Laboratory Phonology 9*, Berlin: Mouton de Gruyter, 577-608. - hal-00244489
- Garnier M., Bailly L., Dohen M., Welby P., Løevenbruck H. (2006). Étude acoustique et articulatoire de la parole Lombard : Effets globaux sur l'énoncé entier. *Actes des XXVIèmes Journées d'Etude sur la Parole*, 12-16 juin 2006, Dinard, France, 69-72. - hal-00214304
- Garnier M., Bailly L., Dohen M., Welby P., Løevenbruck H. (2006). An Acoustic and Articulatory Study of Lombard Speech: Global Effects on the Utterance. *Proceedings of Interspeech 2006: the 9th International Conference on Spoken Language Processing*, 2246-2249, Pittsburgh, Pennsylvania, USA, September 17-22, 2006. - hal-00370947
- Garnier M., Dohen M., Løevenbruck H., Welby P. & Bailly L. (2006). The Lombard Effect: a physiological reflex or a controlled intelligibility enhancement? *Proceedings of the 7th International Seminar on Speech Production (ISSP 2006)*, 255-262, Ubatuba, Brazil, December 13-15. - hal-00214307
- Løevenbruck H. (1999). Articulatory effects of contrastive emphasis on the Accentual Phrase in French. *J. Acous. Soc. Am.*, 106 (4), Pt.2, 2153.
- Løevenbruck 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.
- Løevenbruck H., Ben Jannet M. A., D'Imperio M., Spini M. and Champagne-Lavau M. (2013). Prosodic cues of sarcastic speech in French: slower, higher, wider. *Proceedings of INTERSPEECH 2013, 14th Annual Conference of the International Speech Communication Association*, 25-29 August 2013, Lyon, France, 3537-3541. - hal-00864346
- Rolland G. & Løevenbruck H. (2002). Physiological and perceptual relevance of the Accentual Phrase in French. *Proceedings of the 1st International Conference on Speech Prosody*, Aix-en-Provence, France, 611-614.
- Rolland G., Løevenbruck H. & Cathiard M.-A. (2002). Some perceptual cues to French prosody. *Proceedings of the ISCA International Tutorial and Research Workshop on the Temporal Integration in the Perception of Speech*, Aix-en-Provence, France, 48.
- Welby P. & Løevenbruck 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.). - hal-00371892

Mémoire de DEA et M2R : G. Rolland, 2000 ; M. Ben Jannet, 2012

Post-doctorat : Pauline Welby, 2003-06

Collaborations :

Grenoble : GIPSA-lab, LPNC

Aix-en-Provence : Laboratoire Parole et Langage

II.1.3. Suivi visuel de soi et d'autrui

Nous avons des représentations auditives et visuelles de notre propre voix et de notre propre visage. Nous sommes capables de nous suivre auditivement pour contrôler nos propres productions verbales (cf. I.2.1.1.2). Nous contrôlons-nous aussi visuellement ? Quelles représentations visuelles avons-nous de nous-mêmes et comment se comparent-elles à celles que nous avons d'autrui. Une théorie féconde de la coordination entre action et perception en parole (Schwartz *et al.*, 2012) conçoit la communication comme un échange de gestes perceptivo-moteurs ou de percepts façonnés par l'action. Il en

découle que le suivi visuel de soi devrait être meilleur que le suivi visuel d'autrui. Plusieurs équipes de recherche ont montré qu'il existe un avantage de « soi » sur « autrui » dans la perception visuelle d'actions manuelles comme l'écriture, le lancement de fléchettes ou dans le jeu de piano (Knoblich et al., 2002 ; Knoblich & Flach, 2001 ; Repp & Knoblich, 2004). Des expériences préliminaires initiées par Jean-Luc Schwartz (Arnaud *et al.*, 2008) à GIPSA-lab n'ont toutefois pas réussi à montrer qu'il existerait un avantage de « soi » sur « l'autre » en lecture labiale. Une explication possible de ces résultats dans le domaine de la parole serait que les expériences sur les actions de lancement de fléchettes, sur l'écriture ou sur le piano, portaient sur des stimuli visuels peu courants et pour lesquels la connaissance des régularités motrices d'autrui est faible. Au contraire, la lecture labiale est d'usage permanent, on est habitué à lire sur les lèvres des autres et à apprendre les spécificités des autres (leurs idiosyncrasies). De plus, si la perception de la parole requiert l'action et si des représentations acoustico-articulatoires intégrées existent, cela ne signifie pas nécessairement que ces représentations sont directement transférées dans le domaine visuel. Des recodages entre les modalités articulatoires et proprioceptives sont probablement nécessaires pour obtenir une représentation visuelle de soi parlant. Nous avons donc co-encadré Vanessa Bernier, en Master, dans le but de prolonger les premières expériences menées par Arnaud et al. (2008), en testant la lecture labiale et l'identification visuelle de soi chez des participants ne se connaissant pas. Ces nouvelles expériences n'ont toujours pas permis de déceler un quelconque avantage du suivi visuel de soi par rapport à autrui. Il semblerait ainsi que le suivi visuel d'autrui soit tellement nécessaire que nous apprenions très rapidement les schémas moteurs visuels d'autrui et ne soyons pas meilleurs pour nous percevoir visuellement nous-mêmes.

Publications associées (congrès avec actes, mémoire)

Arnaud L., Schwartz J.-L., Lœvenbruck H. & Savariaux C. (2008). Perception as a (Shaped) Mirror of Action: It Seems Easier to Lipread One's Own Speech Gestures than those of Somebody Else. *Workshop on Speech and Face to Face Communication*, Grenoble, 27-29 Oct. 2008. - hal-00371907

Mémoire de M2R : Vanessa Bernier, 2012

II.2. MONITORING VERBAL POUR LA FONCTION DE PENSEE

II.2.1. Monitoring du langage intérieur

Comme je l'ai décrit plus haut (sections I.1.2 et I.2.2), le langage intérieur joue un rôle central dans le raisonnement, la résolution de problème, la pensée. Comme le langage à voix haute, il requiert un *monitoring*, mais ses liens avec la parole à voix haute font l'objet de débats en neurolinguistique.

II.2.1.1. Hypothèses théoriques

Nous avons mené une revue de littérature, récapitulée dans Perrone-Bertolotti *et al.* (2014, 2016) et étendue dans Lœvenbruck *et al.* (2017, 2018) et Lœvenbruck (2018) qui nous a permis de mieux décrire la nature de la parole intérieure et de faire quelques propositions théoriques (présentées au I.2.2). Pour résumer, il existe beaucoup de similitudes comportementales (existence de lapsus par exemple), physiologiques (taux de respiration, débit d'élocution), cérébrales (activation des aires classiquement observées lors de la production du langage, notamment région de Broca, régions

prémotrices, cortex temporal supérieur, lobule pariétal inférieur) et subjectives (hauteur de la voix, accent, rythme) entre la parole à voix haute et la parole intérieure, qui corroborent l'hypothèse selon laquelle la parole intérieure est une forme de parole à voix haute tronquée. Mais le niveau auquel le processus d'exécution est interrompu (articulatoire, phonologique, lexical, conceptuel) reste débattu, et il est nécessaire de recueillir plus de données expérimentales sur les multiples formes du langage intérieur.

II.2.1.2. Données physiologiques

Sur le plan physiologique, nous avons montré, par plusieurs études en électromyographie (EMG) labiale de surface (cf. Figure II-8) que la parole intérieure délibérée semble recruter le système moteur tout comme la parole à voix haute (Duval, 2012 ; Løevenbruck *et al.*, en préparation ; Nalborczyk *et al.*, en préparation). Nous observons une augmentation de l'activité des muscles orbiculaires des lèvres durant la parole intérieure par rapport à un état de relaxation orofaciale. Cette augmentation d'activité musculaire n'est pas liée à une tension musculaire générale accrue, car elle n'est pas retrouvée dans le muscle du bras. Ces résultats suggèrent que certaines formes de parole intérieure impliquent une planification et un *monitoring*, descendant jusqu'au niveau articulatoire.



Figure II-8. Positionnement des capteurs électromyographiques de surface pour mesurer l'activité des muscles labiaux lors de la production de parole intérieure.

Toujours sur le plan physiologique, mais en étendant les mesures à d'autres muscles faciaux ainsi qu'au clignement des paupières, nous avons cherché à différencier les corrélats physiologiques de la parole intérieure délibérée, du vagabondage mental verbal et de l'imagerie visuelle. Nous avons équipé les participants de capteurs EMG de surface sur cinq muscles du visage, du menton et du cou (au niveau du front, de l'orbiculaire inférieur des lèvres, du grand zygomatique, des muscles de la langue et du platysma), ainsi que de capteurs électro-oculographiques (EOG), comme indiqué sur la Figure II-9. Quatre conditions ont été étudiées, chez 37 participants : parole à voix haute (*Overt Speech* : OS), parole intérieure délibérée (*Inner Speech* : IS), parole intérieure vagabonde (*Mind Wandering* : MW) et visualisation mentale (*Mental Visualisation* : MV). Contrairement à nos hypothèses initiales, la parole intérieure ne recrute pas plus les muscles orofaciaux que la parole vagabonde ni que la visualisation mentale (qui était censée ne pas comporter de parole intérieure). Nous pouvons attribuer cette absence de différence à une limite de nos capteurs EMG, car nous observons cependant des différences sur les taux de clignements. Le taux de clignement diminue dans cet ordre : OS > IS > MW > MV, ce qui est compatible avec l'hypothèse d'une implication décroissante des processus de planification motrice. Ce dernier résultat est en accord avec l'hypothèse du rôle du clignement dans la communication (Descroix *et al.*, 2018). Il semblerait donc que le *monitoring* d'autrui dans la communication passe aussi par ce canal visuel attentionnel. Une publication de ces résultats est en préparation (Clerc, 2018 ; Clerc *et al.*, en préparation).

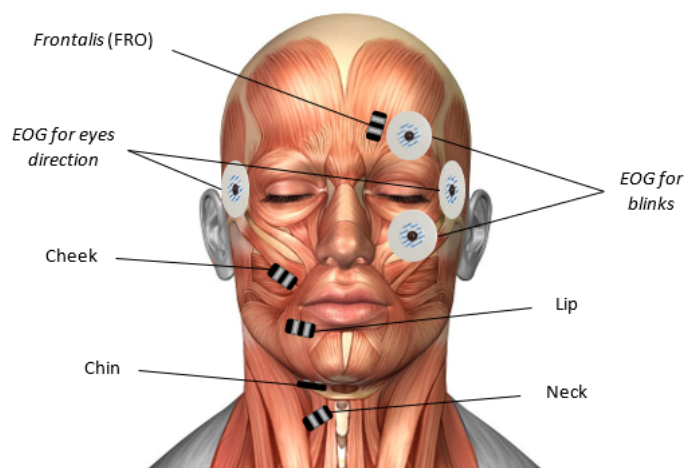


Figure II-9. Positions des capteurs électro-oculographiques et électromyographiques de surface pour mesurer les clignements des yeux et l'activation des muscles orofaciaux.

II.2.1.3. Données de neuroimagerie

Sur le plan neuronal, nous avons montré que la parole intérieure peut utiliser des mécanismes de *monitoring* verbal similaires à ceux de la voix haute. Nous avons recueilli des données d'IRMf chez des participants qui produisaient mentalement des énoncés. Dans la condition contrôle, les participants avaient pour instruction de lire silencieusement la phrase « Madeleine m'amena ». Dans la condition de focalisation prosodique, ils devaient lire silencieusement la phrase comme s'ils répondaient à l'énoncé « Tu as dit Geneviève m'amena ? ». Leur tâche était donc de corriger l'énoncé en apportant une focalisation prosodique sur le prénom : « MADELEINE_F m'amena ». A voix haute, cela correspondrait à une F_0 plus élevée et une durée plus longue sur le constituant focalisé, suivi d'une compression de la F_0 sur les constituants post-focaux (cf. paragraphe II.1.1). Les données d'IRMf montrent que la condition de focalisation prosodique, comparée à la condition contrôle, active plus le cortex prémoteur, le cortex frontal inférieur gauche, l'aire de Wernicke et le gyrus supramarginal gauche (cf. Figure II-10). Ces résultats signifient que lorsque nous lisons silencieusement, nous pouvons utiliser un contour prosodique marqué (typique de la focalisation prosodique), ce qui recrute les régions de la production de la parole à voix haute. Cette activité mentale active également les régions auditives, ce qui corrobore l'impression subjective que la parole intérieure a une qualité auditive (Løevenbruck *et al.*, 2005).

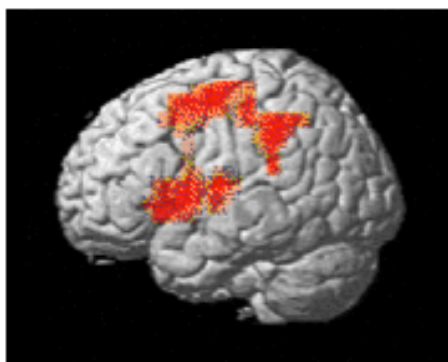


Figure II-10. Réseau cérébral activé lors de la production intérieure avec focalisation prosodique (vs énoncé neutre). Tiré de Løevenbruck, Baci, Segebarth & Abry (2005)

Dans le cadre du projet ANR-13-BSH2-0003-01, INNERSPEECH, nous avons exploré les corrélats physiologiques et neuronaux de la parole intérieure. Nous avons mis en place des protocoles IRMf, iEEG et EMG pour l'étude de différentes formes de parole intérieure, variant sur les dimensions de dialogalité et d'intentionnalité : parole intérieure délibérée (génération mentale de définitions) avec différentes voix (la sienne et celle d'autrui), différentes perspectives (1ère vs. 3ème personne, i.e. monologue vs. dialogue), vagabondage mental verbal (non intentionnel). Nous avons mis au point des tâches permettant d'induire de l'imagerie mentale avec la voix d'autrui et permettant de faire la distinction entre imaginer autrui en train de parler (perspective dialogale allocentrée) et imaginer soi-même imitant autrui (perspective monologique égocentrée).

Le recueil de données iEEG est encore en cours, la rareté des patients implantés dans les aires corticales d'intérêt ayant rendu cette partie du projet difficile à mettre en œuvre.

Nous avons recueilli des données IRMf chez 24 participants. Les résultats font l'objet de publications (Løevenbruck *et al.*, ICPhS 2019 ; Grandchamp *et al.* numéro spécial de *Frontiers* sur la parole intérieure, soumis ; voir aussi Grandchamp *et al.*, 2016). Nous avons comparé 5 tâches : génération de définitions de mots en parole intérieure délibérée avec sa propre voix (*Monological Self voice inner speech: MS*), génération de définitions en parole intérieure délibérée avec la voix d'un avatar (*Monological Other voice inner speech: MO*), imagination de l'avatar s'adressant à soi et produisant des définitions (*Dialogical Other voice inner speech: DO*), perception de parole (*Speech Perception : SP*) et vagabondage mental avec rapports *a posteriori* d'épisodes verbaux (*Verbal Mind Wandering : VMW*). Nos résultats sont compatibles avec le modèle de production de parole intérieure présenté dans la section I.2.2.2 et Figure I-15.

Nous avons observé en particulier que, comparativement à la ligne de base, **la perception de la parole** recrute les GTS bilatéraux, le GSM gauche, le GFI gauche, le gyrus frontal supérieur (GFS) gauche, le cortex PM bilatéral, l'AMS gauche, le cortex moteur gauche, l'hippocampe gauche (cf. Figure II-11 A). Ces résultats sont conformes aux études antérieures sur la perception de parole. Comparée à la ligne de base, **la parole intérieure monologique avec voix propre (MS)** correspond à plus d'activation, latéralisée à gauche, du gyrus frontal inférieur gauche (IFG), du gyrus frontal moyen (GFm), du GFS, du GSM gauche, du GTM postérieur et de l'hippocampe, ainsi qu'à des activations bilatérales de l'AMS, du cortex PM et à une activation du cervelet droit (cf. Figure II-11 B). Comparée à la ligne de base, **la parole intérieure monologique avec voix d'autrui (MO)** est associée à plus d'activation, latéralisée à gauche, du GFI, du GFm, de l'hippocampe, ainsi qu'à des activations bilatérales du cortex PM, de l'AMS et à une activation du cervelet droit (cf. Figure II-11 C). Comparée à la ligne de base, **la parole intérieure dialogale avec voix d'autrui (DO)** correspond à plus d'activation, latéralisée à gauche, du GFm et du gyrus occipital moyen ainsi qu'à des activations bilatérales du cortex PM, du GFI et de l'AMS (cf. Figure II-11 D). Comparé à la ligne de base, **le vagabondage mental verbal (VMW)** correspond à plus d'activation de l'AMS gauche, ainsi qu'à des activations bilatérales du GFI, de l'insula, du GFm, du GFS médian, des cortex pariétaux inférieur et supérieur, du précuneus, et des activations du caudé gauche, du thalamus gauche et du cervelet gauche (cf. Figure II-11 E).

De plus, pour examiner les corrélats neuronaux communs de la parole intérieure et de la perception de parole, des conjonctions entre la condition SP et chacune des 4 conditions

de parole intérieure (MS, MO, DO et VMW) ont été analysées. Elles ont montré que le GFI, le GFS de l'hémisphère gauche et le cortex PM bilatéral étaient activés conjointement par ces 5 conditions. Un recrutement commun du GTM a également été observé dans la conjonction SP et MS. Les conjonctions entre SP et MO, DO et VMW ont cependant révélé une activation temporelle moindre. Dans les conditions MO et DO, les modèles internes à utiliser sont paramétrés pour autrui (la voix de l'avatar) et sont donc probablement moins précis (moins bien maîtrisés) que dans la condition MS. De ce fait, ils produisent probablement des prédictions auditives plus fragiles, ce qui pourrait expliquer l'activation moindre du cortex auditif. Cette interprétation est d'ailleurs confirmée par l'expérience subjective des participants qui ont rapporté une perception affaiblie de la petite voix dans ces deux conditions plus exigeantes cognitivement.

Ces données sont en accord avec le modèle théorique de la production de parole intérieure, présenté dans la section 1.2.2.2, selon lequel la parole intérieure délibérée correspond à des objectifs supramodaux générés dans le LPI et convertis en commandes motrices via le cervelet et les régions frontales gauche. Ces commandes motrices sont inhibées et leur copie d'efférence serait traitée en parallèle par le modèle prédictif cérébelleux, ce qui donnerait lieu à une prédiction sensorielle, la petite voix, dans le GTS/GTM. Ils sont aussi compatibles avec le modèle de la perception de parole, présenté au 1.2.1.2.2, selon lequel la réception de signaux de parole entraîne un traitement par les régions auditives (GTS/GTM), une interprétation et une prédiction par le GFI et le LPI.

Concernant la dimension de dialogalité, le fait d'utiliser la voix d'autrui (MO) vs. sa propre voix (MS) a conduit à une bascule hémisphérique : nous avons observé des activations frontale (GFI, GFS) et pariétale plus importante à gauche pour MS par rapport à MO et à droite pour MO comparé à MS. Le fait de changer de perspective (Figure II-11 C et D), i.e. de passer de l'imitation de la voix d'autrui (MO) à l'imagination d'autrui parlant (DO) a résulté en une activation plus importante du GFI, de l'AMS et du cortex cingulaire antérieur (CCA) à gauche pour MO vs. DO et une activation plus importante du GFI, du GFS, du GFM des lobules pariétaux supérieur et inférieur à droite ainsi que du précuneus bilatéral et du cortex cingulaire postérieur bilatéral pour DO vs. MO. La bascule hémisphérique à droite du GFI lors de la parole intérieure avec la voix d'autrui (par rapport à la voix propre) est en accord avec les études sur la prosodie (Baum & Pell, 1999 ; Løevenbruck *et al.*, 2005 ; Pichon & Kell, 2013). Les activations pariétales et du précuneus observées sont cohérentes avec les études sur la prise de perspective à la 3^{ème} personne (Alderson-Day *et al.*, 2016 ; Blanke, 2012 ; Ruby & Decety, 2001).

Concernant la dimension de l'intentionnalité, le contraste entre les conditions MS et VMW a révélé des activations dans l'AMS, le cortex moteur primaire, le GFI, l'insula, le GTM/GTS, le CCA, le putamen, le noyau caudé, à gauche et dans le cortex PM bilatéral. Le contraste inverse (VMW > MS) a révélé des activations dans le cortex pariétal inférieur, le précuneus, le GFI, le GFS, le GFM, l'insula, le CCA, le thalamus, à droite et le GFS gauche. Ces résultats sont conformes aux études sur le vagabondage mental et le réseau par défaut (Raichle *et al.*, 2015). De plus, l'activation temporelle est moindre que dans les conditions de parole intérieure délibérée, ce qui correspond à l'impression d'évanescence de la parole intérieure vagabonde (et ce qui est conforme à nos propositions théoriques, avec une condensation probable de la parole intérieure lors du vagabondage par rapport aux formes plus délibérées, cf. 1.2.2.2). Des analyses sont en cours sur cette dernière condition pour mieux séparer les activations spécifiques aux épisodes verbaux.

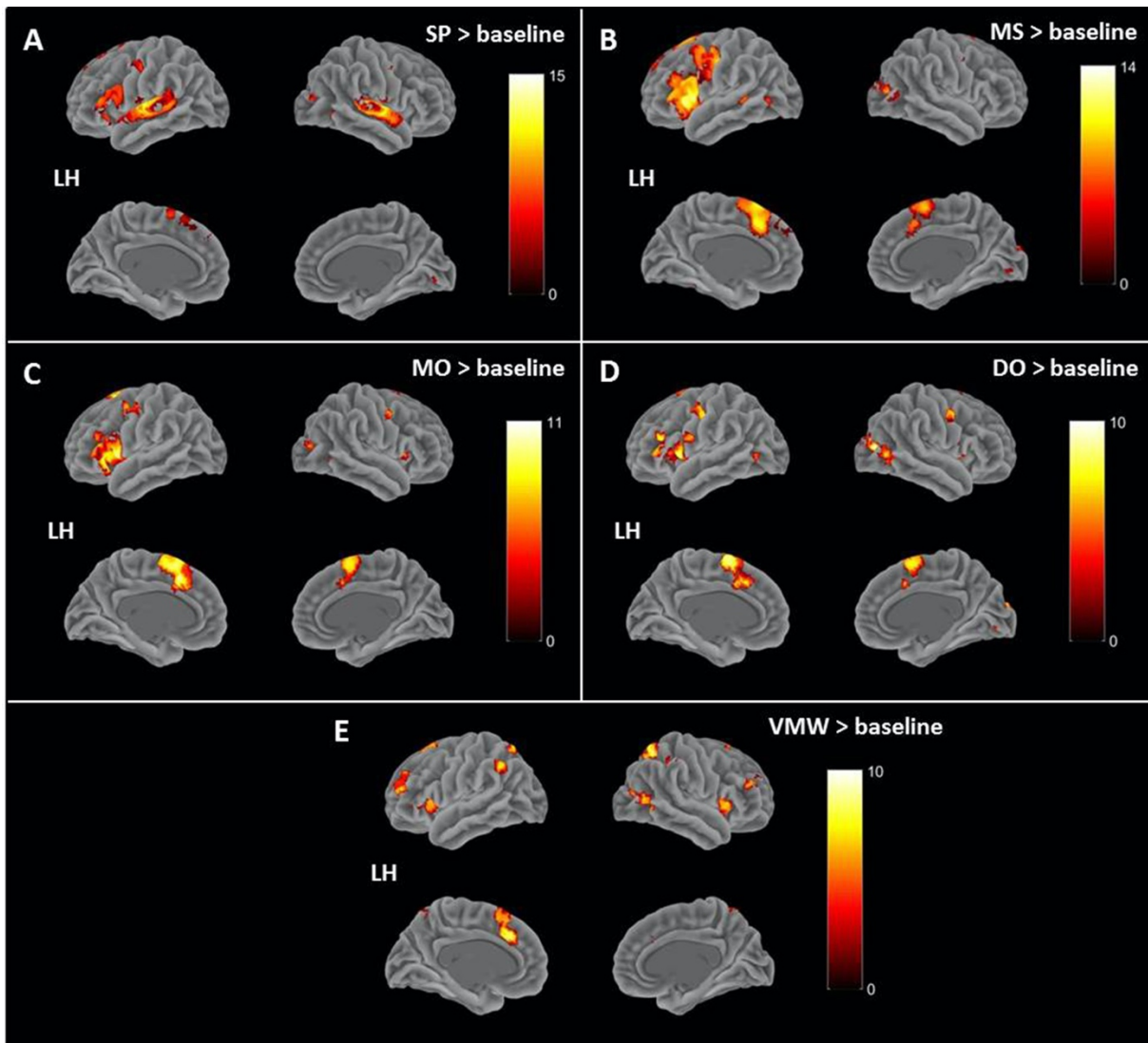


Figure II-11. Contrastes entre chaque condition et la ligne de base, tracés sur une image de cerveau standard (issue de BSPMview, Spunt, 2016).
(A) SP vs. ligne de base (B) MS vs. ligne de base (C) MO vs. ligne de base (D) DO vs. ligne de base (E) VMW vs. ligne de base. Tous les contrastes sont avec correction FWE, $p < 0.05$.

II.2.1.4. Données introspectives

Dans le cadre d'une collaboration avec Stéphanie Smadja (CERILAC, Université Paris Diderot), responsable du projet Monologuer (<http://www.monologuer.fr>), nous avons recueilli des données électromyographiques et des données subjectives introspectives lors de l'induction de rumination ou de parole intérieure abondante. Nous avons ainsi adapté la grille d'entretien d'introspection sur les manifestations de la parole intérieure, utilisée dans le projet *Monologuer* (Smadja, 2018). Nous avons mis en place un protocole en 3 étapes : entretien introspectif préliminaire, puis carnet endophasique introspectif tenu pendant 3 semaines, puis enfin 2ème entretien introspectif et mesures physiologiques lors de l'induction de rumination ou de parole intérieure abondante (EMG faciale et brachiale). Nous avons effectué des enregistrements audio et EMG de 27 participants : en 2 sessions. L'analyse et l'interprétation des données audio (entretiens), textuelles (carnets) et EMG est en cours.

Publications associées (articles de revues, chapitres, congrès, mémoires, sites web)

- Grandchamp R., Rapin L., Løevenbruck H., Perrone-Bertolotti M., Pichat C., Lachaux J.P., Baciú M. (2016). Inner Speech with your own or someone else's voice. Cerebral correlates assessed with fMRI. *Society for the Neurobiology of Language (SNL) Conference 2016*, London 17-20 August 2016, Poster B66.
- Grandchamp R., Rapin L., Perrone-Bertolotti M., Pichat C., Haldin, C., Lachaux J.P., Dohen M., Perrier P., Garnier M., Baciú M., Løevenbruck H. (soumis). The Condialint model: a neurofunctional account of varieties of inner speech along condensation, dialogicality and intentionality dimensions. *Frontiers in Psychology – Exploring the Nature, Content, and Frequency of Intrapersonal Communication*.
- Løevenbruck, H. (2018). What the neurocognitive study of inner language reveals about our inner space. *Épistémocritique, n° 18 : Langage intérieur - Espaces intérieurs / Inner Speech - Inner Space*, Stéphanie Smadja, Pierre-Louis Patoine (eds.) [<http://epistemocritique.org/what-the-neurocognitive-study-of-inner-language-reveals-about-our-inner-space/>] - hal-02039667
- Løevenbruck, H., Nalborczyk, L., Perrone-Bertolotti, M., Baeyens, C., Grandchamp, R. (2017). Un cas d'incarnation linguistique : implication du système moteur orofacial dans les ruminations mentales. *La cognition incarnée*, J. Dokic et D. Perrin (dir.), Recherches sur la philosophie et le langage, N° 33, Vrin, Paris, 297-324, ISBN : 978-2-7116-8423-6. - hal-01887636
- Løevenbruck, H., Grandchamp, R., Rapin, L., Nalborczyk, L., Dohen, M., Perrier, P., Baciú M. & Perrone-Bertolotti, M. (2018). A cognitive neuroscience view of inner language: to predict, to hear, to see and to feel. In *Inner Speech: New Voices*. Peter Langland-Hassan & Agustín Vicente (eds.), Oxford University Press, 131-167 - hal-01898992
- Løevenbruck, H., Grandchamp, R., Rapin, L., Nalborczyk, L., Dohen, M., Perrier, P., Baciú, M., & Perrone-Bertolotti, M. (2018). A neurocognitive predictive model of willful inner speech production. *The Probabilistic Brain Workshop*, Durham, UK, March 23-24, 2018.
- Løevenbruck H., Grandchamp R., Rapin L., Perrone-Bertolotti M., Pichat C., Haldin, C., Dohen M., Perrier P., Lachaux J.P., Garnier M., Baciú M. (accepté). Neural correlates of inner speaking, imitating and hearing: an fMRI study. *19th International Congress of Phonetic Sciences (ICPhS 2019)*.
- Perrone-Bertolotti, M., Rapin L., Lachaux J.-P., Baciú M. & Løevenbruck H. (2014). What is that little voice inside my head? Inner speech phenomenology, its role in cognitive performance and its relation to self-monitoring. *Behavioural Brain Research*, 261, 220-239. Epub: 10.1016/j.bbr.2013.12.034. - hal-00967730
- Perrone-Bertolotti M., Grandchamp R., Rapin L., Baciú M., Lachaux J.P. & Løevenbruck H. (2016). Langage intérieur. In Pinto, S. & Sato, M. (eds.). *Traité de Neurolinguistique : du cerveau au langage*. Collection Neuropsychologie, De Boeck Supérieur, 109-123. - hal-01467107
- Rapin L., Dohen M. & Løevenbruck H. (2016). Les hallucinations auditives verbales. In Pinto, S. & Sato, M. (eds.). *Traité de Neurolinguistique Neurolinguistique : du cerveau au langage*. Collection Neuropsychologie, De Boeck Supérieur, 347-370. - hal-01442463

Mémoire de M2R : Caroline Duval, 2012 ; Olivier Clerc, 2018

Post-doctorat : Romain Grandchamp, 2014-18

Sites web :

Projet ANR InnerSpeech : <https://lpnc.univ-grenoble-alpes.fr/InnerSpeech-2014-2018>

Projet MONOLOGUER, S. Smadja, Univ. Paris Diderot : <https://cerilac.univ-paris-diderot.fr/monologuer>

Impromptu Scientifique « des voix dans la tête » : <http://www.ateliers-du-spectacle.org/spectacle/les-impromptus-scientifiques/#voix>

Financements :

Ce thème a été financé par le projet APN CNRS 1999-2001, No. 7904 (De la prosodie à la syntaxe : Recherche de patrons d'activation cérébrale par IRM fonctionnelle), le projet ANR INNERSPEECH de 4 ans 2014-2018 (<http://lpnc.univ-grenoble-alpes.fr/InnerSpeech>) et par le projet MONOLOGUER porté par Stéphanie Smadja de l'Université Paris Diderot (<http://www.monologuer.fr>).

Collaborations :

Grenoble : GIPSA-lab, LPNC, IRMaGe, GIN

Lyon : CRNL

Paris : CERILAC

Canada : CNoS, University British Columbia

Italie : Univ. Ferrara

II.2.2. Les ruminations mentales : déficit de contrôle de la parole intérieure

La parole intérieure joue la plupart du temps un rôle bénéfique fondamental. Elle peut toutefois jouer un rôle négatif et perturber les performances cognitives, comme dans le cas des ruminations mentales. La rumination mentale est souvent une forme de parole intérieure négative et excessive : les mêmes pensées verbales, en général négatives, sont ressassées. Dans le cadre de la thèse en co-tutelle de Ladislav Nalborczyk, que je co-encadre avec Ernst Koster de l'Université de Ghent, en Belgique, ainsi que Marcela Perrone-Bertolotti, Elsa Spinelli (LPNC) et Céline Baeyens (LIP), nous cherchons à savoir si les ruminations verbales correspondent à une simple évocation de représentations auditives en mémoire ou s'il s'agit de la re-création de percepts acoustiques et articulatoires via une *simulation motrice* du processus de production de parole, comme j'en fais l'hypothèse dans le modèle de parole intérieure décrit dans la section I.2.2.2. Nous avons mesuré les corrélats physiologiques de ruminations verbales induites chez des personnes sans trouble psychiatrique. Nos résultats suggèrent que l'induction de rumination entraîne une augmentation de l'activité EMG labiale, par rapport à un état de repos (Nalborczyk *et al.*, 2017), concomitante avec un score de rumination (échelle *brooding*) élevé (Figure II-12). Nous avons également observé une augmentation de l'activité EMG dans le front, peut-être liée à la présence d'émotions négatives. Ceci semble conforme à l'hypothèse que la production de parole intérieure, notamment lorsqu'elle est excessive, implique le système moteur lié à la production de parole. Nous observons ici les corrélats physiologiques d'un cas de défaillance du *monitoring* verbal égo-centré.

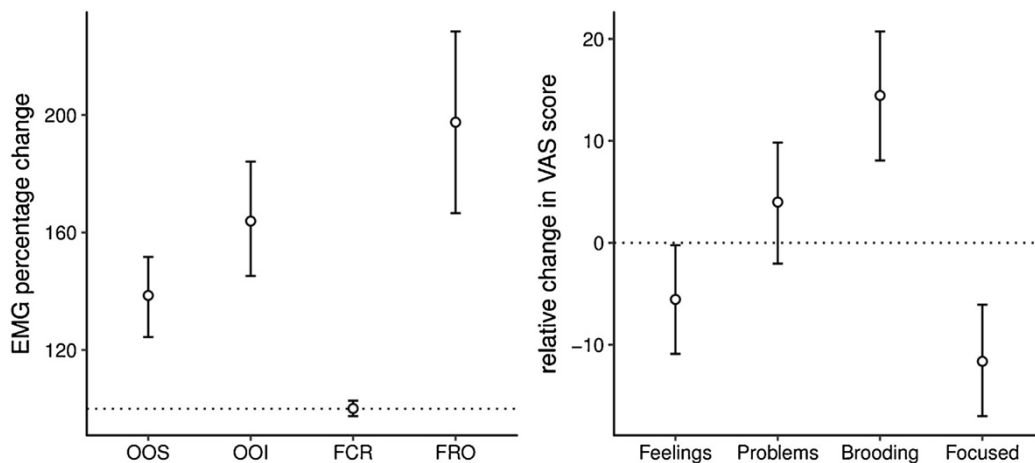


Figure II-12. Activation EMG (lèvre supérieure, lèvre inférieure, bras, front) et scores de rumination après induction de ruminations

A gauche : moyenne a posteriori et intervalle de crédibilité à 95% pour l'amplitude des signaux EMG (exprimé en pourcentage de la ligne de base) ; OOS : Orbicularis Oris Superior, OOI ; Orbicularis Oris Inferior ; FCR : Flexor Carpi Radialis (bras) lèvres, FRO : Frontalis (front) ; à droite : scores de rumination, l'échelle brooding reflète le ressassement négatif

Nous avons mené une étude complémentaire, visant à montrer l'implication du système moteur spécifiquement labial dans les ruminations en utilisant une tâche d'interférence (suppression articulatoire). Nous avons conçu des tâches motrices impliquant différents effecteurs (main, bouche). Pour vérifier que ces tâches étaient similaires sur le plan attentionnel, nous avons effectué un pré-test comparant leurs influences sur les

réponses lors d'une tâche attentionnelle visuelle. Le pré-test étant satisfaisant, nous avons testé deux groupes de participants. L'étude comportait une procédure d'induction de rumination, suivie d'une tâche motrice (manuelle pour le groupe 1 et orale pour le groupe 2), et d'une auto-évaluation des ruminations. L'étude a été pré-enregistrée sur la plate-forme *Open Science Framework* (<https://osf.io/3bh67/>). Les résultats de cette expérience menée chez 106 participants font l'objet d'une publication soumise à *Collabra: Psychology*. Ils montrent que le degré de rumination rapporté diminue après l'exécution de la tâche motrice, quel que soit l'effecteur, et que la diminution est à peine plus importante dans le cas de la suppression articulatoire (notre hypothèse était que la suppression articulatoire aurait un effet beaucoup plus marqué que l'action de la main). D'autres études sont donc nécessaires pour trancher sur l'hypothèse d'une implication du système moteur spécifiquement orofacial dans les ruminations mentales.

Toujours dans le cadre de la thèse de Ladislav Nalborczyk, nous avons cherché à montrer que l'activation des muscles des lèvres lors des ruminations étaient bien liée à la parole. En effet, les effets obtenus pourraient être dus non pas à la simulation de la production de parole, mais simplement à des contractions labiales liées à la contrariété. Une étude EMG est en cours, dans laquelle nous avons fait produire, à voix haute et mentalement, des non-mots impliquant un étirement des lèvres (hypothèse : muscles zygomatiques principalement recrutés) ou un arrondissement des lèvres (hypothèse : muscles orbiculaires principalement recrutés). Nous nous attendons à une dissociation de l'activation des muscles zygomatiques et orbiculaires en fonction des mots, en parole à voix haute et en parole intérieure. Les données recueillies sont en cours d'analyse.

Les études de la littérature, ainsi que les rapports *a posteriori* des participants de nos propres expériences, indiquent que plusieurs types de ruminations mentales existent, correspondant à différents modes de pensée, notamment les ruminations visuelles et verbales. Nous avons cherché à dissocier les corrélats EMG de ces deux types de rumination. Nous avons mis au point un protocole expérimental permettant d'induire des ruminations visuelles vs. verbales. Nous avons mesuré les activités EMG orofaciales chez deux groupes de participants ayant subi une induction visuelle vs. verbale. Nous avons ensuite appliqué différents types de relaxation (orofaciale vs brachiale) et cherché à différencier les effets en fonction du type de rumination. Nos hypothèses étaient que les participants induits à ruminer verbalement devraient montrer plus d'activité EMG dans les lèvres que dans d'autres muscles faciaux (front) et devraient être plus sensibles à la relaxation orofaciale qu'à la relaxation brachiale. Les résultats préliminaires ne semblent pas confirmer ces hypothèses initiales et des interprétations alternatives sont proposées.

Avec le même objectif de montrer l'implication du système moteur dans les ruminations, une expérience d'inhibition du système moteur par TMS est également en préparation, en collaboration avec Ernst Koster du PANLAB (Université de Ghent, Belgique), dans le cadre de la thèse en co-tutelle de Ladislav Nalborczyk dans ce laboratoire.

Publications associées (articles de revues, chapitres d'ouvrage, congrès, mémoire)

Løevenbruck, H., Nalborczyk, L., Perrone-Bertolotti, M., Baeyens, C., Grandchamp, R. (2017). Un cas d'incarnation linguistique : implication du système moteur orofacial dans les ruminations mentales. *La cognition incarnée*, J. Dokic et D. Perrin (dir.), Recherches sur la philosophie et le langage, N° 33, Vrin, Paris, 297-324, ISBN : 978-2-7116-8423-6. - hal-01887636

Nalborczyk, L., Perrone-Bertolotti, M., Baeyens, C., Grandchamp, R., Spinelli, E., Koster, E.H.W. & Løevenbruck H. (soumis). Articulatory suppression effects on induced rumination. *Psychological Research*.

- Nalborczyk L., Perrone-Bertolotti M., Baeyens C., Grandchamp R., Spinelli E., Koster E. H. W. & Løevenbruck H. (2018). Can articulatory suppression disrupt repetitive negative thinking? *The 6th European Workshop on Imagery and Cognition*, Padua, Italy, 7-9 June 2018.
- Nalborczyk L., Perrone-Bertolotti M., Baeyens C., Grandchamp R., Polosan M., Spinelli E., Koster E. H. W. & Løevenbruck H. (2017). Orofacial electromyographic correlates of induced verbal rumination. *Biological Psychology*, 127, 53-63. hal-01910991
- Nalborczyk L., Perrone-Bertolotti M., Baeyens C., Keracheva, E., Grandchamp R., Spinelli E., Koster E. H. W. & Løevenbruck H. (2017). Verbal rumination as simulated speech. *The 10th annual Embodied and Situated Language Processing Conference*, Moscow, Russia, 10-12 September 2017.
- Nalborczyk, L., Løevenbruck, H., Keracheva, E., Koster, E., Baeyens, C., Grandchamp, R., & Perrone-Bertolotti, M. (2017). Articulatory suppression effects on verbal rumination. *Cognitive Science Arena*, Brixen, Italy, 17-18 February 2017.
- Nalborczyk L., Løevenbruck H., Perrone-Bertolotti M., Baeyens C., Grandchamp R. (2016). Verbal rumination as an embodied orofacial action. *First International Conference on Language and Enaction*, Clermont-Ferrand, France, June 1-3 2016.
- Nalborczyk L., Perrone-Bertolotti M., Baeyens C., Keracheva E., Grandchamp R., Spinelli E., Koster E., Løevenbruck H. (2016). Rumination at the crossroads of embodied inner speech and emotion: facial electromyographic correlates. *Consortium of European Research on Emotion (CERE) conference*, Leiden, The Netherlands, July 6-7 2016, Symposium 4, Paper 1, p.101, oral communication.

Mémoire de M2R : Ladislav Nalborczyk, 2015.

Thèse co-encadrée : Ladislav Nalborczyk, co-tutelle avec Université de Gand, Belgique, en cours.

Financements :

Ce thème a été financé par le projet ANR INNERSPEECH de 4 ans 2014-2018 (<http://lpnc.univ-grenoble-alpes.fr/InnerSpeech>) et a bénéficié d'un financement PEPS Université de Grenoble (2014-15) obtenu pour les mesures physiologiques des ruminations par EMG, en collaboration avec Martial Mermillod (LPNC).

Collaborations :

Grenoble : GIPSA-lab, LPNC, IRMaGe, LIP

Paris : CERILAC

Belgique : Université de Gand

II.3. SUIVI DE L'AGENTIVITE POUR LA FONCTION D'AUTONOËSE

II.3.1. Déficit du suivi de l'agentivité : physiologie des hallucinations auditives verbales

Lorsque le suivi de l'agentivité est perturbé, le contrôle de la parole intérieure peut dysfonctionner, de telle sorte que nos dialogues mentaux sont considérés comme des voix externes. Il a été suggéré que ce trouble de la parole intérieure pourrait être à l'origine de l'hallucination auditive verbale (HAV), un trouble qui fait partie des symptômes de la schizophrénie (cf. les hypothèses théoriques présentées au I.2.3) mais qui peut aussi survenir chez des personnes sans trouble psychiatrique (il affecterait entre 4 et 10% de la population saine, cf. Linden *et al.*, 2010 et <http://revfrance.org>). Dans le cadre de la thèse de Lucile Rapin, avec Marion Dohen et Pascal Perrier, nous avons établi une collaboration entre le GIPSA-lab et le Pôle de Psychiatrie et Neurologie du CHU de Grenoble (Pr. Thierry Bougerol, Dr. Mircea Polosan) ainsi que le Centre de Neurosciences Cognitives à Bron (Pr. N. Franck), portant sur la description physiologique et phénoménologique des HAV chez des patients souffrant de schizophrénie.

Un premier objectif de la thèse de Lucile Rapin portait sur la caractérisation physiologique des HAV. Il s'agissait de recueillir les signaux émis lors des HAV chez des patients schizophrènes à partir des activités motrices orofaciales recueillies par EMG de surface.

Onze patients schizophrènes (diagnostiqués avec le DSM-IV) ont été enregistrés. Tous vivaient des HAV fréquentes lors de l'étude. Les signaux EMG ont été recueillis sur les muscles orbiculaires supérieur et inférieur. Des capteurs ont également été posés sur un muscle du bras, pour contrôle. Les séances ont été filmées pour repérer d'éventuelles mimiques faciales. Les résultats (Figure II-13) montrent une augmentation de l'activité du muscle orbiculaire inférieur durant les HAV par rapport au repos (alors qu'aucun mouvement labial n'est visible) reflétant une implication du système moteur.

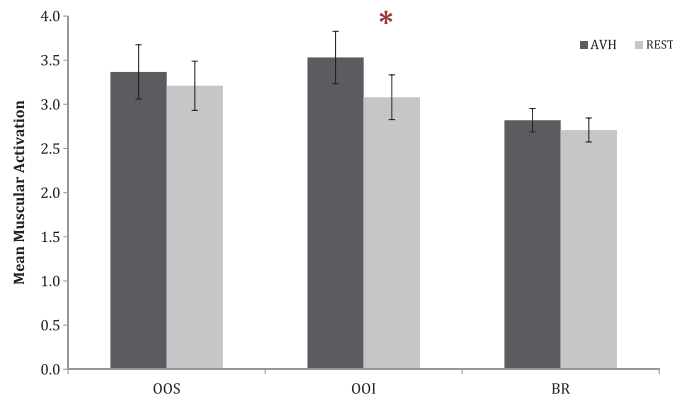


Figure II-13. Corrélats EMG des hallucinations auditives verbales chez des patients souffrant de schizophrénie

Activation EMG des lèvres (OOS : Orbicularis Oris Superior, OOI ; Orbicularis Oris Inferior) et du bras pendant les HAV (barres noires) et au repos (barres grises)

Cette augmentation d'activité musculaire n'est pas liée à une augmentation générale de la tension musculaire, car elle n'est pas retrouvée dans le muscle du bras. Nous émettons l'hypothèse que cette activité musculaire labiale pendant l'hallucination indique que l'hallucination auditive verbale provient bien de parole intérieure. La parole intérieure serait donc mal contrôlée par le patient et serait attribuée à autrui (Rapin *et al.*, 2013).

II.3.2. Déficit du suivi de l'agentivité : neuroimagerie des hallucinations auditives verbales

Un objectif parallèle de la thèse de Lucile Rapin était de mettre en évidence des corrélats cérébraux liés à la présence de HAV via des techniques d'imagerie cérébrale. Lucile Rapin a mené une partie de sa thèse au laboratoire *Cognitive Neuropsychiatry of Schizophrenia* (University British Columbia) à Vancouver dans l'équipe de Todd Woodward. Des données d'IRMf ont été recueillies afin de comparer l'activation cérébrale lors de la génération de pensée verbale (production de parole intérieure) et lors de la perception de parole entre 5 patients hallucinés et 10 sujets sains. Une analyse de connectivité fonctionnelle a été effectuée en utilisant une cPCA (*Constrained Principal Component Analysis*). Des différences entre les groupes de participants (sains et schizophrènes) ont pu être mises en évidence pour deux composantes principales, correspondant à deux réseaux fonctionnels. La première (Figure II-14) reflète une hyper-activité des régions corticales spécialisées dans le traitement de la parole et de la voix (e.g. gyri temporaux supérieurs bilatéraux, STG) lors de la tâche de perception de parole ainsi que lors de la tâche de génération de pensée verbale. La seconde composante correspond à une hyperactivité dans un réseau dit « *task-positive* » qui inclut l'aire de Wernicke, lors de la tâche de génération de pensée verbale. Ces résultats suggèrent une hyperintensité des

réseaux fonctionnels impliquant les régions corticales spécialisées dans le traitement de la voix, qui pourraient contribuer à cet effacement des frontières entre soi et autrui, caractéristique de la schizophrénie (Rapin *et al.*, 2012).

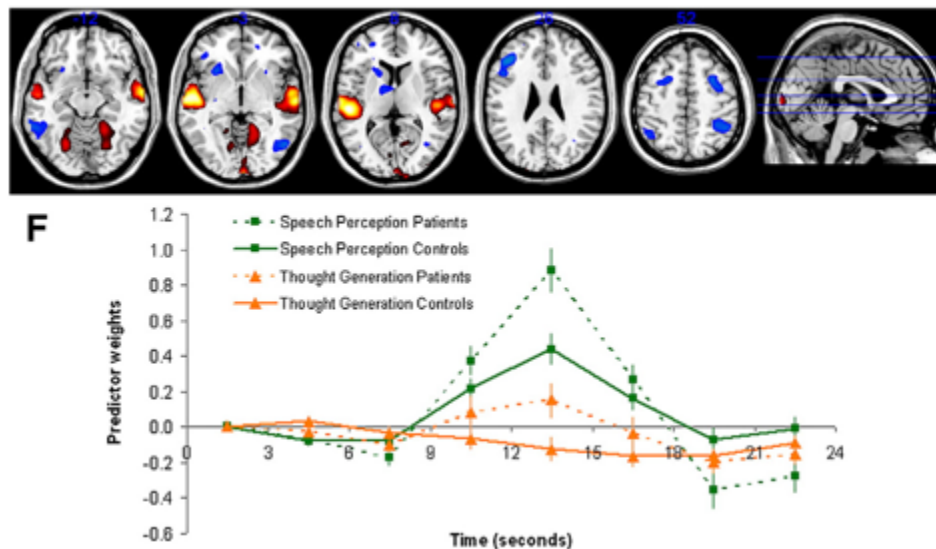


Figure II-14. Réseau impliquant les aires spécialisées dans le traitement de la parole et de la voix et montrant une différence entre patients (hyperactivité) et sujets contrôles dans les deux tâches (Perception de la parole et Génération de pensée verbale)

Ces travaux ont été étendus à un plus grand nombre de participants. 23 patients schizophrènes (10 hallucinés et 13 non-hallucinés), 22 patients bipolaires et 27 participants contrôles ont été examinés, dans les mêmes tâches que précédemment (génération de parole intérieure et perception de parole). Une analyse de connectivité fonctionnelle montre l'implication d'un réseau prédominant à gauche, incluant des régions auditives et motrices liées à la parole. Les patients hallucinés montrent une hyperconnectivité fonctionnelle (observée dans l'étude précédente), mais cette hyperconnectivité n'est significativement supérieure à celle des autres patients psychotiques que dans la tâche de perception (Figure II-15, panneau B). La tâche de perception de parole nécessitant un contrôle exécutif moindre que la tâche de production de parole intérieure, nous émettons l'hypothèse que l'entraînement au contrôle des processus de production de parole intérieure pourrait réduire l'occurrence ou la sévérité des hallucinations (Lavigne *et al.*, 2015). Ces données sont compatibles avec les hypothèses que je pose dans la section I.2.3 sur le dysfonctionnement de l'attribution de l'agentivité et les défaillances des modèles internes d'autrui dans les HAV.

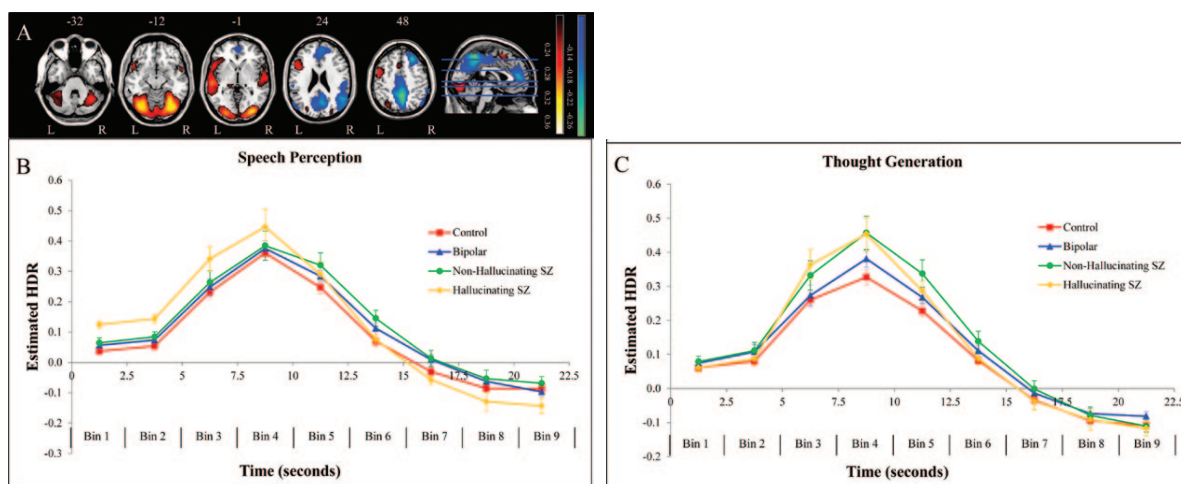


Figure II-15. Réseau impliquant les aires du traitement de la parole (panneau A), montrant une différence entre patients schizophrènes hallucinés et non-hallucinés dans la tâche de perception (panneau B).

Publications associées (articles de revues, chapitres d'ouvrage, mémoire, thèse)

- Lavigne K. M., Rapin L. A. ; Metzak P. D., Whitman J. C., Jung K., Dohen M., Løevenbruck H., Woodward T. S. (2015). Left-dominant temporal-frontal hypercoupling in schizophrenia patients with hallucinations during speech perception. *Schizophrenia Bulletin*, 41 (1), 259-267. doi:10.1093/schbul/sbu004 - PMC4266284 - hal-00966620
- Rapin L., Dohen M. & Løevenbruck H. (2016). Les hallucinations auditives verbales. In Pinto, S. & Sato, M. (eds.). *Traité de Neurolinguistique Neurolinguistique: du cerveau au langage*. Collection Neuropsychologie, De Boeck Supérieur, 347-370. - hal-01442463
- Rapin L., Dohen M., Polosan M., Perrier P., Løevenbruck H. (2013). An EMG study of the lip muscles during covert auditory verbal hallucinations in schizophrenia. *Journal of Speech, Language, and Hearing Research*, 56, S1882–S1893. PMID : 24687444 - hal-00926630
- Rapin, L., Metzak, P. D., Whitman, J. C., Dohen, M., Løevenbruck, H., & Woodward, T. S. (2012). Hyperintensity of functional networks involving voice-selective cortical regions during silent thought in schizophrenia. *Psychiatry Research: Neuroimaging*, 102 (2), 110-117. - hal-00734043, PMID: 22703623

Mémoire de M2R : Lucile Rapin, 2007

Thèse co-encadrée : Lucile Rapin, soutenue en 2011

Financements :

Ce thème a été financé par la Région Rhône-Alpes, Cluster 11 Handicap, Vieillesse, Neurosciences 2009-2011, par un projet PEPS 2014-2015 de l'Université de Grenoble et par une aide du Centre Jacques Cartier, 2009.

Collaborations :

Grenoble : GIPSA-lab, LPNC, CHU

Lyon : CH Bron Vinatier

Canada : CNoS, University British Columbia

II.4. DEVELOPPEMENT DU LANGAGE : MISE EN PLACE DU *MONITORING* VERBAL DE SOI ET D'AUTRUI

II.4.1. Rôle du *monitoring* verbal d'autrui dans le développement phonologique : projet Paidologos

Le développement phonologique met en œuvre, à travers l'imitation, le *monitoring* verbal allocentré. On décrit pourtant souvent le développement phonologique comme reflétant des contraintes universelles « égocentrées », qu'elles soient phonologiques et liées à la marque (e.g. dans le cadre de la Théorie de l'Optimalité, McCarthy, 2007), biomécaniques et liées à la physiologie des appareils de production et perception (e.g. dans le cadre de la théorie *Frame/Content*, Davis *et al.*, 2002) ou idiosyncratiques et liées aux préférences motrices de chaque enfant (dans le cadre de *Templatic Phonology*, Vihman & Croft, 2007).

Parmi les contraintes universelles « égocentrées », la marque imposerait que certains sons soient universellement acquis plus tardivement que d'autres. Ainsi, notamment, il a été proposé que, quelle que soit la langue environnante, la maîtrise des fricatives implique celle des occlusives et des nasales et que la maîtrise des constriction dorsales implique celle des constriction labiales et dentales. Il y aurait donc des étapes dans la maîtrise des consonnes qui semblent être universelles, ce que Jakobson (1941) a nommé la « solidarité irréversible ». L'hypothèse, traduite en phonologie contemporaine, serait donc que les phonèmes marqués sont acquis plus tard. Il est ainsi prédit que les consonnes dorsales comme [k] (marquée [+postérieure]) devraient être acquises plus tardivement que les coronales, telles que [t] (cf. Locke, 1983 ; Jakobson, 1941 ; Yamaguchi, 2008 ; Brandão de Carvalho, Nguyen & Wauquier, 2010).

Une hypothèse quelque peu différente, mais que l'on peut aussi relier à la notion de contrainte universelle, est celle de la contrainte de « cadre ». Elle se fonde sur l'observation de patrons de cooccurrences intrasyllabiques qui seraient universels et reflèteraient des contraintes biomécaniques, émergeant des cycles oscillatoires mandibulaires (Davis *et al.*, 2002). Cette hypothèse prédit en particulier que le cadre antérieur [ti] devrait être maîtrisé avant [tu] et que le cadre postérieur [ku] devrait être maîtrisé avant [ki].

Enfin, selon la *Templatic Phonology* (Vihman & Croft, 2007), chaque enfant produit des sons qui correspondent à un schéma moteur préférentiel (*Vocal Motor Scheme*, VMS), qu'il parvient bien à produire : l'enfant acquiert ainsi un "template", un gabarit propre. Par la suite, pendant la phase des premiers mots, les enfants produisent des mots construits sur ces gabarits. Chaque enfant sélectionne dans le lexique adulte les mots qui correspondent bien à son VMS et qu'il peut ainsi inclure facilement dans son propre lexique. Puis il s'aventure vers des mots nouveaux en commettant des erreurs de production qui reflètent la façon dont il adapte les mots adultes à son gabarit. Selon cette théorie, les schémas moteurs préférentiels sont essentiellement égocentrés, et les productions reflètent la façon dont ce *monitoring* verbal égocentré interagit avec le *monitoring* verbal allocentré, l'enfant sélectionnant dans la langue ambiante les patrons adaptés à ses propres capacités.

J'ai participé, avec Anne Vilain (GIPSA-lab), à un vaste projet interlinguistique (intitulé « Paidologos ») initié par Mary Beckman (Ohio State University) et Jan Edwards (Univ.

Wisconsin), visant à mieux comprendre les interactions entre *les contraintes de marque universelles* et *les influences spécifiques à la langue*, lors de l'acquisition de syllabes chez les enfants de 2 à 5 ans. Les langues étudiées étaient l'anglais américain, le cantonais de Hong-Kong, le mandarin, le coréen, le grec et le japonais. Ces langues diffèrent par la fréquence de certaines consonnes en début de mot. Avec Julia Monnin, doctorante en convention multi-sceaux entre l'INP Grenoble et l'Université de la Nouvelle-Calédonie (Laboratoire Transcultures), nous avons étendu le projet Paidologos au français et au drehu, langue vernaculaire parlée par environ 15000 locuteurs, sur l'île de Lifou, la plus grande et la plus peuplée des îles Loyauté en Nouvelle-Calédonie (Figure II-16). Cette langue kanaque présente en effet l'intérêt de posséder un système consonantique éloigné du français et des répartitions phonologiques différentes.



Figure II-16. à gauche : Langues de Nouvelle-Calédonie ; à droite : Enregistrement d'une enfant à domicile, dans la tâche de répétition de mots illustrés

Nous avons recueilli un corpus de productions audio d'enfants francophones et locuteurs du drehu de 2 ans à 5 ans (Figure II-16, à droite). 158 enfants monolingues du français et 48 enfants bilingues drehu-français ont été enregistrés, en répétition de mots et de non-mots. Huit groupes d'âge de 20 enfants ont été constitués en français, en respectant des écarts de 6 mois entre chaque groupe, de 2 ans jusqu'à 5 ans. En drehu, trois groupes d'âge ont été retenus, allant de 3 à 5 ans, chaque groupe étant constitué de 15 à 20 enfants. Les groupes d'enfants ont chacun répété des mots dans leur langue (en français pour les enfants monolingues, en drehu pour les enfants bilingues) et des non-mots typiques du français. Les mots et non-mots commençaient tous par une des consonnes étudiées dans le projet. Pour permettre une comparaison des deux langues, nous nous sommes focalisées sur des syllabes formées d'une consonne suivie d'une voyelle (CV). Les consonnes étaient choisies parmi /t/, /d/, /k/, /g/, /s/ et /tʃ/ et suivies d'une voyelle parmi /i, a, u/. Nous avons également recueilli des données sur la fréquence des sons étudiés dans les mots du français et du drehu, à partir de notre corpus d'enregistrements d'adultes s'adressant à des enfants de 2 ans. Les résultats préliminaires ont été publiés dans Monnin *et al.* (2015). On observe que les scores de production correcte augmentent avec l'âge mais qu'il existe des différences interlinguistiques, certaines consonnes étant maîtrisées plus tôt en français qu'en drehu et réciproquement. La contrainte de marque phonologique ne peut donc expliquer ces résultats.

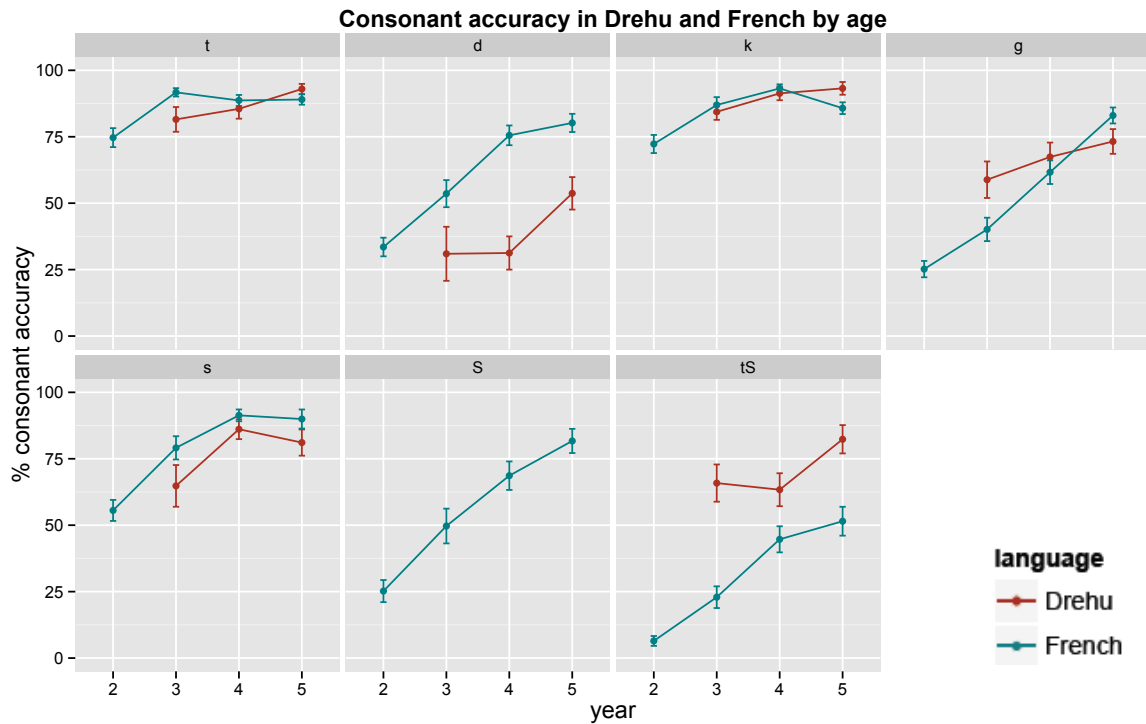


Figure II-17. Scores de production correcte pour les consonnes /t, d, k, g, s, ʃ et tʃ/, en fonction de l'âge, en drehu (rouge) et français (bleu)

Nous avons mesuré les scores de production correcte en fonction du contexte vocalique. La justesse de la production de la consonne dépend de la voyelle qui suit (Figure II-18). Mais contrairement à ce que prédit la contrainte de cadre, les cadres antérieurs ne sont pas mieux produits que les non-cadres ([ti] est moins bien produit que [tu], en français comme en drehu, et [si] est moins bien produit que [su] en drehu). De même les cadres postérieurs ne sont pas systématiquement mieux produits ([gu] n'est pas mieux produit que [ga] en français). Et des différences en fonction de la langue ambiante sont notées.

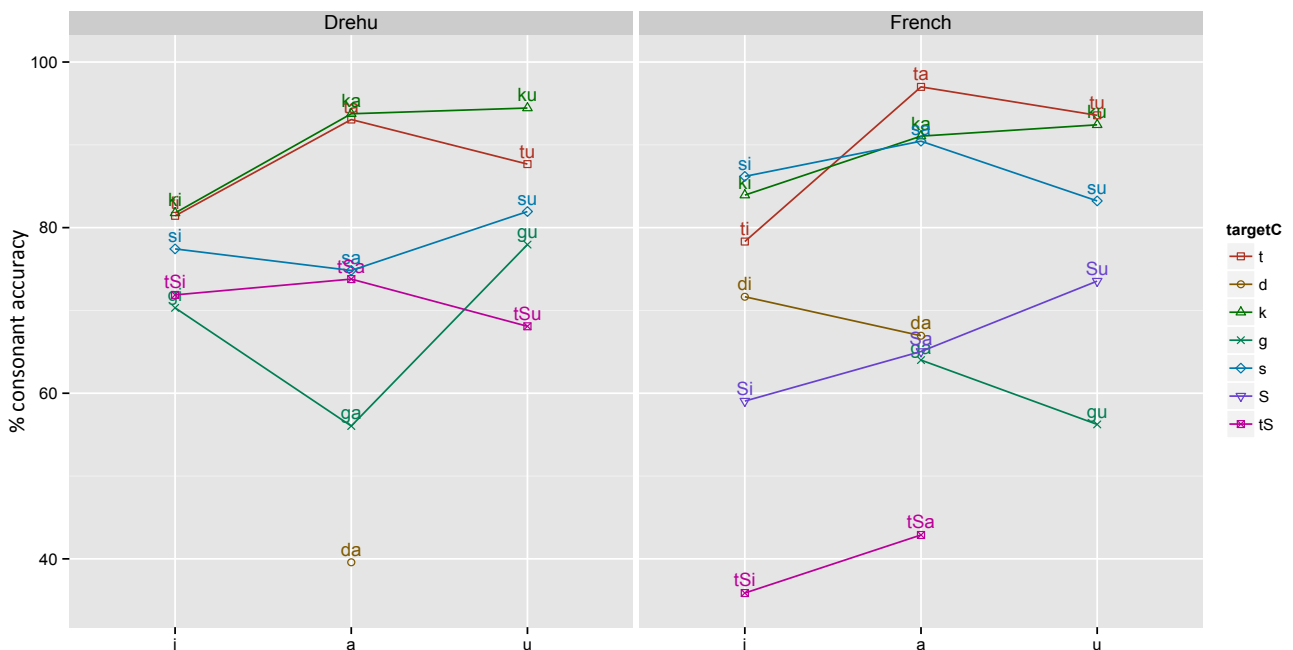


Figure II-18. Scores de production correcte pour les consonnes /t, d, k, g, s, ʃ et tʃ/, en fonction de la voyelle, en drehu (à gauche) et français (à droite)

Nous avons donc relié les scores de production correcte des syllabes CV d'intérêt à la fréquence de ces syllabes dans la langue ambiante. Les résultats (Figure II-19) indiquent que la fréquence de la syllabe CV contribue à la justesse de la production de la consonne, quelle que soit la langue. Ces travaux indiquent donc que les productions des jeunes enfants reflètent une modulation des contraintes de marque universelles (ou du *monitoring* verbal égocentré) par des spécificités phonotactiques propres à la langue ambiante (*monitoring* verbal allocentré).

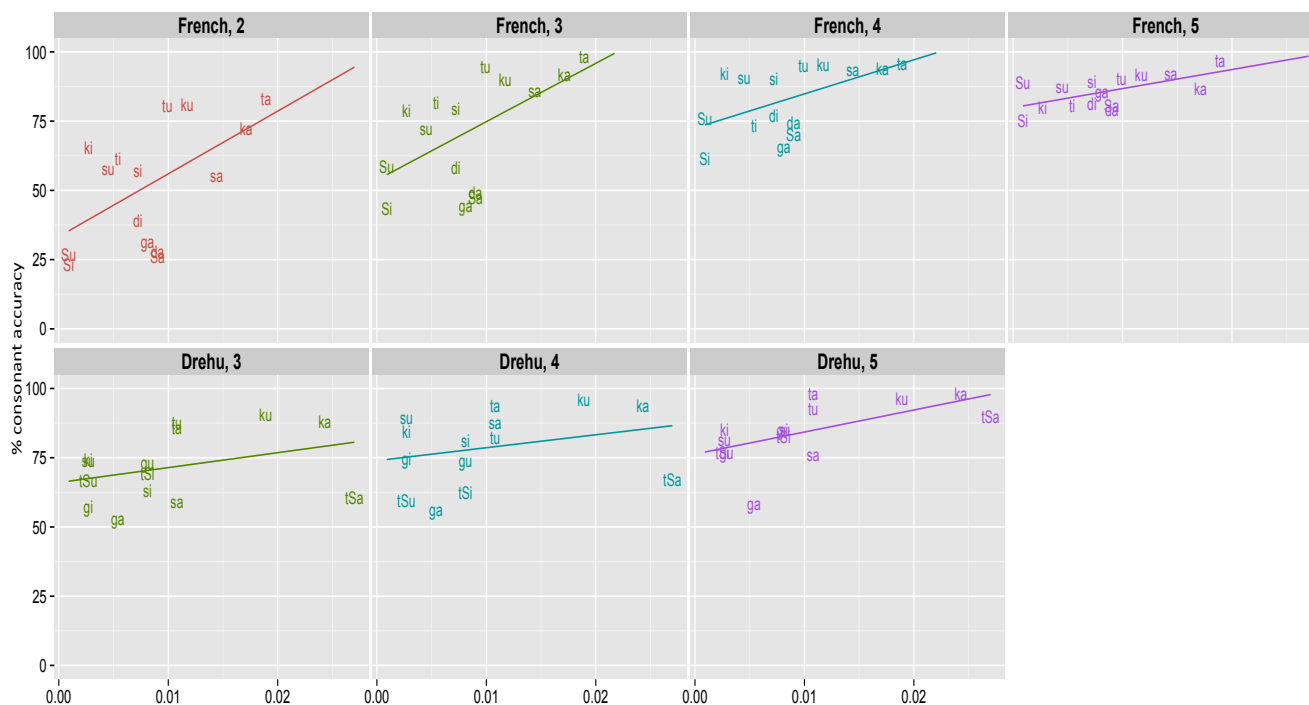


Figure II-19. Scores de production correcte pour les consonnes /t, d, k, g, s, f, tʃ/ en fonction de la fréquence de la syllabe CV dans la langue, en français (en haut) pour les tranches d'âge de 2 à 5 ans et drehu (en bas) pour les tranches d'âge de 3 à 5 ans

Cet axe de recherche a des retombées fondamentales originales dans le domaine du développement phonologique. Tout d'abord, il fournit les premières données chronologiques sur les étapes du développement phonologique dans une langue kanaque. Ensuite, si les travaux sur la perception précoce des langues commencent à fournir des résultats intéressants, la production précoce des enfants est encore peu explorée. Nous envisageons à l'avenir d'étendre ce projet à d'autres langues, notamment au portugais du Brésil, dans le cadre de la collaboration initié par Rafael Laboissière du LPNC, avec l'Universidade Federal de Parana au Brésil (projet CMIRA-COOPERA).

Ces travaux, comparant plusieurs langues de familles variées, permettent de mieux cerner les liens entre capacités de production et influence de la langue ambiante au cours du développement du langage, i.e. entre *monitorings* verbaux égo- et allo-centrés.

Publications associées (articles de congrès avec actes, mémoire, thèse)

Monnin, J., Løevenbruck, H., & Beckman, M. E. (2007). The influence of frequency on word-initial obstruent acquisition in Hexagonal French. Proceedings of the XVth International Congress of Phonetic Sciences, 6-10 August 2007, Saarbruecken, 1569-1572. - hal-00275720

Monnin, J., Løevenbruck, H. & Wacalie, F. (2007). Données fréquentielles du français et du drehu dans le langage adressé au jeune enfant. Proceedings of the Seventh international Conference on Oceanic Linguistics, July 2 – 6 2007, Noumea, New-Caledonia, 23. - hal-00371960

- Monnin, J., Løevenbruck, H. (2008). Influence des fréquences lexicales des langues française et drehu sur l'acquisition des consonnes initiales de mots. *Actes des XXVIIèmes Journées d'Etude sur la Parole (JEP'08)*, Avignon, 9 – 13 juin. - hal-00346917
- Monnin J. & Løevenbruck H. (2010). Spécificités de l'acquisition des consonnes en français et en drehu : influence de la langue ambiante. *Actes des XXVIIIèmes Journées d'Etude sur la Parole (JEP'10)*, Mons, 25 – 28 mai, 29-32. - hal-00557748
- Monnin J. & Løevenbruck H. (2010). Language-Specific Influence on Phoneme Development: French and Drehu Data. *Proceedings of Interspeech 2010*, Makuhari, Japan, 26-30 September 2010, 1882-1885. - hal-00557749
- Monnin J., Løevenbruck H., Beckman M. E. & Edwards J. (2011). Vowel context and frequency effects in dorsal and coronal acquisition in Drehu and French. *Proceedings of the 9th International Seminar on Speech Production (ISSP 2011) Abstracts*, 91, June 20-23 2011, Montreal, Canada. - hal-00613845
- Monnin J., Løevenbruck H., Beckman M. E. & Edwards J. (2012). Dorsal and coronal acquisition in Drehu and French is unexpected by markedness constraints alone. *International Child Phonology Conference*, June 4-6 2012, Minneapolis, USA. - hal-00781641
- Monnin J., Løevenbruck H. & Vilain A. (2015). Consonant and syllable frequency effects in stop and fricative acquisition in preschool Drehu- and French-acquiring children. *Proceedings of the 18th International Congress of Phonetic Sciences (ICPhS 2015)*, Glasgow, UK, 10-14 August 2015, Glasgow, UK: the University of Glasgow. ISBN 978-0-85261-941-4. Paper number 665. hal-01250518

Mémoire de M2R : Julia Monnin, 2005

Thèse co-encadrée : Julia Monnin, convention multi-sceaux avec l'Université de la Nouvelle-Calédonie, soutenue en 2010

Financements :

Ce thème a été financé par le projet « Paidologos : cross-language investigation of phonological development » (<http://www.ling.ohio-state.edu/~edwards/>), porté par M. Beckman et J. Edwards.

Collaborations :

Grenoble : GIPSA-lab, LPNC
Nouméa : Université de la Nouvelle-Calédonie
USA : Ohio State University

II.4.2. Rôle du *monitoring* verbal égo-centré dans la perception précoce

Le développement du langage fait émerger des capacités de catégorisation des sons produits par autrui, qui sont elles-mêmes façonnées par nos propres capacités de production. Ainsi le *monitoring* verbal égo-centré joue un rôle sur le *monitoring* verbal allocentré. Je participe au projet ERC Speech Unit(e)s de Jean-Luc Schwartz (GIPSA-lab), impliquant également Anne Vilain (GIPSA-lab) et Olivier Pascalis (LPNC), et dont l'un des axes porte sur l'influence fonctionnelle des compétences motrices du nourrisson dans le développement du langage. Dans une étude menée au Babylab du LPNC, nous avons exploré l'hypothèse que les propres capacités de production de parole du nourrisson pourraient l'aider à découvrir les catégories phonétiques de sa langue, malgré les effets de la coarticulation qui introduisent de la variabilité dans les sons (le signal acoustique d'un [t] diffère en fonction de la voyelle qui suit, pourtant en tant qu'adulte, nous percevons le son /t/ dans tous les cas). Nous avons examiné l'influence du babillage sur la catégorisation des consonnes chez des nourrissons de 6 et 9 mois. En nous appuyant sur une procédure d'association intersensorielle, nous avons examiné la capacité des nourrissons à faire l'association entre les signaux *acoustiques* correspondant à une consonne donnée, produite dans des contextes vocaliques variés, et les signaux *visuels* correspondant à cette même consonne. En parallèle, nous avons examiné les répertoires

de production vocales de ces nourrissons, à partir de questionnaires remplis par leurs parents. Dans une première expérience, sur le contraste /b-/d/, nous avons pu montrer que seuls les nourrissons capables de produire une des deux consonnes (/b/ ou /d/) dans leur babillage, quel que soit leur âge, étaient capables de faire l'association entre les signaux intersensoriels d'une même consonne. Ceci suggère que les capacités de production induisent des habiletés de perception et notamment de catégorisation phonétique, telles que celles observées chez l'adulte. Pour vérifier que ce résultat n'était pas uniquement lié à des compétences linguistiques d'ordre général chez les enfants capables de babillage, nous avons mené une deuxième expérience, utilisant la paire /v-/z/, qui n'est généralement pas produite à ces âges. Avec cette paire de sons, nous n'avons pas trouvé de capacité à faire l'association, quel que soit l'âge ou la capacité de babillage (celle-ci étant mesurée sur les consonnes /b/ et /d/). Toutefois, de façon intéressante, les rares nourrissons capables de produire soit /v/ soit /z/ étaient justement capables d'associer les stimuli auditifs et visuels.

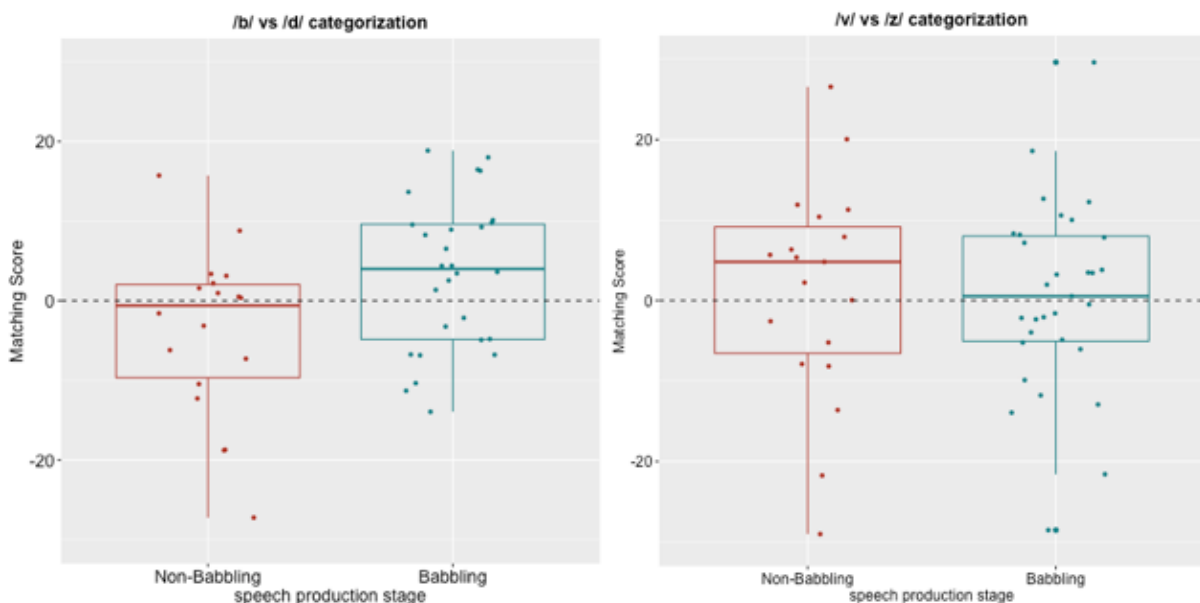


Figure II-20. Scores d'association audiovisuelle correcte pour /b, d/ à gauche et /v, z/ à droite, en fonction des capacités de babillage des nourrissons de 6-9 mois

Ces résultats font l'objet d'une publication dans *Developmental Science* (Vilain *et al.*, sous presse). Ils suggèrent que le babillage joue un rôle dans l'extraction des représentations du lieu d'articulation des occlusives, en d'autres termes que le *monitoring* verbal égo-centré intervient très précocement dans le *monitoring* verbal allocentré.

Publications associées (articles de revues, congrès avec actes, site web)

- Vilain, A., Dole, M., Løevenbruck, H., Pascalis, O., & Schwartz J.-L. (sous presse). The role of production abilities in the perception of consonant category in infants. *Developmental Science*.
- Dole, M., Løevenbruck, H., Pascalis, O., Schwartz, J.L, Vilain, A. (2016). Phoneme categorization depends on production abilities during the first year of life. *International Conference on Infant Studies (ICIS)*, New Orleans, Louisiana, USA, May 26-28 2016, Poster 1095528.
- Dole, M., Løevenbruck, H., Pascalis, O., Schwartz, J.L, Vilain, A. (2015). Perceptual abilities in relation with motor development in the first year of life. *Workshop on Infant Language Development (WILD) 2015*, Stockholm, Sweden, 10-12 June 2015. hal-01251120

Site web :

Babylab du LPNC : <http://babylab-grenoble.fr>

Financement :

Ce thème a bénéficié du financement du projet ERC Speech Unit(e)s –2014-2019 porté par J.L. Schwartz.

Collaborations :

Grenoble : GIPSA-lab, LPNC

II.4.3. Monitoring verbal de soi pour autrui : pointage chez l'enfant

Comme mentionné ci-dessus (cf. II.1.1), il a été montré que la capacité de pointage, d'abord avec l'œil, puis avec le doigt et enfin avec l'intonation et la syntaxe, joue un rôle crucial dans le développement de la communication chez le nourrisson et l'enfant. Je me suis penchée en particulier sur les pointages prosodique et manuel au cours du développement, chez l'enfant francophone.

II.4.3.1. Corrélats acoustico-articulatoires du pointage prosodique chez l'enfant

La focalisation prosodique est un phénomène intéressant car elle implique des contrôles d'effecteurs maîtrisés à des stades différents. Par exemple, les nourrissons ont un contrôle relativement précis de la durée, de la F_0 et de l'intensité (aspects laryngés), très précocement alors que la dynamique des articulateurs supralaryngés n'est maîtrisée que tardivement dans l'enfance (Vihman, 1996). J'ai participé à un projet dirigé par Lucie Ménard de l'Université du Québec à Montréal. Un corpus audio-visuel a été enregistré pour 10 locuteurs (trois enfants de 4 ans, trois enfants de 8 ans, et quatre adultes). Des répétitions de la séquence [baba] ont été produites dans deux conditions prosodiques : assertion neutre et focalisation contrastive. La séquence [baba] était insérée dans deux phrases : « J'ai vu Baba qui mange un gâteau » pour la condition neutre et « Non, c'est BABA qui mange un gâteau » pour la condition de focalisation. Les résultats indiquent que pour les gestes d'ouverture ([b] vers [a]) et de fermeture ([a] vers [b]), bien que le focus sur [baba] corresponde à une augmentation des durées articulatoires et acoustiques pour tous les locuteurs, la dynamique articulatoire n'est pas la même pour les enfants que les adultes. De plus, pour les enfants de 4 ans, les valeurs du formant F1 varient moins en fonction de la condition que pour les adultes. Les enfants seraient capables d'hyper-articuler les constituants focalisés mais auraient plus de difficultés à hypo-articuler les autres, rendant le contraste focalisé/non-focalisé moins saillant.

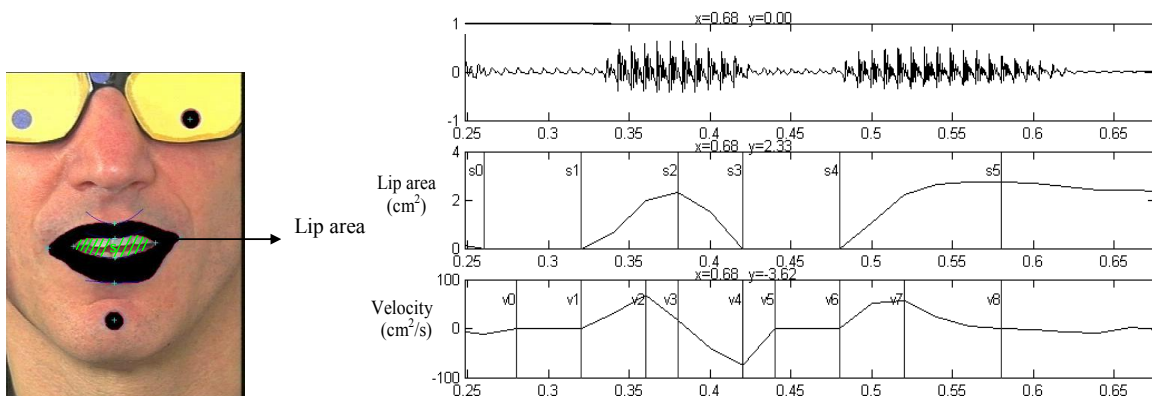


Figure II-21. Données articulatoires extraites de données vidéo

A gauche : mesure de l'aire interlabiale à partir de données vidéo ; A droite : événements articulatoires repérés sur l'aire interlabiale pour la séquence [baba]

En ce qui concerne l'intonation, le focus contrastif est marqué pour tous les locuteurs par un pic de F_0 sur le constituant focalisé et un plateau de F_0 peu élevée pour les constituants pré- et post-focaux. En résumé, ces données suggèrent que la modalité articulatoire de la focalisation prosodique est maîtrisée relativement tardivement, alors que la modalité acoustique semble déjà relativement bien acquise (Ménard *et al.*, 2006). Ces résultats indiquent qu'à l'âge de 8 ans, la capacité de véhiculer la focalisation prosodique, i.e. le *monitoring* verbal égocentré dédié à autrui, mis en œuvre pour faciliter à autrui la compréhension, n'a pas encore atteint les caractéristiques de celle des adultes.

Publications associées (articles de revue, congrès avec actes)

- Ménard L., Løevenbruck H., Savariaux C. & Cedergren H. (2003). Articulatory and acoustic correlates of contrastive emphasis in French: a developmental study. *Sixth Seminar on Speech Production*, Sydney, 8-10 Décembre 2003.
- Ménard L., Løevenbruck H. & Savariaux C. (2005). A study of articulatory and acoustic correlates of contrastive emphasis in French children and adults. *Actes du colloque "un siècle de phonétique expérimentale : histoire et développement. De Théodore Rosset à John Ohala."*, Grenoble, France, 24-25 février 2005, 46, [Résumé]. - hal-00371965
- Ménard L., Løevenbruck H. & Savariaux C. (2006). Articulatory and acoustic correlates of contrastive focus in French: a developmental study, in Harrington, J. & Tabain, M. (eds), *Speech Production: Models, Phonetic Processes and Techniques*, Psychology Press: New York, 227-251. - hal-00371878.

Financement :

Ce thème a bénéficié d'un financement du centre Jacques Cartier 2002.

Collaboration :

Montréal : Laboratoire de Phonétique de l'UQAM

II.4.3.2. Corrélats acoustico-corporels du pointage prosodique chez l'enfant

Je participe actuellement à un projet initié par Nuria Esteve-Gibert et Mariapaola D'Imperio au Laboratoire Parole et Langage, à Aix-en-Provence, en collaboration avec Marion Dohen, GIPSA-lab. Il existe peu de travaux sur le développement de l'intégration des marqueurs prosodiques et corporels chez l'enfant. Les hochements de tête font partie des marqueurs pragmatiques utilisés chez l'adulte pour rendre plus saillante la structure informationnelle du message et faciliter la compréhension chez autrui (Ambranzaitis & House, 2017). Toutefois, la plupart des travaux sur les marqueurs pragmatiques de la structure informationnelle au cours du développement se sont centrés sur la modalité prosodique acoustique et il existe très peu d'études sur les gestes corporels. Nous avons enregistré et filmé 24 enfants de 5 ans participant à un jeu animé sur ordinateur, utilisant des objets de différentes couleurs. Ce jeu permettait d'éliciter des productions orales semi-spontanées dans trois conditions prosodiques : focalisation large (neutre), focalisation étroite contrastive, focalisation étroite corrective. La focalisation portait sur le nom ou l'adjectif. Les données audio ont été analysées sous le logiciel *Praat* et les données vidéo sous *ELAN* (Figure II-22). Les mesures acoustiques de durée et d'étendue de F_0 sur les syllabes et les mots montrent que la condition de focalisation n'a pas d'effet sur les productions orales spontanées des enfants. Toutefois, les mesures des mouvements de la tête indiquent un effet de la condition de focalisation. Les gestes de hochement de la tête sont alignés avec les constituants focalisés. De plus, un taux plus important de hochements a été observé dans les conditions les plus focalisées, dans cet ordre : focalisation corrective > contrastive > large.

Ces résultats suggèrent que le *monitoring* verbal égocentré (de nos propres productions) mais orienté vers autrui, i.e. visant à faciliter à autrui la compréhension, est tardif et semble se mettre en place d'abord par l'utilisation de gestes corporels, notamment par les hochements de la tête (Esteve-Gibert et al, 2018 ; Esteve-Gibert *et al.*, ICPhS 2019).

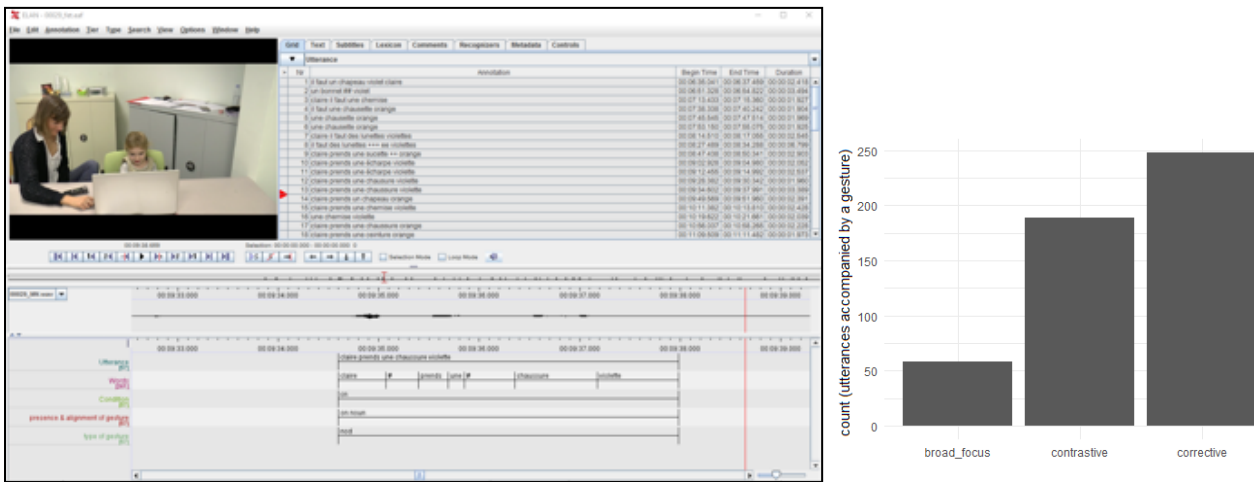


Figure II-22. Analyse des données vidéo des enfants enregistrés au cours d'un jeu élicitant une focalisation prosodique semi-spontanée

A gauche : Annotation des gestes corporels (mains, corps, tête, sourcils) et de leur alignement avec le signal audio ; à droite : Nombre d'énoncés accompagnés d'un geste de la tête dans les 3 conditions prosodiques (focalisation large, étroite contrastive et étroite corrective)

Publications associées (articles de congrès avec actes)

- Esteve-Gibert N., Løevenbruck H., Dohen M. and D'Imperio M. (accepté). Pre-schoolers use head gestures rather than duration or pitch range to signal narrow focus in French. *19th International Congress of Phonetic Sciences (ICPhS 2019)*.
- Esteve-Gibert, N., Løevenbruck, H., Dohen, M. & D'Imperio M. (2018). French pre-schoolers use head nods more than prosodic cues as a marker of information structure. *Laboratory Phonology 16 Conference*, Lisbon, Portugal - 22 June 2018.
- Esteve-Gibert, N., Legou, T., Løevenbruck, H., Dohen, M. & D'Imperio M. (2018). Head movements highlight important information in speech: an EMA study with French speakers. *XIVth AISV conference*, Bozen-Bolzano, Italy, 25-27 January 2018.
- Esteve-Gibert N., Løevenbruck H., Dohen M. and D'Imperio M. (2017). Audiovisual correlates of focus production in French-speaking 4 and 5 year old. *14th International Congress for the Study of Child Language (IASCL)*, Lyon, 17-21 July 2017.
- Esteve-Gibert N., Løevenbruck H., Dohen M., Legou T. and D'Imperio M. (2017). The development of prosodic and gesture cues to focus in French pre-schoolers. *Phonetics and Phonology in Europe Conference (PaPE)*. Cologne, Germany, 12-14 June 2017.
- Esteve-Gibert N., Løevenbruck H., Dohen M. and D'Imperio M. (2017). The use of prosody and gestures for the production of contrastive focus in French-speaking 4 and 5 year olds. *Workshop on Abstraction, Diversity and Speech Dynamics*, Herrsching, Germany, May 3-5, 2017.
- Esteve-Gibert N., Løevenbruck H., Dohen M., Legou T. and D'Imperio M. (2017). Adults' and children's use of prosody and head gestures to mark contrastive focus in French. *iGesto'17, International Conference on Gesture and Multimodality*, Porto, Portugal, 02-03 February 2017.
- Esteve-Gibert N., Løevenbruck H., Dohen M., Legou T. and D'Imperio M. (2017). Adults' and children's use of prosody and head gestures to mark contrastive focus in French. *XIIIth AISV conference*, Pisa Italy, 25-27 January 2017.

Collaborations :

Aix-en-Provence : Laboratoire Parole et Langage
Barcelone : Univ. Catalunya

II.4.3.3. Production de pointage et de gestes manuels chez les enfants présentant des déficits langagiers

J'ai participé à une étude menée par Marion Dohen (GIPSA-lab) sur la production de gestes manuels communicatifs, notamment de pointage, chez les enfants présentant des déficits langagiers. On sait que les gestes manuels, et particulièrement le pointage, jouent un rôle dans l'acquisition du langage chez les enfants. On peut ainsi se poser la question de ce qu'il en est chez les enfants présentant des troubles spécifiques du développement du langage oral (TSLO ou dysphasie). Produisent-ils plus de gestes manuels que les autres enfants ? Cette étude, menée en commun avec le Centre Référent des Troubles du Langage et des Apprentissages scolaires du CHU de Grenoble (CRTLA), visait à examiner si les enfants utilisent spontanément les gestes pour compenser leur déficit et comment (nature des gestes produits, combinaison et coordination avec la parole). Nous avons recueilli un corpus de données audio et vidéo de narration spontanée chez des enfants tout-venant et présentant des TSLO. Nos résultats indiquent que les enfants présentant des troubles du langage importants utilisent plus de gestes que les enfants tout-venant, et en particulier plus de gestes interactifs et discursifs et plus de gestes se substituant à la parole. Des analyses plus approfondies sont en cours. A long terme, ce projet permettra de mettre en place des thérapies compensatoires adaptées, utilisant la modalité gestuelle. Dans une autre étude préliminaire, menée par Marion Dohen (GIPSA-lab), et toujours en collaboration avec le CRTLA du CHU de Grenoble, nous avons examiné si les enfants retiennent mieux à court terme des pseudo-mots lorsque ceux-ci sont présentés à l'oral et accompagnés de gestes représentationnels. Les résultats dans des tâches d'identification et de dénomination chez 10 enfants présentant TSLO suggèrent que lorsqu'un mot est associé à un geste représentationnel durant l'apprentissage, il est mieux appris : les scores d'identification et de dénomination sont plus élevés que lorsque le mot est appris sans geste (Dohen *et al.*, 2016).

Publication associée (article de congrès avec actes)

Dohen M., Vilain A., Løevenbruck H., Rochet-Cappellan A. & Gillet-Perret E. (2016). The Role of Manual Representational Gestures in Novel Word Acquisition in Children with Specific Language Impairments. 7th Conference of the International Society for Gesture Studies (ISGS7). Paris, France, 18-22 July 2016, p. 166, Oral communication.

Financements :

Ce thème a bénéficié de financements de la SFR Santé Société Grenoble, sur l'étude de la compensation potentielle des déficits de langage par l'utilisation de gestes manuels communicatifs chez les enfants, projets portés par Marion Dohen, 2011-12, 2012-13.

Collaborations :

Grenoble : GIPSA-lab, LPNC, CRTLA du CHU

II.4.4. Développement phonologique typique et atypique : *monitoring verbal de soi complexifié*

Les Troubles du Développement des Sons de Parole (TDSP, en anglais *speech sound disorders*) correspondent à un délai d'acquisition des sons de la langue. Ces troubles sont associés à des difficultés de perception, de production de la parole et/ou de représentation phonologique, qui affectent l'intelligibilité et l'acceptabilité, et causent des retards de langage, et des difficultés scolaires et sociales. Cependant, les causes sous-

jaçentes à ces troubles restent mal comprises, et les marqueurs cliniques mal définis. En français, il n'existe pas d'outil standardisé pour l'évaluation des TDSP. On constate également un manque de données de référence sur le développement phonologique des enfants francophones. Ce manque d'outils et de données est en contradiction avec la pratique fondée sur des données probantes recommandée aux cliniciens. Dans ce contexte, je co-dirige avec Anne Vilain de GIPSA-lab, en collaboration avec Estelle Gillet-Perret du CRTLA du CHUGA, le projet EULALIES qui rassemble chercheur·se·s et clinicien·ne·s dans le but de produire (1) un outil standardisé pour l'évaluation des TDSP en français, et (2) un ensemble de données normatives sur le développement phonologique dans deux variétés linguistiques : en France et au Québec. Ce projet est mené en partenariat avec Sophie Kern et Mélanie Canault du Laboratoire Dynamique du Langage à Lyon et Andrea MacLeod de l'Université de Montréal au Canada. Nous sommes également en contact avec Yvan Rose de *Memorial University*, Canada et Christelle Maillart de l'Université de Liège. La thèse de Geneviève Meloni se déroule dans le cadre de ce projet, en co-tutelle avec l'Université Grenoble Alpes et l'Université de Montréal. Ses premiers travaux de master et de thèse nous ont permis de développer une batterie de tests établie à partir du modèle psycholinguistique de traitement de parole de Stackhouse et Wells (1997), dans le cadre phonologique de Munson *et al.* (2005). Cette batterie est construite de façon à faciliter l'analyse linguistique des productions enfantines et l'établissement de profils psycholinguistiques. Elle contient cinq tâches.

La tâche de jugement de lexicalité teste les représentations phonologiques perceptives. Il est demandé à l'enfant de juger si les mots prononcés sont corrects ou non. L'épreuve comprend 45 paires de mots/pseudomots. Les pseudomots présentent des altérations au niveau de la structure ('livre' vs. 'lirve') et des segments ('pantalon' vs. 'pankalon').

La tâche de dénomination vise les représentations phonologiques motrices (programmes moteurs). L'enfant doit nommer les images affichées sur un écran d'ordinateur. Cette épreuve comprend des mots de longueurs syllabiques différentes (une à cinq syllabes) avec une représentation segmentale exhaustive des phonèmes du français hexagonal dans différentes position dans les mots.

La tâche de répétition de pseudomots cible la programmation motrice. On demande à l'enfant de répéter des mots qui n'existent pas. Ces pseudomots sont construits avec des variations de longueurs, de complexité syllabique et de vraisemblance (présence de morphème lexical ou grammatical).

La tâche de diadococinésies teste la planification motrice. Elle consiste à répéter rapidement les séquences de syllabes /pa-ta-ka/, /pa-ba/ et /pa/, chacune pendant 10 s.

La tâche de répétition de syllabes juge de l'exécution motrice. L'enfant doit répéter les syllabes simples et complexes présentées à l'audio.

La collecte des données chez des enfants typiques et présentant des TDSP est en cours. Un premier recueil a été effectué par Valérie Schott-Brua, étudiante en orthophonie en 2017-18 auprès de 27 enfants tout-venant et 8 enfants avec TDSP, dont certains présentaient un trouble phonologique et d'autres une dyspraxie verbale (trouble plus sévère). Les résultats montrent que la batterie de tests est opérante pour dresser des profils langagiers, permettant à terme d'envisager des pistes de rééducation (Meloni *et al.*, 2017, 2018 ; Schott-Brua, *et al.*, 2018). Le type de représentation en radar proposé (cf. Figure II-23) permet de décrire les scores par groupes d'enfants ou par enfant, pour mettre en avant les différents profils de TDSP ou les spécificités de chaque enfant.

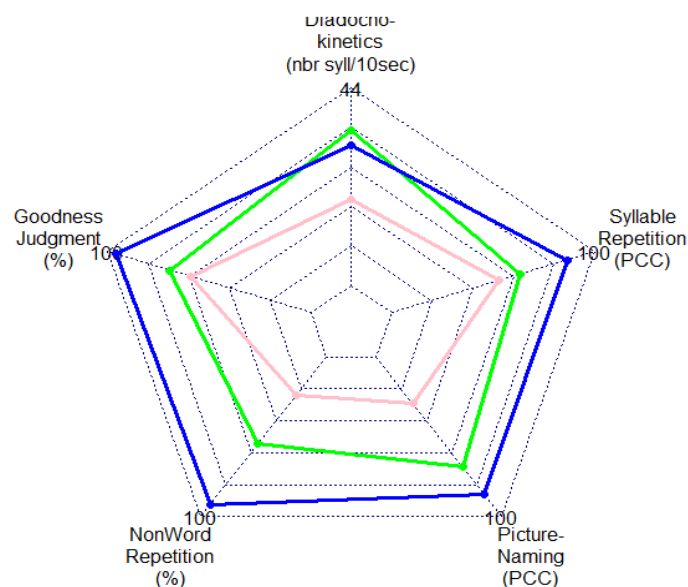


Figure II-23. Analyse en radar des scores des enfants à la batterie de tests EULALIES, sur un petit nombre d'enfants tests.

En Bleu, les enfants typiques, en vert, les enfants présentant un trouble phonologique, en rouge, les enfants présentant une dyspraxie verbale.

De plus, nous avons montré (Meloni, 2018, Meloni *et al.*, 2018) qu'à partir de ces profils radars individuels, une analyse phonologique multilinéaire, inspirée des travaux de Bernhardt & Stoel-Gammon (1994), Bernhardt *et al.* (2015) et Brosseau-Lapré *et al.* (2018), permet de pointer spécifiquement sur les difficultés individuelles de chaque enfant (telles que nasalité, voisement, groupes consonantiques), ce qui devrait permettre à terme de proposer des pistes de rééducation ciblée. Nous espérons, grâce à ce projet, contribuer à mieux décrire le développement typique et atypique du *monitoring* verbal égocentré, notamment au niveau phonologique, chez les enfants de 3 à 11 ans.

Publications associées (articles de revues, chapitres d'ouvrage, congrès avec actes, mémoire, thèse)

- Meloni, Geneviève, Løevenbruck, Hélène, Vilain, Anne & MacLeod, Andrea (2018). Integrating production and perception tasks in a psycholinguistic approach to assessment of Speech Sound Disorder in French: a case study. *17th International Clinical Phonetics and Linguistics Association conference*, Malta, 23-25 October 2018. hal-01950372
- Meloni, G., Vilain, A., Løevenbruck, H. & MacLeod A. (2017). The France-Canada Speech Sound Disorders Project. *14th International Congress for the Study of Child Language (IASCL)*, Lyon, 17-21 July 2017.
- Meloni, G., Løevenbruck, H. & Vilain, A. (2017). Evaluation de la perception des sons de parole chez les populations pédiatriques : réflexions sur les épreuves existantes. *7èmes Journées de Phonétique Clinique*, Paris, 29-30 juin 2017.
- Meloni G., Vilain A., Løevenbruck H., Maillart C., MacLeod A. (2016). The France-Belgium-Canada SSD Project. *International Clinical Phonetics and Linguistics Association Conference (ICPLA) 2016*, Halifax, Canada, June 15-18 2016, Poster 14.
- Meloni, G., Vilain, A., Løevenbruck, H., MacLeod, A. (2015). Projet TDSP France-Québec : évaluation et dépistage des troubles du développement des sons de parole. *Colloque international Atylang, Atypies langagières : mais de quoi parle-t-on vraiment ?* Paris, France, 27-28 November 2015. hal-01251090
- Schott-Brua V., Meloni G., Løevenbruck H., Vilain A., MacLeod A.A.N. (2018). Differential diagnostic of childhood apraxia of speech: a study on French-speaking children. *2018 American Speech-Language-Hearing Association (ASHA) Convention*, Boston, USA, 15-17 November 2018. hal-02044352

Mémoire de M2R : Geneviève Meloni, 2015

Thèse co-encadrée : Geneviève Meloni, co-tutelle avec l'Université de Montréal, en cours.

Site web : <http://www.gipsa-lab.fr/projet/EULALIES/accueil.html>

Financements :

Ce thème bénéficie d'un financement ARC2 de la région AURA 2016-19 (Évaluation et dépistage des troubles du développement des sons de parole) porté par moi-même, ainsi que de financements de la SFR Santé Société (2015 et 2018) et du CNRS-INS2I PICS (projet international de coopération scientifique, porté par Anne Vilain 2018-20).

Collaborations :

Grenoble : GIPSA-lab, LPNC, CRTLA du CHU

Lyon : Dynamique Du Langage

Belgique: Université de Liège

Canada : Université de Montréal, Memorial University Newfoundland

II.4.5. Production de parole chez les enfants implantés cochléaires : le monitoring verbal de soi et d'autrui altéré

Les difficultés de langage des enfants sourds pré-linguaux porteurs d'implant cochléaire (IC) sont peu étudiées. Les rares études disponibles révèlent l'existence de difficultés de production et de perception de certains sons de parole, même après plusieurs années d'implantation. Or les troubles de la parole peuvent induire des difficultés dans les apprentissages scolaires et avoir un retentissement affectif et social pour l'enfant.

J'ai participé à un projet mené par Anne Vilain du GIPSA-lab portant sur l'évaluation des difficultés de production de parole persistantes chez les enfants sourds pré-linguaux, après plusieurs années d'implantation cochléaire, avec pour but, à long terme, de proposer des stratégies de rééducation adaptées.

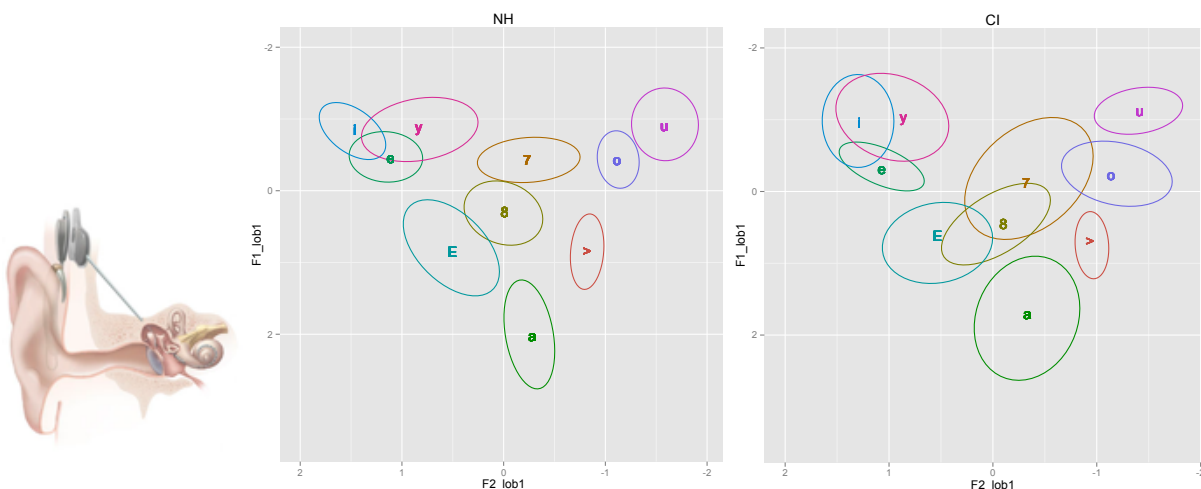


Figure II-24. A gauche : Schéma d'un implant cochléaire (IC) ; au centre : espaces vocaliques des enfants normo-entendants ; à droite : espaces vocaliques des enfants porteurs d'IC.

Les données formantiques, issues d'une tâche de dénomination, sont normalisées.

Nos résultats sur des tâches de dénomination ou de répétition de mots indiquent que les scores de production correcte sont plus faibles chez les enfants porteurs d'IC que chez

les enfants tout-venant. L'intelligibilité de ces enfants est également réduite (Aubert, 2016). Concernant la production des voyelles en particulier, nous avons montré (Figure II-24) que les enfants porteurs d'IC produisent plus de variation intra-catégorielle et moins de distances inter-catégorielles que les enfants du groupe contrôle, du même âge : leurs réalisations vocaliques acoustiques sont donc moins précises. Concernant les consonnes, des différences sont également observées en termes de contrôle du voisement et du bruit de friction des consonnes fricatives (Grandon *et al.*, 2017).

Nous poursuivons ces premiers travaux dans le cadre du projet ANR *HearCog*, porté par Pascal Barone, du CERCO à Toulouse, en collaboration avec Olivier Pascalis (LPNC) et Anne Vilain (GIPSA-lab). De nouvelles données sont actuellement recueillies, en utilisant le protocole développé dans le cadre du projet EULALIES (voir II.4.4).

Nous espérons que ces travaux contribueront à apporter des stratégies de rééducation pour les enfants présentant des TDSP liés à un déficit auditif et dont le *monitoring* verbal de soi et d'autrui est altéré.

Publications associées (articles de revues, chapitres d'ouvrage, congrès avec actes, mémoire, thèse)

- Grandon, B., Vilain, A., Løevenbruck, H., Schmerber, S. & Truy, E. (2017). Realisation of voicing by French-speaking CI children after long-term implant use: An acoustic study. *Clinical Linguistics & Phonetics*, 31, 1-14. doi: 10.1080/02699206.2017.1302511 - PMID: 28362227, halshs-01500526-1
- Grandon B., Vilain A., Løevenbruck H. (2016). Realization of voicing by French-speaking CI children after long-term implant use: an acoustic study. *International Clinical Phonetics and Linguistics Association Conference (ICPLA) 2016*, Halifax, Canada, June 15-18 2016. Oral presentation.
- Grandon B., Vilain A., Løevenbruck H. (2016). Acoustic characterization of coarticulation in stop-vowel sequences by French speaking adults vs. children: A preliminary study. *International Clinical Phonetics and Linguistics Association Conference (ICPLA) 2016*, Halifax, Canada, June 15-18 2016. Poster 2.
- Grandon B., Vilain A., Løevenbruck H. & Schmerber S. (2014). Vowel spaces in French children wearing cochlear implants. *Workshop on Late Stages in Speech and Communication Development*, 3-4 April 2014, London, UK. hal-01079964
- Scarbel L., Vilain A., Løevenbruck H. (2013). Phonetic characteristics of speech production by French children wearing cochlear implants. *International Child Phonology Conference*, Nijmegen, The Netherlands, June 10-12 2013. - hal-00838770.
- Scarbel L., Vilain A., Løevenbruck H., Schmerber S. (2012). An acoustic study of speech production by French children wearing cochlear implants. *3rd Early Language Acquisition Conference*, Lyon, 5-7 December 2012. - hal-00781655

Mémoire d'orthophonie : Corinne Aubert, 2016.

Mémoire de M2R : Lucie Scarbel, 2012.

Thèse co-encadrée : Laura Machart, en cours.

Financements :

Ce thème a été financé par la Région Rhône-Alpes, ARC2 Handicap sensoriel et moteur 2013-2016, porté par Anne Vilain (GIPSA-lab) et l'ANR (HearCog 2017-2021 ANR-17-CE28-0006-05, porté par Pascal Barone).

Collaborations :

Grenoble : GIPSA-lab, LPNC, CRTLA du CHU

Toulouse : CerCo

Canada : Université de Montréal, Memorial University Newfoundland, Université du Québec à Montréal

II.4.6. Développement du *monitoring* inter-modal du langage et du visage d'autrui

Projet CROSSMOD (Grenoble - Giessen, Allemagne) : nourrissons

J'ai participé au projet ANR-DFG international *CROSSMOD*, mené par Olivier Pascalis, du LPNC, qui visait à explorer à partir de quand et comment les nourrissons sont capables de lier indices de parole auditifs et faciaux.



Figure II-25. A gauche : les stimuli présentés aux bébés ; à droite : un bébé en plein monitoring visuel d'autrui

Nos résultats (Kubicek *et al.*, 2014a) indiquent que dès 4.5 mois, les nourrissons allemands sont capables d'associer des stimuli audio et visuels *présentés successivement*, que ce soit dans une langue non-maternelle ou maternelle. Toutefois, à 6 mois, les bébés ne sont capables d'associer les stimuli audio et visuels que pour leur langue maternelle, reflétant un affinage perceptif (« perceptual narrowing »). Nous avons examiné l'influence du style de parole (Kubicek *et al.*, 2014b). Les résultats indiquent que les nourrissons allemands de 12 mois ne sont capables de détecter l'association entre les stimuli audio et visuels que pour leur langue maternelle et pour le style de parole adressé au nourrisson (*Infant-Directed*). Nous avons aussi étudié les mouvements oculaires des nourrissons. La détection de la langue maternelle dans les stimuli audio semble orienter l'attention visuelle (vers les yeux plutôt que la bouche) et influencer le suivi visuel d'autrui chez les nourrissons dès 12 mois (Kubicek *et al.*, 2013). Nous avons examiné si les nourrissons sont capables de percevoir la cohérence entre corrélats auditifs et visuels du genre. Nos résultats indiquent que la perception multisensorielle intégrée du genre émerge vers 9 mois (Hillairet de Boisferon *et al.*, 2015). En résumé, ces travaux suggèrent que le *monitoring* verbal allocentré visuel émerge précocement (dès 6 mois).

Nous bénéficions d'un nouveau financement de l'ANR-DFG (projet *BabyFaceSpeech*, porté par Olivier Pascalis, LPNC) qui nous permettra d'étendre ces premiers travaux.

Publications associées (articles de revues, thèse)

- Hillairet de Boisferon, A., Dupierrix, E., Quinn, P.C., Løevenbruck, H., Lewkowicz, D.J., Lee, K., & Pascalis, O. (2015). Perception of Multisensory Gender Coherence in 6 and 9-month-old Infants. *Infancy*, 20 (6), 661-674. doi: 10.1111/inf.12088 - PMC4637175 - hal-01203120
- Kubicek C., Hillairet de Boisferon A., Dupierrix E., Løevenbruck H., Gervain J., and Schwarzer G. (2013). Face-scanning behavior to silently talking faces in 12-month-old infants: The impact of pre-exposed auditory speech. *International Journal of Behavioral Development*, 37 (2), 106-110. doi:10.1177/0165025412473016 - hal-00864344

- Kubicek, C., Gervain, J., Hillairet de Boisferon, A., Pascalis, O., Løevenbruck, H., & Schwarzer, G. (2014). The influence of infant-directed speech on 12-month-olds' intersensory perception of fluent speech. *Infant Behavior & Development*, 37(4), 644-651. doi:10.1016/j.infbeh.2014.08.010. PMID: 25238663 - hal-1099560
- Kubicek C., Hillairet de Boisferon A., Dupierrix E., Pascalis O., Løevenbruck H., Gervain J. & Schwarzer G. (2014). Cross-modal matching of audiovisual German and French fluent speech in infancy. *PLoS ONE* 9(2): e89275. doi:10.1371/journal.pone.0089275 - hal-00964518
- Kubicek, C., Gervain, J., Løevenbruck, H., Pascalis, O., & Schwarzer, G. (2018). Goldilock vs. Goldlöckchen: Visual speech preference for same-rhythm-class languages in 6-month-old infants. *Infant and Child Development*. doi: 10.1002/icd.2084 - hal-01741631
- Pascalis O., Løevenbruck H., Quinn P. C., Kandel S., Tanaka J. W. & Lee K. (2014). On the Links Among Face Processing, Language Processing, and Narrowing during Development. *Child Development Perspectives*. 8 (2), 65-70. doi: 10.1111/cdep.12064. PMC4164271 - hal-00964721
- Richoiz, A.-R., Quinn, P. C., Hillairet de Boisferon, A., Berger, C., Løevenbruck, H., Lewkowicz, D. J., Lee, K., Dole, M., Caldara, R. and Pascalis, O. (2017). Audio-Visual Perception of Gender by Infants Emerges Earlier for Adult-Directed Speech. *PLOS ONE*, 12, 1-15. doi:10.1371/ journal.pone.0169325 - PMC5218491 - hal-01519114

Thèse co-encadrée : Olivier Clerc, en cours.

Financements :

Ce thème a été financé par un projet ANR-DFG 2011-2013, porté par Olivier Pascalis, intitulé Crossmod : Cross-modal integration of language and face cues in German and French infants. Il est de nouveau financé par un projet ANR-DFG BabyFaceSpeech, 2018-21, porté par Olivier Pascalis.

Collaborations :

Grenoble : GIPSA-lab, LPNC, Maternité du CHU
Paris : LPP
Allemagne : Univ. Giessen

Chapitre III.

APPLICATIONS THERAPEUTIQUES, EDUCATIVES ET TECHNOLOGIQUES

Dans cette dernière partie, je décris succinctement les applications thérapeutiques, éducatives et technologiques possibles de ces propositions théoriques, dans le domaine de la remédiation des troubles du langage, visant à maintenir, rétablir ou réparer ces trois fonctions essentielles : communication, pensée et auto-nécessaire. Ce sera mon objectif dans les années à venir de les tester effectivement.

III.1. REMEDIATION ET PREVENTION DES TROUBLES DU LANGAGE CHEZ L'ENFANT

III.1.1. Apport de la modalité gestuelle dans les troubles développementaux du langage oral

Les résultats préliminaires des études sur les enfants présentant des troubles du langage oral auxquelles j'ai participé avec Marion Dohen, Amélie Rochet-Capellan et Anne Vilain au GIPSA-lab, en collaboration avec Estelle Gillet-Perret du CRTLA du CHUGA (présentés dans la section II.4.3) montrent que ces enfants utilisent plus de gestes compensatoires que les enfants tout-venant. De plus, les gestes produits par les adultes semblent les aider lors de l'apprentissage de mots nouveaux. Il me semble donc adapté de proposer des pistes de rééducation ou de remédiation s'appuyant sur les gestes pour compenser les déficits phonologiques, lexicaux ou grammaticaux de ces enfants. L'utilisation de gestes manuels et corporels (cf. II.4.3.2) de la part de l'enfant comme de la part de ses interlocuteurs, pourraient en effet contribuer à dynamiser les capacités langagières de ces enfants.

III.1.2. Intervention orthophonique fondée sur une analyse différentielle des profils langagiers

Dans le cadre du projet EULALIES (cf. II.4.4), et notamment de la thèse de Geneviève Meloni, co-dirigée avec Anne Vilain de GIPSA-lab et Andrea MacLeod de l'Université de Montréal, nous avons montré que les données langagières récoltées avec notre batterie de tests peuvent faciliter l'analyse linguistique des productions enfantines et la construction de profils psycholinguistiques, permettant à terme d'établir un diagnostic différentiel et de construire un plan d'intervention. Nous avons en particulier montré (Meloni, 2018, Meloni *et al.*, 2018) qu'à partir des profils radars issus des scores individuels à la batterie de tests (cf. II.4.4), une analyse phonologique multilinéaire permet de révéler les difficultés spécifiques de chaque enfant.

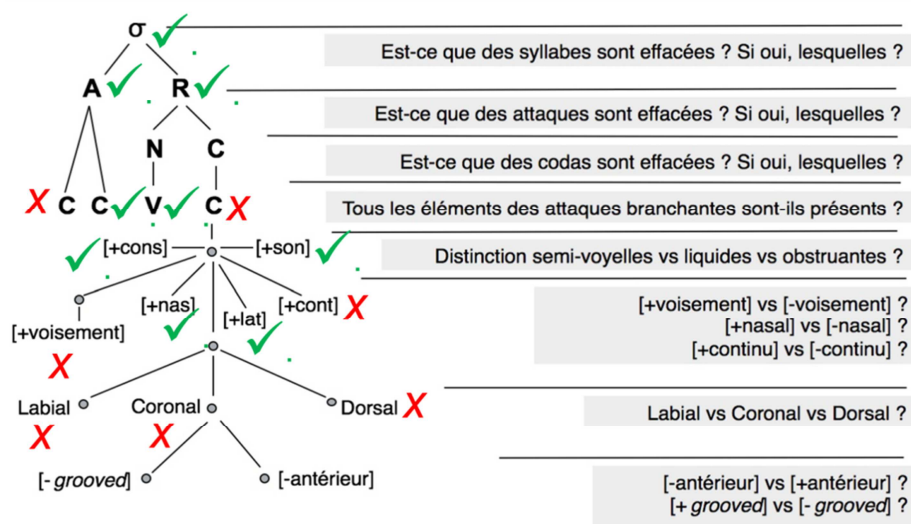


Figure III-1. Analyse multilinéaire des productions d'un enfant présentant un trouble phonologique

En s'appuyant sur le cas d'un enfant diagnostiqué avec un trouble phonologique, nous avons montré la pertinence d'une analyse multilinéaire des scores de cet enfant aux épreuves de production et de perception proposées. L'analyse, proposée par Geneviève Meloni, est résumée sur la Figure III-1. Elle indique que le cœur des difficultés de l'enfant semble être au niveau des représentations des groupes consonantiques en attaque, en coda ou en frontière de syllabe. On observe que les patrons d'erreurs segmentaux (notamment les difficultés au niveau du lieu d'articulation, mais également au niveau du mode ou de la sonorité) sont en lien avec la position des segments dans la syllabe. On peut conclure à une suppression de la ligne d'association entre le segment et la position syllabique dans une attaque ou une coda tranchante. Ce serait donc la difficulté à produire des groupes consonantiques qui entraînerait les erreurs observées.

Cette analyse fournit une stratégie de rééducation. On pourrait en effet envisager de travailler en premier lieu la production des groupes consonantiques (particulièrement en position médiane) avant de s'intéresser au traitement des erreurs sur les traits segmentaux. L'hypothèse est que l'établissement de la ligne d'association permettra la suppression des patrons d'erreurs segmentaux. Sans cette analyse multilinéaire, la rééducation aurait ciblé trop d'éléments à la fois, avec le risque de ne pas être efficace.

Collaborations :

Grenoble : GIPSA-lab, LPNC, CRTLA du CHU
Lyon : Dynamique Du Langage
Belgique: Université de Liège
Canada : Université de Montréal, Memorial University Newfoundland

III.1.3. Prévention des troubles langagiers : interventions langage et musique chez le nourrisson

Dans le cadre d'un projet art-science exploratoire du Babylab-LPNC, intitulé « les Langues de Babylab », financé par l'IDEX Université Grenoble Alpes, en partenariat avec l'artiste Aurélie Maisonneuve et l'association MÉDIARTS (<http://mediarts38.fr>), nous avons amorcé un paradigme de recherche novateur pour l'étude du nourrisson visant à examiner l'impact d'interventions artistiques, musicales et vocales, sur le développement communicatif du nourrisson. Un documentaire donnant à voir la construction du projet sera diffusé en ligne prochainement. Si les résultats expérimentaux de ce projet sont encore trop minces pour aboutir à des conclusions scientifiques, les retours positifs des parents et des professionnels de la petite enfance nous ont incités à ancrer davantage ce projet science-art-parentalité dans le territoire grenoblois, en travaillant avec des artistes locaux et en initiant des expérimentations du Babylab-LPNC hors-les-murs-universitaires.

Nous avons donc lancé un nouveau projet, « Langues et Musiques de Babylab », qui articule recherche scientifique et action culturelle et sociale. Sur le plan scientifique, en échangeant avec les artistes, le projet vise à mettre en place une situation expérimentale permettant d'évaluer l'apport, sur le développement du langage du nourrisson, d'une série d'interventions musicales précoces, en environnement collectif familial. Nous collaborons avec Myriam Roulet et Bertille Puissat, deux musiciennes grenobloises centrées sur la voix. Nous nous attendons à ce que, entre le début et la fin de l'étude, la taille du vocabulaire des nourrissons ayant bénéficié de cette initiation à la musique et à l'art vocal augmente plus que celle d'un groupe témoin. Nous nous attendons également à ce que leurs productions vocales et gestuelles soient plus variées. Sur le plan de la parentalité, les familles seront encouragées à participer aux échanges avec artistes et scientifiques sur les liens entre musique, langage et éveil.

Publications et retombées associées (documentaire vidéo, diffusion)

“Les Langues de Babylab”, documentaire de Simon Barral-Barron, 2019. Diffusion en ligne prochaine.
Animations grands publics dans des salles de spectacle de la région grenobloise ainsi que dans des espaces petite enfance (crèches, médiathèques).

Financements :

Ce thème a été financé par un projet IDEX RSC de l'Université Grenoble Alpes 2018 et bénéficie d'un nouveau financement de l'IDEX en 2019.

Collaborations :

Grenoble et sa région : GIPSA-lab, LPNC, Maternité du CHU, Association Médiarts, Ville de Fontaine

III.2. PRISE EN CHARGE DES TROUBLES DE LA PAROLE INTERIEURE : HAV ET RUMINATIONS

Nos recherches sur les dysfonctionnements de la parole intérieure, notamment les ruminations mentales (cf. II.2.2) et les hallucinations auditives verbales (cf. sections I.2.3 et II.3) suggèrent que ces troubles sont liés à une production inadaptée de la parole

intérieure : excessivement productive dans le premier cas, et liées à un défaut d'agentivité dans l'autre. Elles semblent de plus montrer que le processus de production de parole est engagé jusqu'au niveau de la planification articulatoire, le système moteur semblant bien impliqué.

Les premiers travaux de Ladislav Nalborczyk dans le cadre de sa thèse indiquent de plus que le désengagement du système moteur orofacial par la relaxation est bénéfique dans le cas de ruminations mentales induites. Nous avons donc suggéré qu'il pourrait être approprié de proposer des stratégies de suppression ou de diminution des HAV ou des ruminations fondées sur la relaxation. Cela reste à démontrer pratiquement.

Collaborations :

Grenoble : CHUGA, LIP

III.3. TROUBLE DE LA VOIX : SYNTHÈSE DE PAROLE AUDIBLE À PARTIR DE PAROLE SILENCIEUSE

J'ai participé à un projet du GIPSA-lab financé par le Bonus Qualité Recherche (BQR) de l'Institut National Polytechnique de Grenoble (INPG). Ce projet intitulé « la voix du silence » portait sur la caractérisation acoustique et physiologique de la voix silencieuse et du murmure inaudible. La voix silencieuse, le murmure inaudible et le chuchotement sont divers modes de production de parole qui représentent des défis scientifiques et technologiques pour le traitement du signal avancé ainsi que pour la compréhension des mécanismes de phonation et des représentations mentales de la parole. Ils sont aussi de forts enjeux de recherche appliqués en interface homme-machine et en télécommunications. En effet, si l'on est capable de synthétiser de la parole audible à partir de la parole inaudible, on peut communiquer à distance en toute confidentialité et sans polluer l'espace sonore. On peut également remédier à certains troubles affectant la voix, telles que la dysphonie, la laryngectomie, etc.

L'objectif principal du projet était de rendre audible la parole murmurée. Le projet s'est fait en collaboration avec des chercheurs du *Nara Institute of Science and Technology* (NAIST) au Japon qui ont mis au point un dispositif, appelé capteur NAM (*Non Audible Murmur*) permettant l'acquisition de signaux sonores de très faible intensité, par conduction des tissus (Nakajima *et al.*, 2003). Le capteur est un microphone contact placé sur la peau juste en dessous de l'oreille (cf. Figure III-3, photo de droite).

Un premier système de conversion de voix de la parole murmurée vers la parole à haute voix a été proposé par l'équipe japonaise. Ce système utilise un modèle GMM (*Gaussian Mixture Model* ou modèle de mélange gaussien), fondé sur un apprentissage de l'appariement (*matching*) entre énoncés chuchotés et énoncés à voix haute (Nakajima *et al.*, 2003 ; Toda & Shikano, 2005). Mais l'intelligibilité et le naturel de la voix convertie par ce système étaient assez médiocres. Le travail de thèse proposé à Viet-Anh Tran en septembre 2006 (thèse co-dirigée avec Gérard Bailly, soutenue le 28 janvier 2010), visait à travailler en collaboration avec les chercheurs du NAIST pour proposer des améliorations au système existant. Trois pistes ont été explorées (Tran *et al.*, 2010).

La première piste a concerné le rendu de l'intonation. L'intonation est en effet essentielle pour la démarcation et la hiérarchisation des unités du message parlé, pour la transmission des modalités (question ou affirmation par exemple) ou la spécification de

la position du·de la locuteur·rice vis-à-vis de son discours, par exemple. Elle est véhiculée par la hauteur de la voix ou la fréquence fondamentale (F_0) du signal audio ainsi que par le rythme (la durée des syllabes). Nous avons pour cela enregistré un corpus de phrases prononcées par un locuteur francophone en mode normal (voix haute) et chuchoté. Nous avons analysé les signaux audio captés par un microphone tête standard ainsi que par le capteur microphone contact NAM et avons amélioré l'apprentissage du système GMM en respectant mieux l'alignement temporel des paires d'énoncés et en introduisant des connaissances phonétiques sur le voisement dans le modèle. La Figure III-2 ci-dessous illustre les améliorations de la courbe de F_0 apportée par cette nouvelle méthode, par rapport au système de référence. Nous avons évalué ces diverses améliorations de deux façons. Nous avons mené des tests objectifs, s'appuyant sur des mesures de distorsion spectrale et d'écart de F_0 . Nous avons également mené des tests perceptifs subjectifs, en demandant à des participants français d'évaluer le naturel et l'intelligibilité des signaux convertis. Les résultats indiquent que nos améliorations sont perceptibles.

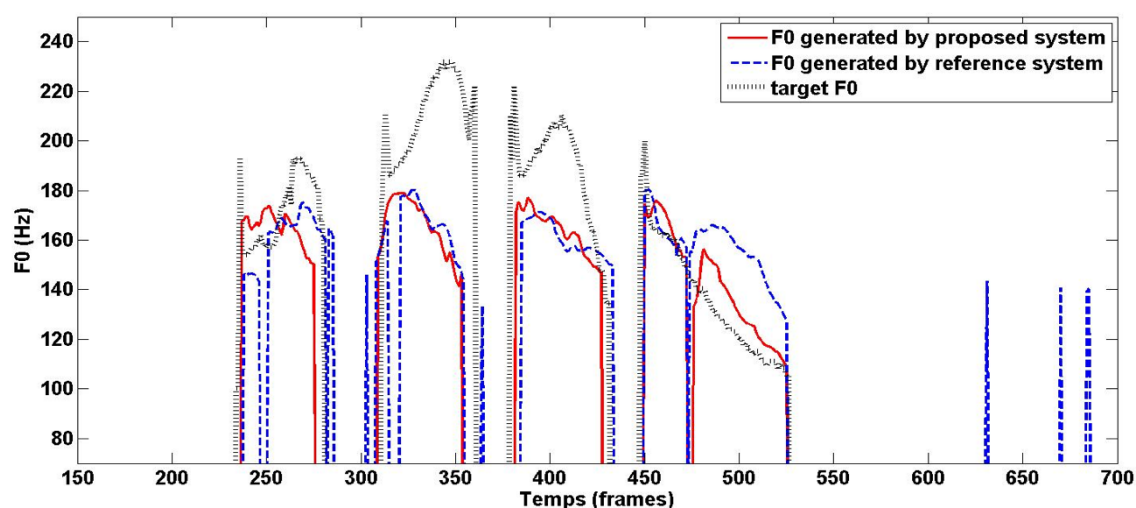


Figure III-2. Courbe de F_0 originale et resynthétisées par les deux méthodes pour l'énoncé « Armstrong tombe et s'envole. ». Les erreurs de détection sont moins nombreuses avec notre méthode.

La deuxième piste a été celle de l'intégration de données visuelles, complémentaires aux données acoustiques. En effet, il est démontré que certains paramètres visuels liés à des mouvements orofaciaux (étirement, ouverture ou arrondissement des lèvres, ouverture de la mâchoire, visibilité des dents et de la langue, mouvements du larynx) contribuent grandement à l'intelligibilité de la parole, notamment dans le bruit, ou bien lorsque la parole est chuchotée. Viet-Anh Tran a ainsi examiné la contribution de paramètres visuels en entrée et en sortie du système de conversion pour l'intelligibilité et le naturel de la parole convertie. En premier lieu, nous avons recueilli un corpus de phrases prononcées par un locuteur en mode normal (voix haute) et chuchoté. Le dispositif expérimental comportait le capteur microphone contact NAM ainsi que deux caméras (face et profil) pour capturer le mouvement de 142 billes colorées collées sur le visage (cf. Figure III-3).



Figure III-3. Enregistrements NAM et vidéo.

*Des billes colorées sont collées sur le visage du locuteur pour capturer les mouvements faciaux.
Le microphone NAM est fixé sur le cou du locuteur sous l'oreille*

En utilisant la méthode de suivi facial mise au point par Gérard Bailly au Département Parole et Cognition de GIPSA-lab, nous avons ainsi récupéré pour ce locuteur, cinq paramètres correspondant à la rotation de la mâchoire, aux mouvements verticaux et horizontaux des lèvres et aux mouvements du larynx. Ensuite, ces paramètres visuels ont été combinés aux paramètres acoustiques. Viet-Anh Tran a testé différentes valeurs de pondération des paramètres visuels. Il a appliqué alors la méthode de conversion améliorée à cette combinaison de paramètres audio et visuels pour convertir la parole audiovisuelle chuchotée en de la parole audiovisuelle à voix haute. Il a ensuite mené des tests objectifs et subjectifs comparatifs sur la parole audio et audiovisuelle convertie à partir des données audio et audiovisuelles. Les résultats (Tran *et al.*, 2010) montrent la contribution des données visuelles au système de conversion.

Enfin, la troisième piste explorée a consisté à évaluer l'apport des modèles de Markov cachés (HMM) à la conversion de voix. Au lieu d'utiliser un apprentissage par appariement entre phrases chuchotées et à voix haute par GMM, Viet-Anh Tran a testé une nouvelle démarche consistant à combiner une phase de reconnaissance statistique de la parole audiovisuelle chuchotée (par HMM), tenant compte d'informations phonétiques, suivie d'une phase de synthèse de parole à voix haute. Les résultats obtenus sont inférieurs à ceux obtenus avec le système GMM amélioré, mais des perspectives intéressantes dans ce domaine peuvent être proposées.

Dans le cadre des travaux sur la parole intérieure présentés ci-dessus (II.2.1) une autre piste intéressante est celle de l'utilisation de signaux EMG issus de capteurs positionnés de façon adéquate sur le visage. Des travaux prometteurs ont été menés sur ce terrain par Kapur *et al.* (2018), avec l'interface de parole silencieuse *AlterEgo*.

Nos propres travaux sur le recueil de traces physiologiques de la parole intérieure ou silencieuse méritent donc d'être étendus à des applications technologiques et cliniques.

Publications associées (articles de revues, congrès avec actes, mémoires, thèse)

- Heracleous P., Beutemps D., Tran V.-A., Løevenbruck H. & Bailly G. (2009). Exploiting visual information for NAM recognition. *IEICE Electronics Express*, 6, 2, 77-82. - hal-00357985
- Tran V.-A., Bailly G., Løevenbruck H. & Toda T. (2010). Improvement to a NAM-captured whisper-to-speech system. *Speech Communication*. 52 (4), 314-326. - hal-00459973
- Tran V.-A., Bailly G., Løevenbruck H. & Toda T. (2009). Multimodal HMM-based NAM-to-speech conversion. *Proceedings of Interspeech 2009*, Brighton, UK, 6 – 10 September 2009, 656-659. - hal-00419232
- Tran V.-A., Bailly G., Løevenbruck H. & Jutten C. (2008). Amélioration de la conversion de voix chuchotée enregistrée par capteur NAM vers la voix audible. *Actes des XXVIIèmes Journées d'Etude sur la Parole (JEP'08)*, Avignon, 9 – 13 juin. - hal-00339058

- Tran V.-A., Bailly G., Lœvenbruck H. & Toda T. (2008). Predicting F0 and voicing from NAM-captured whispered speech. *Proceedings of the 4th conference on Speech Prosody 2008*, Campinas, Brazil, May 6 – 9 2008. - hal-00339045
- Tran V.-A., Bailly G., Lœvenbruck H. & Jutten C. (2008). Improvement to a NAM captured whisper-to-speech system. *Proceedings of Interspeech 2008*, Brisbane, Australia, 22-26 September 2008. - hal-00333288

Thèse co-encadrée : Viet-Anh Tran, soutenue en 2010.

Financements :

Ce thème a bénéficié d'un financement de la fédération ELESa 2005 et du BQR de l'INPG 2006.

III.4. APPORT D'INFORMATIONS VISUELLES ARTICULATOIRES DANS LA REEDUCATION DE PATIENTS PRESENTANT UNE APHASIE DE BROCA NON-FLUENTE

Je participe à un projet mené par Monica Baciú au LPNC sur la rééducation des patients aphasiques de Broca par l'apport d'informations articulatoires (linguales et labiales) visuelle, via le logiciel *Ultraspeech player*, développé par Thomas Hueber à GIPSA-lab. Nous avons évalué cette méthode de rééducation chez deux patientes présentant une aphasie non fluente chronique post-AVC. L'évaluation de l'effet de cette méthode s'est fondée sur la comparaison, avant et après rééducation, des données multimodales permettant (a) une analyse quantitative des paramètres acoustiques du signal de parole, (b) une analyse qualitative des performances langagières et des scores de langage, et (c) une évaluation neurophysiologique des réseaux anatomo-fonctionnels du langage. Une partie des analyses a été menée par Célise Haldin au cours de son M2R, co-encadré par Monica Baciú et moi-même. Nos résultats montrent l'efficacité de la méthode chez la patiente P1 mais un bénéfice limité chez la patiente P2 (qui s'est avérée présenter une lésion pariétale en plus de la lésion de la région de Broca). Qualitativement, la dénomination, la lecture, la répétition des mots et le jugement de rimes s'améliorent chez P1, tandis que P2 décline en lecture et jugement de rimes. Quantitativement, P1 améliore sa production de voyelles et consonnes alors que P2 améliore très partiellement la qualité de quelques consonnes et voyelles uniquement et décline pour certaines épreuves. En termes de corrélats cérébraux, nous observons, chez P1, une modulation de l'activité des réseaux langagiers, en accord avec l'amélioration comportementale de la parole et du langage ainsi que des patrons de réorganisation fonctionnelle efficaces. Dans Haldin *et al.* (sous presse), nous discutons ces résultats, dans le cadre des modèles théoriques du contrôle de la parole et nous présentons les conditions particulières de l'application de cette méthode chez les patients présentant une aphasie non-fluente.

Ces résultats encourageants, nous ont conduites à appliquer de nouveau cette méthode de rééducation chez un nouveau patient présentant une aphasie de Broca non fluente typique. Les données, recueillies par Célise Haldin sont encore en cours d'analyse, mais il semble que les bénéfices de cette méthode soient de nouveau démontrés. Nous envisageons donc de poursuivre dans cette voie.

Publications associées (articles de revues)

- Acher, A., Fabre, D., Hueber, T., Badin, P., Detante, O., Cousin, E., Pichat, C., Lœvenbruck, H., Haldin, C. & Baciú, M. (2016). Retour visuel en aphasologie : résultats comportementaux, acoustiques et en neuroimagerie. Nathalie Joyeux, Silvia Topouzkhian. *XVIèmes Rencontres Internationales d'Orthophonie. Orthophonie et technologies innovantes*, Décembre 2016, Paris, France. Ortho Edition, 227-260. - hal-01419292

Haldin, C., Acher, A., Kauffmann, L., Hueber, T., Cousin, E., Badin, P., Perrier, P., Fabre, D., Perennou, D., Detante, O., Jaillard, A., Løevenbruck, H. & Baciú, M. (sous presse). Effet de la rééducation perceptivo-motrice sur la récupération de la parole chez deux patientes avec aphasie non fluente chronique post-AVC. *Revue de Neuropsychologie*.

Haldin, C., Acher, A., Kauffmann, L., Hueber, T., Cousin, E., Badin, P., Perrier, P., Fabre, D., Perennou, D., Detante, O., Jaillard, A., Løevenbruck, H. & Baciú, M. (2017). Speech recovery and language plasticity can be facilitated by Sensori-Motor Fusion (SMF) training in chronic non-fluent aphasia. A case report study. *Clinical Linguistics & Phonetics*, Taylor & Francis, 2017, pp.1 - 27. <10.1080/02699206.2017.1402090> - hal-01651198

Thèse co-encadrée : Céline Haldin, en cours.

Financements : Ce thème bénéficie d'un financement IRS IDEX UGA, .projet SEMO 2019-22, porté par Monica Baciú

Collaborations :

Grenoble : LPNC, GIPSA-lab, CHUGA

Chapitre IV.

PROGRAMME DE RECHERCHE

IV.1. VERS UN MODELE NEUROCOMPUTATIONNEL INTEGRE DE LA CYBERNETIQUE VERBALE SOUTENANT LA TRIADE COMMUNICATION-COGNITION-AUTONOESE

Dans la lignée des travaux de recherche que j'ai menés depuis maintenant vingt ans, je propose de contribuer à mieux caractériser les substrats neurobiologiques de la cybernétique verbale qui sous-tend la triade communication-cognition-autonoèse. Comme je l'ai développé au Chapitre I, je défends l'hypothèse que le langage remplit trois fonctions majeures et intriquées : la fonction communicative, qui nous permet d'échanger avec nos congénères, la fonction cognitive, qui nous permet d'élaborer et de structurer notre pensée, notre raisonnement et enfin la fonction autoéotique qui contribue à la conscience de soi dans le temps. Pour que chacune de ces trois fonctions se déploie de façon adéquate, des dispositifs de *monitoring* verbal et d'attribution de l'agentivité sont nécessaires. J'ai fait quelques propositions théoriques au Chapitre I, sur ces dispositifs cybernétiques pour chacune des fonctions de la triade, indépendamment. Ces modélisations théoriques s'appuient sur des données et des connaissances en neuropsycholinguistique. J'ai recueilli des données expérimentales qui contribuent à les tester expérimentalement (cf. Chapitre II). Il reste toutefois de nombreux points à éclaircir ou compléter dans ces premières modélisations.

De plus, ces trois fonctions du langage sont intriquées et le contrôle de l'une joue sur le contrôle de l'autre. Je cherche donc à construire une modélisation neuro-computationnelle intégrée de la cybernétique verbale, qui rassemblera en un seul formalisme les modèles de *monitoring* verbal et d'attribution de l'agentivité que j'ai proposés pour chaque fonction dans mes travaux antérieurs.

Je compte continuer de mener une démarche triple : théorique – empirique – appliquée.

Sur le plan théorique, je chercherai un formalisme neurocomputationnel intégré visant à décrire la cybernétique de la triade communication-cognition-autozoïque. J'ai décrit les premières briques de ce formalisme encore en construction au 1.2.4. J'envisage, en collaboration notamment avec mes collègues de l'équipe Langage du LPNC, de le compléter et de le développer.

Sur le plan empirique, dans le but d'étayer ces propositions théoriques, je continuerai aussi de recueillir des données phénoménologiques, comportementales, acoustiques, articulatoires, physiologiques et neurales, sur la production et la perception du langage, chez le nourrisson, l'enfant ou l'adulte, mono- ou plurilingues, sains ou présentant des troubles affectant la triade.

Sur le plan appliqué, j'envisage de développer les retombées thérapeutiques et éducatives de ces résultats théoriques dans les domaines de la remédiation et de la prévention des troubles du langage, en visant à maintenir, rétablir ou réparer la triade. Je continuerai également de m'investir dans la diffusion des savoirs, notamment dans la médiatisation et la participation aux événements scientifiques destinés au grand public (Semaine du Cerveau, Fête de la Science, Forum de Sciences Cognitives) et dans la contribution à des projets art-science.

Enfin, je compte continuer à participer activement à la structuration des sciences du langage et de la cognition : outre mon engagement dans l'encadrement d'étudiants de master et de thèse, je compte maintenir mon implication dans le Master de Sciences Cognitives de Grenoble, et ma participation au *Pôle Grenoble Cognition* et au programme grenoblois *NeuroCog*.

IV.2. MISE A L'ÉPREUVE DE LA MODELISATION PROPOSEE PAR LE RECUEIL DE DONNEES EXPERIMENTALES

IV.2.1. Cybernétique verbale pour la fonction de communication

Dans le cadre de financements en cours (ARC région AURA, ANR HearCog et ANR-DFG BabyFaceSpeech), je continue de travailler sur le développement des capacités de *monitoring* verbal égocentré et allocentré chez l'enfant.

J'envisage ainsi dans le court terme de continuer à recueillir et analyser des données sur la production de parole chez les enfants de 3 à 11 ans et de mieux expliquer comment se développe le *monitoring* verbal égocentré et allocentré chez les enfants tout-venant, les enfants présentant un trouble du développement des sons de parole (TDSP) et les enfants sourds porteurs ou non d'implants cochléaires. Ceci se fera dans le cadre du projet EULALIES, que je co-dirige avec Anne Vilain du GIPSA-lab, et en collaboration avec Elsa Spinelli du LPNC, Marion Dohen et Maëva Garnier du GIPSA-lab, Estelle Gillet-Perret du CRTLA du CHU de Grenoble, Sophie Kern et Mélanie Canault du Laboratoire Dynamique du Langage à Lyon, Andrea MacLeod de l'Université de Montréal au Canada. Nous sommes également en contact avec Yvan Rose de la *Memorial University of Newfoundland*, St John's, Canada et Christelle Maillart de l'Unité de Logopédie Clinique, Université de Liège, Belgique. Ce projet intègre les travaux de thèses en cours de Geneviève Meloni et Laura Machart, que je co-encadre toutes les deux, et utilise la batterie d'épreuves complémentaires que nous avons mise au point.

Le projet EULALIES nécessite de recueillir un nombre conséquent de données chez les enfants des 8 tranches d'âges considérées (de la petite section de maternelle au CM2 inclus), pour obtenir des scores normatifs qui permettront d'évaluer les enfants présentant des TDSP. Nous sommes donc à la recherche de financements supplémentaires pour étendre les projets actuels.

Par conséquent, j'ai participé, toujours avec Anne Vilain du GIPSA-lab, au montage du projet Européen ITN Comm4CHILD (Communication for Children with Hearing Impairment to optimise Language Development) porté par Cécile Colin et Jacqueline Leybaert de l'ULB à Bruxelles. Si ce projet est financé, alors nous devrions obtenir une nouvelle allocation de recherche de doctorat qui nous permettra d'étendre le champ des données de parole collectées chez les enfants. Nous avons également déposé un projet ANR, porté par Anne Vilain, visant à obtenir des allocations de recherche et des ressources complémentaires.

Je participe actuellement au projet ANR-DFG *BabyFaceSpeech* porté par Olivier Pascalis du LPNC et je co-dirige avec lui la thèse de Olivier Clerc qui vient de démarrer. Nous étudions les capacités de *monitoring* audio et visuel des nourrissons monolingues et bilingues. Bien qu'un parallèle soit souvent évoqué entre l'émergence de la parole et celle du traitement des visages, aucune étude systématique du développement de ces deux compétences cognitives n'existe. Notre but est de répondre à ce manque en étudiant l'interaction entre le traitement de la parole et des visages chez le nourrisson. Nous évaluerons dans quelle mesure l'affinage perceptif (*perceptual narrowing*), qui existe pour la parole et le traitement des visages, module cette interaction et nous examinerons l'influence du bilinguisme. Sur le plan théorique, le projet *BabyFaceSpeech* vise à mieux comprendre les interactions entre traitements de la parole et des visages au cours du développement et le rôle modulateur du bilinguisme sur l'affinage perceptif multimodal. Notre projet peut également avoir des retombées cliniques, en produisant des normes développementales du traitement de la parole et des visages.

Je continuerai d'explorer les mécanismes de *monitoring* de la fonction de communication, notamment les outils par lesquels nous pouvons aider autrui à nous suivre et réciproquement. Ces outils sont d'ordre pragmatique et utilisent des marqueurs prosodiques (structure informationnelle), ainsi que des marqueurs gestuels (manuels et corporels). J'envisage pour ce faire de continuer à collaborer avec Marion Dohen et Anne Vilain du GIPSA-lab, ainsi que Mariapaola D'Imperio du Laboratoire Parole et Langage à Aix-en-Provence. Nous avons en effet des projets en cours (dont certains sont décrits au Chapitre II et au Chapitre III), dans le cadre desquels nous avons recueilli des données articulatoires (électromagnéto-métriques), vidéo et audio encore en cours d'analyse et qui nous guideront dans la mise en place d'expériences futures.

Dans mes travaux de modélisation du contrôle de la production de la parole, j'ai fait un certain nombre d'hypothèses et de conjectures neuroanatomiques, qui s'appuient sur des arguments théoriques, mais pour lesquelles il n'existe pas encore de données expérimentales probantes.

En particulier, en ce qui concerne le niveau linguistique de planification articulatoire, j'ai proposé un modèle du rôle du cervelet dans le *Monitoring* par Contrôle Prédicatif (MCP) de la parole. Ce modèle se fonde sur des travaux très récents qui montrent que le cervelet possède une représentation en deux exemplaires (voire trois) de la quasi-totalité du télencéphale (cortex et structures sous-corticales). Cette double représentation me

paraît être la clef de la correspondance entre modèles prédicteur (direct) et contrôleur (inverse), mais ceci reste à démontrer.

Je propose donc de mettre en place des expériences en neuroimagerie qui viseront à tester cette hypothèse. Il s'agira d'utiliser deux tâches, impliquant l'une plus le modèle contrôleur (inverse) et l'autre plus le modèle direct (prédicteur) et de vérifier si les activations cérébelleuses suivent le parcours cortico–ponto–cerebello–dento–thalamo–cortical prévu.

En ce qui concerne les plus hauts niveaux linguistiques (formulation prosodico-syntaxico-phonologique et conceptualisation), j'ai également fait des hypothèses sur les réseaux neuraux impliqués, en m'inspirant notamment des propositions de Baciú & Perrone-Bertolotti (2015), Duffau *et al.* (2014), Keller & Kell (2016), Llorens *et al.* (2011), Price (2012), Rauschecker & Scott (2009), Tian & Poeppel (2013) ou Tremblay & Dick (2016). Mais ces hypothèses restent à vérifier. J'envisage de le faire dans le cadre du projet multi-équipes *InLang* proposé par Monica Baciú dans le cadre de l'appel d'offre NeuroCog 2018. Ce projet a été retenu et nous bénéficions d'un financement qui devrait nous permettre d'avancer vers un connectome du langage, en nous appuyant sur les nombreuses bases de données de neuroimagerie fonctionnelle et d'électrophysiologie intracrânienne disponibles à Grenoble.

Je compte également, en collaboration avec Philippe Kahane (Laboratoire d'Epilepsie du CHU Grenoble Alpes) et Jean-Philippe Lachaux (CRNL, Lyon) recueillir de nouvelles données en électrophysiologie de scalp (EEG) et intracrânienne (iEEG) pour mieux comprendre la dynamique temporelle des activations des réseaux cérébraux liés à la production et la perception du langage.

IV.2.2. Cybernétique verbale pour la fonction de pensée

Une partie conséquente de mes travaux, dans la lignée du projet ANR INNERSPEECH que j'ai conduit de 2014 à 2018, sera dévolue à tenter de résoudre les questions qui restent ouvertes sur la parole intérieure.

En particulier, en ce qui concerne le *monitoring* des verbalisations imaginées, j'ai fait l'hypothèse qu'il pouvait s'exercer à tous les niveaux de la hiérarchie linguistique : planification articulatoire, formulation et conceptualisation. Des études en psycholinguistique montrent que nous sommes capables de détecter des lapsus lors de la production mentale de virelangues, mais ceci reste un cas tout à fait particulier. Il reste donc à mieux vérifier expérimentalement que le *monitoring* de la parole intérieure s'exerce à tous les niveaux et avec quel degré il s'exerce. Il a en effet été proposé qu'il soit plus relâché en parole intérieure qu'en parole à voix haute (« Dans la parole intérieure, il suffit que nous soyons compris de nous-mêmes », Egger, 1881). Ceci pourra se faire, par exemple, avec des expérimentations psycholinguistiques inspirées des travaux sur les lapsus.

En ce qui concerne le *monitoring* de la parole intérieure appliqué au raisonnement et à la cognition, une question importante est celle de la parole intérieure dans l'aphasie de Broca (non-fluente). Il existe en effet des dissociations entre déficit de parole intérieure et déficit de parole à voix haute dans l'aphasie. De nombreuses études rapportent que les troubles de la parole à voix haute dans l'aphasie sont accompagnés de trouble de la parole intérieure (cf. e.g. Levine *et al.*, 1982). Il existe aussi des cas de troubles de la

parole à voix haute sans trouble de la parole intérieure (Baddeley & Wilson, 1985), ce qui peut s'expliquer aisément si on considère la parole intérieure comme de la parole à voix haute tronquée à un niveau de programmation supérieur. Mais il existe aussi (Geva *et al.*, 2011 ; Langland-Hassan *et al.*, 2017) des cas, plus difficiles à expliquer, de trouble de la parole intérieure sans trouble de la parole à voix haute (selon nous, ce résultat s'explique par la tâche en jeu qui est cognitivement trop complexe à effectuer en parole intérieure car elle fait appel à de l'imagination auditive, Lœvenbruck *et al.*, 2018). Les résultats variés de la littérature s'expliquent essentiellement parce qu'il n'existe pas à l'heure actuelle de tâche adaptée permettant de tester l'utilisation efficiente (pour des tâches cognitives, pour la résolution de problème, pour le raisonnement) du langage intérieur chez les personnes présentant une aphasie. Il n'existe pas non plus de tâche adaptée à d'autres populations, notamment les personnes souffrant de troubles du spectre autistique et ayant des difficultés de production de parole.

Je co-encadre actuellement, avec Monica Baciú et Marcela Perrone-Bertolotti du LPNC, une étudiante en Master avec qui nous sommes en train de mettre au point une tâche permettant d'évaluer l'utilisation de parole intérieure spontanée dans ce type de population.

A terme, nous envisageons de déployer cette tâche auprès de diverses populations : patients aphasiques, personnes souffrant de troubles du spectre autistique (avec Cécile Souchay du LPNC et Stéphanie Smadja du CERILAC à Paris, porteuse du projet *Monologuer*, auquel je participe), jeunes enfants. Ceci devrait permettre de mieux comprendre les liens entre parole intérieure et parole à voix haute, ainsi que le rôle du langage intérieur dans le raisonnement et la cognition.

En ce qui concerne le *monitoring* du déclenchement de la parole intérieure, le projet ANR INNERPSEECH et les travaux de Ladislav Nalborczyk que j'ai encadrés sur les ruminations mentales nous ont permis de progresser sur la nature des ruminations et les processus cérébraux sous-jacents. Mais ils n'ont pas permis d'expliquer pourquoi et comment la parole intérieure devient parfois délétère. Les mécanismes qui font que la parole intérieure se met à être générée en boucle sur un même thème restent obscurs. Dans le cadre de la thèse de Ladislav Nalborczyk, nous avons projeté une expérience en stimulation magnétique transcrânienne (TMS) qui devrait nous éclairer sur certains des processus de contrôle cognitif en jeu. C'est là encore une piste de recherche que je compte explorer dans le futur.

En ce qui concerne le dispositif d'attribution de l'agentivité (en parole intérieure et en parole extériorisée), qui est crucial pour mieux comprendre le phénomène hallucinatoire, j'ai fait des hypothèses sur les comparaisons entre signaux sensoriels désirés, prédits et effectifs qui reposent sur des temporisations très fines entre des différents signaux. Les prédictions sont censées arriver avant les signaux effectifs dans le cas de la production de parole. Elles sont censées arriver après dans le cas de la perception de parole. Il existe quelques arguments en ce sens. Notamment les travaux en électrocorticographie (ECoG) de Flinker *et al.* (2015) qui montrent un décalage temporel des activations cérébrales lors de la production de parole compatible avec mes hypothèses (les pics d'activités cérébrales démarrent dans le GTS et le STS (il s'agit de répéter un mot), puis se propagent à la région de Broca (GFI) avant d'atteindre les cortex prémoteur et moteur). Les travaux de Shergill *et al.* (2004) sur les patients schizophrènes souffrant d'hallucinations auditives verbales montrent également un décalage temporel compatible avec de la production de

parole intérieure, *i.e.* une activation dans la région de Broca, suivie d'une activation des régions temporales auditives. Mais tout ceci reste bien sûr à confirmer de façon plus détaillée. J'envisage donc de poursuivre les travaux initiés dans le cadre du projet ANR INNERSPEECH en EEG intracrânienne (avec Philippe Kahane, du Laboratoire d'Epilepsie du CHU Grenoble Alpes et Jean-Philippe Lachaux du CRNL à Lyon) qui devraient permettre d'obtenir des résultats très précis sur le décours temporel des activations entre régions motrices et auditives et mieux expliquer les éventuelles temporisations de signaux sensoriels (cf. p.100). Je bénéficierai aussi des données du projet multi-équipes *InLang* cité ci-dessus, qui vise à décrire un connectome du langage et pour lequel des données de connectivité dynamique seront à terme disponibles.

IV.2.3. Cybernétique verbale pour la fonction d'autoïté

Mes recherches théoriques sur le rôle du langage dans l'autoïté me poussent à étendre mes travaux dans le domaine de la mémoire autobiographique et la mémoire prospective. Je suis en contact avec plusieurs chercheurs de la région (notamment Chris Moulin du LPNC et Thierry Atzeni du LIP de Chambéry) qui s'interrogent sur le rôle de la parole intérieure délibérée et vagabonde dans la réactivation de tâches dans la mémoire prospective ou dans le rappel de souvenirs autobiographiques. Nous envisageons de mettre en place des collaborations prochainement pour explorer ces questions.

Je m'intéresse également à la métacognition au sens plus large, et au rôle qu'y joue le langage. J'ai été contactée par Denis Perrin, du laboratoire Philosophie Pratiques et Langage à Grenoble, qui a déposé un projet ANR intitulé CORCOGART. Ce projet, qui rassemble philosophes et spécialistes des sciences cognitives, porte sur le problème de la nature et de la validité des fondements d'une théorie incorporée des facultés cognitives. Il pose un certain nombre de questions, dont celle du rôle du corps dans la cognition et la métacognition, que je rapproche de ma question sur l'implication du système moteur dans la parole intérieure et dans l'autoïté. Si ce projet est financé, cela nous permettra d'avancer sur ces questions d'incarnation et d'incorporation dans une démarche théorique pluridisciplinaire.

IV.3. APPLICATIONS THERAPEUTIQUES ET EDUCATIVES ENVISAGEES

Enfin, j'envisage de m'investir dans les applications thérapeutiques et éducatives envisagées dans mes travaux actuels, mais que je n'ai pas encore pu tester de façon rigoureuse (cf. Chapitre III). Je les résume ici succinctement et les complète par d'autres pistes potentielles.

IV.3.1. Rééducation et prévention des troubles du langage

Il s'agit, tout d'abord, de la remédiation des troubles du langage chez l'enfant.

Dans la suite des projets menés notamment avec Marion Dohen au GIPSA-lab, Il me semble pertinent de proposer des pistes de rééducation ou de remédiation s'appuyant sur les gestes pour compenser les déficits lexicaux, grammaticaux ou phonologiques des enfants présentant des troubles du langage oral. L'utilisation de gestes manuels et corporels de la part de l'enfant comme de la part de ses interlocuteurs, pourraient en effet contribuer à dynamiser les capacités langagières de ces enfants.

Toujours concernant la rééducation de la parole chez l'enfant présentant des troubles langagiers ou des déficits auditifs, dans le cadre du projet EULALIES, avec Anne Vilain au GIPSA-lab, nous pourrions proposer des pistes d'intervention orthophonique fondées sur une analyse différentielle des profils langagiers, comme expliqué ci-dessus (cf. III.1.2).

Et dans un but de prévention des troubles langagiers, j'envisage également de poursuivre la collaboration entre le Babylab-LPNC et l'association Mediarts, en m'engageant dans des projets art-science-parentalité explorant l'apport des interventions artistiques précoces et des interactions scientifiques-artistes-parents sur le développement communicatif et langagier du tout-petit.

Il s'agit, ensuite, de la rééducation de la parole chez l'adulte.

Je continuerai de m'investir dans la rééducation de la parole chez les patients aphasiques, en utilisant des technologies variées, dont l'apport de la modalité visuelle articulatoire (projet en cours porté par Monica Baciu, décrit au III.4).

Dans le cadre du projet IRS IDEX BeatBox, porté par Nathalie Henrich du GIPSA-lab, et auquel je participe, nous co-encadrons Annalisa Paroni qui vient de démarrer sa thèse et envisageons de proposer de faire évoluer les pratiques de rééducation orthophonique en s'inspirant des comportements respiratoires, phonatoires et articulatoires produits dans le *Human Beatbox*, un art vocal émergent. Le projet se centre sur la production de sons percussifs dans le *Human Beatbox*. Il ambitionne de décrire la phonétique de ce langage musical et les comportements physiologiques associés, en s'appuyant sur une comparaison systématique avec la production de séquences de consonnes occlusives et fricatives dans la parole. A terme, il permettra une nouvelle approche thérapeutique ludique et innovante fondée sur des exercices de sons beatboxés occlusifs et fricatifs. Ces exercices devraient améliorer l'articulation des consonnes occlusives (/p, b, t, d, k, g/) et fricatives (/f, v, s, z, ʃ, ʒ /) et permettre d'accroître l'intelligibilité des patients, compte tenu du rôle crucial de ces sons dans la parole. Les exercices seront développés et validés en laboratoire et leur pertinence sera ensuite testée en contexte clinique de rééducation orthophonique chez les enfants et les adultes.

IV.3.2. Hallucinations auditives verbales et ruminations mentales : à l'écoute des entendeurs de voix et des ressasseurs de mots

Dans la continuité des thèses de Lucile Rapin et Ladislav Nalborczyk, et en collaboration avec Marcela Perrone-Bertolotti et Monica Baciu (LPNC), je continuerai de m'investir sur la prise en charge des hallucinations auditives verbales et des ruminations mentales et notamment sur l'amélioration du contrôle cognitif, en tirant parti des connaissances que nous avons acquises non seulement sur les mécanismes de ces troubles mais sur leur déclenchement. J'ai pris contact avec Charles Fernyhough de l'université de Durham (UK), qui dirige le projet *Hearing the Voice*, portant sur le phénomène d'entente de voix. J'envisage d'échanger davantage avec les membres de ce projet.

Ces travaux bénéficieront aussi des collaborations avec Chris Moulin et Céline Souchay (LPNC) sur les aspects métacognitifs de la parole intérieure et ses dysfonctionnements.

J'ai échangé avec Anne-Julie Rollet et Anne-Laure Pigache, deux artistes qui ont créé un spectacle donnant à ressentir les expériences auditives des entendeurs de voix : « voix magnétiques » (<http://annelaurepigache.fr/spectacles-et-performances/voix-magnetiques>).

Leurs questions et leur spectacle m'ont poussée à m'intéresser davantage aux entendeurs de voix (qui préfèrent parfois qu'on les appelle les « écouteurs de voix ») et j'envisage de prendre contact avec le réseau des entendeurs de voix de Grenoble (<https://revgrenoble.wordpress.com>) pour écouter leurs expériences et questionner nos propres recherches et théories.

IV.3.3. Synthèse vocale à partir de la parole silencieuse

Dans la continuité des travaux menés dans le cadre du projet ANR INNERSPEECH et du projet BQR La voix du Silence (cf. III.3), portant sur le recueil de traces physiologiques de la parole intérieure ou silencieuse, des applications technologiques et cliniques sont envisageables pour convertir ces traces en signaux de parole audible. On peut ainsi envisager de remédier à certains troubles affectant la voix, telles que la dysphonie, la laryngectomie, etc.

IV.3.4. Apport thérapeutique de la pratique du langage intérieur

Comme expliqué au I.1.3.2.7, dans le cadre du projet *Monologuer*, porté par Stéphanie Smadja du CERILAC, Université Paris Diderot, j'ai eu l'occasion de mener des entretiens phénoménologiques auprès de participants amenés à l'introspection sur leur langage intérieur via la tenue d'un carnet pendant quelques semaines. Les personnes interrogées ont souligné le caractère bénéfique et plaisant de cette pratique et certains ont ajouté qu'elle a été pour eux thérapeutique. Il me semble donc que la pratique de l'énonciation, à destination d'un-e thérapeute ou de soi-même, à voix haute ou silencieuse, écrite ou non, peut constituer une piste de recherche clinique intéressante, à développer. J'envisage de continuer à collaborer sur ce thème avec Stéphanie Smadja, qui défend d'ailleurs la création d'une « linguistique clinique ».

J'envisage également de continuer à participer aux projets art-science du groupe n+1 et de la Compagnie *les Ateliers du Spectacle*, qui pourraient faire suite à « l'impromptu scientifique » sur les « Voix dans la tête » (cf. I.1.3.2.7). Ce spectacle sous une forme théâtrale à la fois ludique, scientifique et interactive invite le public à réfléchir et se questionner sur son langage intérieur et semble avoir une portée bénéfique. Je compte donc continuer de m'engager aux côtés de ce groupe de comédiens dans l'élaboration de nouveaux spectacles participatifs, mêlant art, science et introspection.

CONCLUSION

J'ai défendu l'hypothèse que le langage sous-tend la communication avec autrui, fait partie du processus de construction de la pensée et participe à l'autoconscience, d'une part, et qu'une cybernétique verbale commune est à l'œuvre, gouvernant la production et la réception des messages verbaux externes et internes et leur attribuant une agentivité, d'autre part.

Sur le plan théorique, je propose de contribuer à mieux caractériser les substrats neurobiologiques de la cybernétique verbale qui sous-tend la triade communication-cognition-autoconscience, dans l'objectif de construire un formalisme neurocomputational intégré.

Sur le plan empirique, je compte étendre le champ des données en m'investissant davantage dans l'électrophysiologie intracrânienne, en examinant plus en détail les étapes du développement de la parole chez l'enfant, en menant des enquêtes plus approfondies sur la parole intérieure et ses liens avec la métacognition, en m'intéressant à diverses populations pathologiques, afin d'étayer ces propositions théoriques ou de les corriger.

Sur le plan appliqué, j'envisage de développer les retombées thérapeutiques et éducatives de ces résultats théoriques, dans le domaine du développement, de la remédiation ou de la rééducation du langage, gardant à l'esprit que c'est par le langage que nous communiquons, pensons et existons.



RÉFÉRENCES BIBLIOGRAPHIQUES

- Ackermann, H. (2008), 'Cerebellar contributions to speech production and speech perception: psycholinguistic and neurobiological perspectives', *Trends in Neurosciences* 31(6), 265 - 272.
- Agnew, Z. K.; McGettigan, C.; Banks, B. & Scott, S. K. (2013), 'Articulatory movements modulate auditory responses to speech', *NeuroImage* 73, 191 - 199.
- Akhmatova A. (1992). *Poèmes*. Traduction de Jean Malaplate, Editions Librairie du Globe, Paris.
- Alderson-Day, B. & Fernyhough, C. (2015), 'Inner Speech: Development, Cognitive Functions, Phenomenology, and Neurobiology.', *Psychological Bulletin* 141(5), 931-965.
- Alderson-Day, B.; Mitrenga, K.; Wilkinson, S.; McCarthy-Jones, S. & Fernyhough, C. (2018), 'The varieties of inner speech questionnaire - Revised (VISQ-R): Replicating and refining links between inner speech and psychopathology', *Consciousness and Cognition* 65, 48 - 58.
- Alderson-Day, B.; Weis, S.; McCarthy-Jones, S.; Moseley, P.; Smailes, D. & Fernyhough, C. (2016), 'The Brains Conversation with Itself: Neural Substrates of Dialogic Inner Speech', *Social Cognitive and Affective Neuroscience*.
- Altmann, G. & Steedman, M. (1988), 'Interaction with context during human sentence processing', *Cognition* 30(3), 191 - 238.
- Ambrazaitis, G. & House, D. (2017), 'Multimodal prominences: Exploring the patterning and usage of focal pitch accents, head beats and eyebrow beats in Swedish television news readings', *Speech Communication* 95, 100-113.
- Amici, F.; Aureli, F. & Call, J. (2008), 'Fission-Fusion Dynamics, Behavioral Flexibility, and Inhibitory Control in Primates', *Current Biology* 18(18), 1415 - 1419.
- Anderson, J. R. (1996), 'ACT: A simple theory of complex cognition.', *American Psychologist* 51(4), 355.
- Anderson, S. R. (2010), 'How many languages are there in the world', *Linguistic Society of America*.
- Apostol, L.; Perrier, P. & Bailly, G. (2004), 'A model of acoustic interspeaker variability based on the concept of formant-cavity affiliation', *The Journal of the Acoustical Society of America* 115(1), 337-351.
- Armstrong, K. (1993), *A History of God: The 4000 Year Quest of Judaism, Christianity and Islam*, New York: New York: Knopf.
- Arnaud, L.; Schwartz, J.-L.; Loevenbruck, H. & Savariaux, C. (2008), Perception as a (Shaped) Mirror of Action: It Seems Easier to Lipread One's Own Speech Gestures than those of Somebody Else, in 'Workshop on Speech and Face to Face Communication'.
- Arnauld, A. & Lancelot, C. (1660), *Grammaire générale et raisonnée contenant les fondemens de l'art de parler, expliqués d'une manière claire et naturelle*, P. Le Petit, Paris.
- Asoulin, E. (.), 'Language as an instrument of thought', *Glossa: a journal of general linguistics* 1(1), 1-23.
- Atran, S. (1990), *Cognitive foundations of natural history: Towards an anthropology of science*, Cambridge University Press.
- Augustin, S. *Confessions, livre X, chapitre VIII*, Traduction de E. Tréhorel et G. Bouissou. http://www.augustinus.it/francese/confessioni/conf_10_libro.htm
- Augustin, S. *Sermon 288*. (<http://www.abbaye-saint-benoit.ch/saints/augustin/sermons>)
- Baars, B. (1988), *A Cognitive Theory of Consciousness*, Cambridge University Press.
- Baciu, M. & Perrone-Bertolotti, M. (2015), 'What do patients with epilepsy tell us about language dynamics? A review of fMRI studies', *Reviews in the Neurosciences* 26(3), 323-341.

- Baciu, M.; Rubin, C.; Décorps, M. & Segebarth, C. (1999), 'Hemispheric Language Dominance Testing by Means of fMRI', *Journal of Neuroimaging* 9(4), 246-247.
- Baciu, M. V.; Watson, J. M.; Maccotta, L.; McDermott, K. B.; Buckner, R. L.; Gilliam, F. G. & Ojemann, J. G. (2005), 'Evaluating functional MRI procedures for assessing hemispheric language dominance in neurosurgical patients', *Neuroradiology* 47(11), 835-844.
- Baddeley, A. (1992), 'Working memory', *Science* 255(5044), 556-559.
- Baddeley, A. D. & Hitch, G. (1974), Working Memory, in Gordon H. Bower (ed.). Academic Press, 47 - 89.
- Baddeley, A. D. & Wilson, B. (1985), 'Phonological coding and short-term memory in patients without speech'. *Journal of Memory and Language* 24(4), 490-502.
- Bahrlick, L. E. & Watson, J. S. (1985), 'Detection of intermodal proprioceptive-visual contingency as a potential basis of self-perception in infancy.', *Developmental Psychology* 21(6), 963.
- Bailly, G. (1986), 'Un modèle de congruence relationnel pour la synthèse de la prosodie du français', *Actes des 15es Journées d'Étude sur la Parole*, 75-78.
- Bain, A. (1855), *The senses and the intellect*, London: John W. Parker and Son, West Strand.
- Bakhtin, M. M. (1986), *Speech genres and other late essays*, Austin, TX: University of Texas Press.
- Baldo, J. V.; Dronkers, N. F.; Wilkins, D.; Ludy, C.; Raskin, P. & Kim, J. (2005), 'Is problem solving dependent on language?', *Brain and Language* 92(3), 240-250.
- Baldo, J. V.; Paulraj, S. R.; Curran, B. C. & Dronkers, N. F. (2015), 'Impaired reasoning and problem-solving in individuals with language impairment due to aphasia or language delay', *Frontiers in Psychology* 6, 1523.
- Ballet, G. (1886), *Le langage intérieur et les diverses formes de l'aphasie*, Alcan.
- Baron, F. (1999), *Le livre des Je t'aime*, www.lesjetaime.com.
- Baron-Cohen, S. (1995), *Mindblindness: An essay on autism and theory of mind*, MIT press.
- Barsalou, L. W. (1999), 'Perceptions of perceptual symbols', *Behavioral and Brain Sciences* 22(4), 637-660.
- Barsalou, L. W.; Simmons, W. K.; Barbey, A. K. & Wilson, C. D. (2003), 'Grounding conceptual knowledge in modality-specific systems', *Trends in Cognitive Sciences* 7(2), 84 - 91.
- Bates, E. (1990), 'Language about me and you: Pronominal reference and the emerging concept of self', *The self in transition: Infancy to childhood*, 165-182.
- Bates, E. & Dick, F. (2002), 'Language, gesture, and the developing brain', *Developmental psychobiology* 40(3), 293-310.
- Bauer, P. J. (2007), *Remembering the times of our lives: Memory in infancy and beyond*, Mahwah, NJ: Lawrence Erlbaum..
- Baum, S. R. & Pell, M. D. (1999), 'The neural bases of prosody: Insights from lesion studies and neuroimaging', *Aphasiology* 13(8), 581-608.
- Beavan, V.; Read, J. & Cartwright, C. (2011), 'The prevalence of voice-hearers in the general population: A literature review', *Journal of Mental Health* 20(3), 281-292.
- Beckman, M. E. (1986), *Stress and non-stress accent*, Vol. 7, Foris publications: Dordrecht, The Netherlands.
- Benveniste, E. (1966), *De la subjectivité dans le langage*, Paris, Gallimard, 258-266.
- Bergounioux, G. (2001a), 'Esquisse d'une histoire négative de l'endophasie [avec une attention presque exclusive pour les productions en langue française consacrées ? cette question]', *Langue française*, 3-25.
- Bergounioux, G. (2004), *Le moyen de parler*, Verdier, Paris.
- Bergounioux, G. (2014), *Dominos*, Champ Vallon, Seyssel.
- Bergson, H. (1889), *Essai sur les données immédiates de la conscience*, Félix Alcan, Paris, coll. Bibliothèque de philosophie contemporaine.
- Bergson, H. (1934), *La pensée et le mouvant. Essais et conférences*, Félix Alcan, Paris, coll. Bibliothèque de philosophie contemporaine.
- Bermúdez, L. J. (2001), 'Nonconceptual self-consciousness and cognitive science', *Synthese* 129(1), 129-149.
- Bernhardt, B. & Stoel-Gammon, C. (1994), 'Nonlinear phonology: Introduction and clinical application', *Journal of Speech, Language, and Hearing Research* 37(1), 123-143.
- Bernhardt, B. M.; Stemberger, J. P. & Bérubé, D. (2015), 'Phonology and fun-ology website: Free assessment tools, tutorials and activities', .
- Berwick, R. C. & Chomsky, N. (2011), *The biolinguistic program: The current state of its development*, Oxford University Press Oxford, 19-41.
- Best, C. (1995), 'A direct realist view of cross-language speech perception'. *Speech perception and linguistic experience: Issues in cross-language speech research*. Strange, W. (ed.), 171-206.

- Bickerton, D. (2002), 'Language in the modular mind? Its a no-brainer!', *Behavioral and Brain Sciences* 25(6), 677-678.
- Binet, A. (1886), *La psychologie du raisonnement, recherches expérimentales par l'hypnotisme*, Félix Alcan, éditeur, Paris.
- Blackmer, E. R. & Mitton, J. L. (1991), 'Theories of monitoring and the timing of repairs in spontaneous speech', *Cognition* 39(3), 173 - 194.
- Blakemore, S. J. (2003), 'Deluding the motor system', *Consciousness and Cognition* 12(4), 647-655.
- Blakemore, S.-J.; Wolpert, D. M. & Frith, C. D. (1998), 'Central cancellation of self-produced tickle sensation', *Nature neuroscience* 1(7), 635-640.
- Blakemore, S.-J.; Wolpert, D. M. & Frith, C. D. (2002), 'Abnormalities in the awareness of action', *Trends in cognitive sciences* 6(6), 237-242.
- Blank, H. & Davis, M. H. (2016), 'Prediction Errors but Not Sharpened Signals Simulate Multivoxel fMRI Patterns during Speech Perception', *PLOS Biology* 14(11), 1-32.
- Blank, S. C.; Scott, S. K.; Murphy, K.; Warburton, E. & Wise, R. J. S. (2002), 'Speech production: Wernicke, Broca and beyond', *Brain* 125(8), 1829-1838.
- Blanke, O. (2012), 'Multisensory brain mechanisms of bodily self-consciousness', *Nature Reviews Neuroscience* 13(8), 556-571.
- Bock, K. (1987), 'Exploring levels of processing in sentence production'. *Natural language generation*. Kempen, G. (ed.), Dordrecht: Martinus Nijhoff, 351-63.
- Boroditsky, L. (2011), 'How Language Shapes Thought', *Scientific American* 304(2), 62-65.
- Bowerman, M. & Levinson, S. C. (2001), *Language acquisition and conceptual development*, Cambridge University Press.
- Brachet, T. (2007), 'Entretien avec André Jacob', *Le Philosophoire*, 28, 47-65.
- Braitenberg, V. (1967), 'Is the cerebellar cortex a biological clock in the millisecond range', *Prog Brain Res* 25, 334-346.
- Brando de Carvalho, J.; Nguyen, N. & Wauquier, S. (2010), *Comprendre la phonologie*, Presses Universitaires de France, collection Linguistique Nouvelle, Paris.
- Bres, J. (2005), 'Savoir de quoi on parle: dialogue, dialogal, dialogique ; dialogisme, polyphonie?'. *Dialogisme, polyphonie : approches linguistiques*, Bres, J.; Haillet, P. P.; Mellet, S.; Nølke, H. & Rosier, L., ed., De Boeck-Duculot, 47-62.
- Bres, J. & Dufour, F. (2014), 'Du fonctionnement dialogique monologal des marqueurs dialogaux oui, non, si', *Arena Romanistica-Journal of Romance studies*(17), 58-69.
- Bretherton, I. & Beeghly, M. (1982), 'Talking about internal states: The acquisition of an explicit theory of mind.', *Developmental psychology* 18(6), 906.
- Broca, P. (1861), 'Remarques sur le siège de la faculté du langage articulé, suivies d'une observation d'aphémie (perte de la parole)', *Bulletin et Memoires de la Societe anatomique de Paris* 6, 330-357.
- Brosseau-Laprè, F.; Rvachew, S.; MacLeod, A. A. N.; Findlay, K.; Bérubé, D. & Bernhardt, B. M. (2018), 'Une vue d'ensemble: les données probantes sur le développement phonologique des enfants francophones canadiens An Overview of Data on the Phonological Development of French-Speaking Canadian Children'. *Canadian Journal of Speech-Language Pathology and Audiology*, 42 (1), 1-19.
- Browman, C. P. & Goldstein, L. (1989), 'Articulatory gestures as phonological units', *Phonology* 6(02), 201-251.
- Brown, R. & McNeill, D. (1966), 'The tip of the tongue phenomenon', *Journal of Verbal Learning and Verbal Behavior* 5(4), 325 - 337.
- Bruner, J. (1987), 'Life as Narrative', *Social Research* 54(1), 11-32.
- Buckner, R. L.; Krienen, F. M.; Castellanos, A.; Diaz, J. C. & Yeo, B. T. T. (2011), 'The organization of the human cerebellum estimated by intrinsic functional connectivity', *Journal of Neurophysiology* 106(5), 2322-2345.
- Caplan, D.; Alpert, N.; Waters, G. & Olivieri, A. (2000), 'Activation of Brocas area by syntactic processing under conditions of concurrent articulation', *Human Brain Mapping* 9(2), 65-71.
- Caplan, D. & Hanna, J. E. (1998), 'Sentence production by aphasic patients in a constrained task', *Brain and Language* 63(2), 184-218.
- Caron, J. (1989), *Précis de psycholinguistique*, Presses Universitaires de France.
- Carruthers, P. (2002), 'The cognitive functions of language', *Behavioral and Brain Sciences* 25(6), 657-674.
- Carruthers, P. (2009), 'How we know our own minds: The relationship between mindreading and metacognition', *Behavioral and brain sciences* 32(02), 121-138.
- Carruthers, P. (2016), 'Two Systems for Mindreading?', *Review of Philosophy and Psychology* 7(1), 141-162.
- Carruthers, P. (), 'The causes and contents of inner speech', , 31-52.

- Cary, P. (2011), 'The Inner Word Prior to Language', *Philosophy today*, 192-198.
- Charcot, J. M. (1880-93), *Œuvres Complètes*, Paris, Delahaye et Lecrosnier.
- Charcot, J. M. (1889), *Leçons du mardi à la Salpêtrière. Notes de cours de MM. Blin, Charcot, Colin*, Paris, Lecrosnier et Babé.
- Chomsky, N. (1969), *La linguistique cartésienne: un chapitre de l'histoire de la pensée rationaliste; suivi de La nature formelle du langage*, Seuil, Paris.
- Chomsky, N. (1977), *Réflexions sur le langage*, trad. J.C. Milner, B. Vautherin et P. Fiala, Maspero, Paris.
- Chomsky, N. (1995), *The minimalist program*, MIT press.
- Chomsky, N. (2002), *On nature and language*, Cambridge University Press.
- Chomsky, N. (2014), 'Language Architecture and its Import for Evolution'.
- Chouinard, M. M. & Clark, E. V. (2003), 'Adult reformulations of child errors as negative evidence', *Journal of child language* 30(3), 637-669.
- Christoffels, I. K., van de Ven, V., Waldorp, L. J., Formisano, E., & Schiller, N. O. (2011). 'The sensory consequences of speaking: parametric neural cancellation during speech in auditory cortex.' *PLoS One*, 6(5), e18307.
- Clark, A. (1998), 'Magic words: How language augments human computation', *Language and thought: Interdisciplinary themes*, 162-183.
- Clark, A. (2002), 'Anchors not inner codes, coordination not translation (and hold the modules please)', *Behavioral and Brain Sciences* 25(6), 681-681.
- Cleary, M. & Pisoni, D. B. (2001), 'Speech perception and spoken word recognition: research and theory', *Blackwell Handbook of Sensation and Perception*, 499-534.
- Clerc, O. (2018), 'Neurophysiological correlates of deliberate inner speech and verbal mind wandering'.
- Clowes, R. (2007), 'A self-regulation model of inner speech and its role in the organisation of human conscious experience', *Journal of Consciousness Studies*, 14 (7), 59-71.
- Collins, A. M. & Loftus, E. F. (1975), 'A spreading-activation theory of semantic processing.', *Psychological review* 82(6), 407.
- Collins, A. M. & Quillian, M. R. (1969), 'Retrieval time from semantic memory', *Journal of Verbal Learning and Verbal Behavior* 8(2), 240 - 247.
- Coltheart, M. (1999), 'Modularity and cognition', *Trends in Cognitive Sciences* 3(3), 115 - 120.
- Conrad, R. (1971), 'The chronology of the development of covert speech in children.', *Developmental Psychology* 71.
- Conway, M. A. (2005), 'Memory and the self', *Journal of Memory and Language* 53(4), 594 - 628.
- Conway, M. A. & Rubin, D. C. (1993), 'The structure of autobiographical memory', *Theories of memory* 103, 137.
- Conway, M. A.; Singer, J. A. & Tagini, A. (2004), 'The self and autobiographical memory: Correspondence and coherence', *Social cognition* 22(5: Special issue), 491-529.
- Cooley, C. H. (1922), *Human nature and the social order*, New Brunswick, NJ: Transaction Publishers, new edition 1992.
- Corley, M.; Brocklehurst, P. H. & Moat, H. S. (2011), 'Error biases in inner and overt speech: Evidence from tongue twisters.', *Journal of Experimental Psychology: Learning, Memory, and Cognition* 37(1), 162.
- Cosmides, L. & Tooby, J. (1992), 'Cognitive adaptations for social exchange', *The adapted mind: Evolutionary psychology and the generation of culture* 163, 163-228.
- Craik, F. I. M. & Lockhart, R. S. (1972), 'Levels of processing: A framework for memory research', *Journal of Verbal Learning and Verbal Behavior* 11(6), 671 - 684.
- Cummins, R. (1975), 'Functional Analysis', *The Journal of Philosophy* 72, 741-765.
- Cutler, A. (1983), 'Speakers Conceptions of the Function of Prosody'. *Prosody: Models and Measurements*. Cutler, A. & Ladd, D. R. (ed.), Springer Berlin Heidelberg, Berlin, Heidelberg, 79-91.
- Dale, R. & Spivey, M. (2002), 'A linguistic module for integrating the senses, or a house of cards?', *Behavioral and Brain Sciences* 25(6), 681-682.
- Davidson, D. (1975), 'Thought and talk', *Mind and language* 1975, 7-23.
- Davis, B. L.; MacNeilage, P. F. & Matyear, C. L. (2002), 'Acquisition of serial complexity in speech production: A comparison of phonetic and phonological approaches to first word production', *Phonetica* 59(2-3), 75-107.
- De Saussure, F. (1995), *Cours de linguistique générale*, Payot, Paris.
- De Saussure, F. (2002), *Écrits de linguistique générale*, Gallimard, Paris.
- De Zeeuw, C. I. & Ten Brinke, M. M. (2015), 'Motor Learning and the Cerebellum', *Cold Spring Harbor Perspectives in Biology* 7(9).

- Deacon, T. W. (1997), *The symbolic species: The co-evolution of language and the brain*, WW Norton & Company, New York, NY.
- Decety, J. (2005), 'Perspective taking as the royal avenue to empathy', *Other minds: How humans bridge the divide between self and others*, 143-157.
- Decety, J. & Grèzes, J. (2006), 'The power of simulation: Imagining ones own and others behavior', *Brain Research* 1079(1), 4 - 14.
- Decety, J. & Jackson, P. L. (2004), 'The Functional Architecture of Human Empathy', *Behavioral and Cognitive Neuroscience Reviews* 3(2), 71-100.
- Dehaene, S. (1997), *The number sense*, Oxford University Press.
- Dehaene, S. & Naccache, L. (2001), 'Towards a cognitive neuroscience of consciousness: basic evidence and a workspace framework', *Cognition* 79(1), 1 - 37.
- Dell, F. (1984), 'L'accentuation dans les phrases en français', *Forme sonore du langage*, 65-122.
- Dell, G. S. (1986), 'A spreading-activation theory of retrieval in sentence production.', *Psychological review* 93(3), 283.
- Dell, G. S. (2013), 'Cascading and feedback in interactive models of production: A reflection of forward modeling?', *Behavioral and Brain Sciences* 36(4), 23-24.
- Dell, G. S., & Oppenheim, G. M. (2015), 'Insights for Speech Production Planning from Errors in Inner Speech'. *The Handbook of Speech Production*, Redford, M. (Ed.), John Wiley & Sons, West Sussex, UK, 404-418.
- Dell, G. S., & Repka, R. J. (1992), 'Errors in inner speech'. *Experimental Slips and Human Error*. B. J. Baars (ed.), Springer, NY, USA, 237-262.
- Dell, G. S.; Schwartz, M. F.; Martin, N.; Saffran, E. M. & Gagnon, D. A. (1997), 'Lexical access in aphasic and nonaphasic speakers.', *Psychological review* 104(4), 801.
- Dennett, D. (1991), *Consciousness explained*, New York: Little Brown & Co.
- Dennett, D. C. (1988), 'Why everyone is a novelist', *The Times Literary Supplement*(4459), 1016.
- Derrida, J. (1972), *La dissémination*, Le Seuil, Paris.
- Descartes, R. (1646-1649), *Œuvres complètes*, Vrin, Paris.
- Descartes, R. (1637), *Discours de la méthode*, Edition Jan Maire, Leyde.
- Descartes, R. (1646), *Lettre au marquis de Newcastle du 23 novembre 1646*, La Pléiade, Gallimard, 1953, 1254-1256.
- Desmurget, M.; Reilly, K. T.; Richard, N.; Szathmari, A.; Mottolose, C. & Sirigu, A. (2009), 'Movement Intention After Parietal Cortex Stimulation in Humans', *Science* 324(5928), 811-813.
- Di Cristo, A. (2000), 'Vers une modélisation de l'accentuation du français (seconde partie)', *Journal of French language studies* 10(1), 27-44.
- Diaz, R. & Berk, L. (1992), *Private speech: From social interaction to self-regulation*, Lawrence Erlbaum: Hillsdale, NJ.
- Diedrichsen, J. & Zotow, E. (2015), 'Surface-Based Display of Volume-Averaged Cerebellar Imaging Data', *PLOS ONE* 10(7), 1-18.
- Diehl, R. L.; Lotto, A. J. & Holt, L. L. (2004), 'Speech perception', *Psychology* 55(1), 149-179.
- Dohen, M. & Lœvenbruck, H. (2004), Pre-focal rephrasing, focal enhancement and postfocal deaccentuation in French, in 'Eighth International Conference on Spoken Language Processing, Jeju, Korea, 4-8 October 20014', 785-788.
- Dohen, M. & Lœvenbruck, H. (2005), Audiovisual Production and Perception of Contrastive Focus in French: a multispeaker study, in 'Ninth International Conference on Speech Communication and Technology, September 4-8, Lisboa, Portugal', 2413-2416.
- Dohen, M. & Lœvenbruck, H. (2009), 'Interaction of audition and vision for the perception of prosodic contrastive focus', *Language and Speech* 52(2-3), 177-206.
- Dohen, M.; Lœvenbruck, H.; Roustan, B. & Vilain, C. (2011), 'Pointage manuel et vocal: corrélats sensorimoteurs et neurocognitifs, coordination geste manuel/parole', *Faits de langue* 37, 97-116.
- Droege, P. (2013), 'Memory and consciousness', *Philosophia Scientiæ* 17(2), 171-193.
- Dronkers, N. (1987), Chelsea s cognitive and neuropsychological status, in 'Orton Dyslexia Society Annual Meeting, San Francisco, CA'.
- Dronkers, N. F.; Ludy, C. A. & Redfern, B. B. (1998), 'Pragmatics in the absence of verbal language: Descriptions of a severe aphasic and a language-deprived adult', *Journal of Neurolinguistics* 11(1), 179 - 190.
- Duffau, H. (2008), 'The anatomo-functional connectivity of language revisited: New insights provided by electrostimulation and tractography', *Neuropsychologia* 46(4), 927 - 934.

- Duffau, H. (2013), 'Nouveautés dans la chirurgie des gliomes cérébraux: vers un acte personnalisé', *e-mémoires de l'Académie Nationale de Chirurgie* 12(1), 019-024.
- Duffau, H.; Moritz-Gasser, S. & Mandonnet, E. (2014), 'A re-examination of neural basis of language processing: proposal of a dynamic hodotopical model from data provided by brain stimulation mapping during picture naming', *Brain and language* 131, 1-10.
- Dujardin, E. (1887), *Les lauriers sont coupés*, Revue indépendante.
- Duval, C. (2012), 'Mise en évidence d'une activité myoélectrique labiale pendant la parole intérieure et les hallucinations auditives verbales : apport de la relaxation'.
- D Imperio, M.; Espesser, R.; Loevenbruck, H.; Menezes, C.; Nguyen, N. & Welby, P. (2007), *Are tones aligned with articulatory events? Evidence from Italian and French*, Mouton de Gruyter, 577-608.
- Egger, V. (1881), *La parole intérieure: essai de psychologie descriptive*, G. Baillièrre et Cie, Paris.
- Emerson, M. J. & Miyake, A. (2003), 'The role of inner speech in task switching: A dual-task investigation', *Journal of Memory and Language* 48(1), 148-168.
- Epting, K. L. & Critchfield, T. S. (2006), 'Self-Editing: On the Relation Between behavioral and Psycholinguistic Approaches', *The Behavior Analyst* 29(2), 211-234.
- Evans, J. S. B. T. (2003), 'In two minds: dual-process accounts of reasoning', *Trends in Cognitive Sciences* 7(10), 454 - 459.
- Evans, J. S. B. T. & Frankish, K. E. (2009), *In two minds: Dual processes and beyond.*, Oxford University Press.
- Evans, J. S. B. T. & Over, D. E. (1996), *Rationality and reasoning*, Psychology Press.
- Fadiga, L.; Craighero, L.; Buccino, G. & Rizzolatti, G. (2002), 'Speech listening specifically modulates the excitability of tongue muscles: a TMS study', *European Journal of Neuroscience* 15(2), 399-402.
- Farrer, C. & Frith, C. D. (2002), 'Experiencing oneself vs another person as being the cause of an action: the neural correlates of the experience of agency', *Neuroimage* 15(3), 596-603.
- Fauconnier, G. (1985), *Mental spaces: Aspects of meaning construction in natural language*, MIT Press.
- Feinberg, I. (1978), 'Efference copy and corollary discharge: implications for thinking and its disorders.', *Schizophrenia Bulletin* 4(4), 636-640.
- Fernyhough, C. (2004), 'Alien voices and inner dialogue: towards a developmental account of auditory verbal hallucinations', *New Ideas in Psychology* 22(1), 49-68.
- Fernyhough, C. (2012), 'Pieces of light : The New Science of Memory', *London: Profile*.
- Fernyhough, C. (2016), *The Voices Within: The history and science of how we talk to ourselves*, Profile Books/Wellcome Collection, UK.
- Fivush, R. (2011), 'The Development of Autobiographical Memory', *Annual Review of Psychology* 62(1), 559-582.
- Flinker, A.; Korzeniewska, A.; Shestiyuk, A. Y.; Franaszczuk, P. J.; Dronkers, N. F.; Knight, R. T. & Crone, N. E. (2015), 'Redefining the role of Broca's area in speech', *Proceedings of the National Academy of Sciences* 112(9), 2871-2875.
- Fodor, J. A. (1983), *The modularity of mind*, MIT press.
- Fodor, J. A. & Pylyshyn, Z. W. (2015), *Minds without meanings: An essay on the content of concepts*, MIT Press, Cambridge, MA.
- Fodor, J.; Bever, A. & Garrett, T. G. (1974), 'The psychology of language: An introduction to psycholinguistics and generative grammar', .
- Ford, J. M. & Mathalon, D. H. (2004), 'Electrophysiological evidence of corollary discharge dysfunction in schizophrenia during talking and thinking', *Journal of psychiatric research* 38(1), 37-46.
- Forster, K. I. (1976), 'Accessing the mental lexicon'. *New approaches to language mechanisms*. Wales, R. J. & Walker, E. W., ed., Amsterdam: North-Holland.
- Fowler, C. A. (1986), 'An event approach to the study of speech perception from a direct-realist perspective', *Journal of Phonetics* 14, 3-28.
- Franck, N. (2006), *Schizophrénie (La): La reconnaître et la soigner*, Odile Jacob.
- Franck, N. & Thibaut, F. (2003), 'Les Hallucinations', *Encyclopédie médico-chirurgicale. Psychiatrie*, 37-120.
- Francois, J. & Nespoulous, J.-L. (2011), 'L'architecture des processus de production et de réception: aspects (neuro) psycholinguistiques', *Mémoires de la Société de Linguistique de Paris, l'architecture des théories linguistiques, les modules et leurs interfaces, Nouvelle Série* 20, 205-239.
- Frankish, K. (1998), 'Natural language and virtual belief', *Language and thought: interdisciplinary themes*, 248.
- Frankish, K. (2010), 'Dual-process and dual-system theories of reasoning', *Philosophy Compass* 5(10), 914-926.

- Frazier, L. & Rayner, K. (1982), 'Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences', *Cognitive Psychology* 14(2), 178 - 210.
- Freud, S. & Breuer, J. (1895/1956), *Studien über Hysterie. Études sur l'hystérie*, Vienne, 1895. Paris, Presses universitaires de France.
- Friederici, A. D. (2002), 'Towards a neural basis of auditory sentence processing', *Trends in cognitive sciences* 6(2), 78-84.
- Friederici, A. D. (2012), 'The cortical language circuit: from auditory perception to sentence comprehension', *Trends in Cognitive Sciences* 16(5), 262 - 268.
- Frith, C. (1992), *The cognitive neuropsychology of schizophrenia*, Psychology Press.
- Frith, C. (2002), 'Attention to action and awareness of other minds', *Consciousness and Cognition* 11(4), 481-487.
- Frith, C. (2005), 'The self in action: Lessons from delusions of control', *Consciousness and cognition* 14(4), 752-770.
- Frith, C. (2010), 'What is consciousness for?', *Pragmatics & Cognition* 18(3), 497-551.
- Frith, C. (2012), 'Explaining delusions of control: The comparator model 20 years on', *Consciousness and cognition* 21(1), 52-54.
- Frith, C. D.; Blakemore, S. J. & Wolpert, D. M. (2000), 'Explaining the symptoms of schizophrenia: abnormalities in the awareness of action', *Brain Research Reviews* 31(2-3), 357-363.
- Frith, C. D. & Done, D. J. (1989), 'Experiences of alien control in schizophrenia reflect a disorder in the central monitoring of action', *Psychological Medicine* 19(02), 359-363.
- Frith, C. D.; Friston, K. J.; Herold, S.; Silbersweig, D.; Fletcher, P.; Cahill, C.; Dolan, R. J.; Frackowiak, R. S. & Liddle, P. F. (1995), 'Regional brain activity in chronic schizophrenic patients during the performance of a verbal fluency task.', *The British Journal of Psychiatry* 167(3), 343-349.
- Frith, C. D. (1987), 'The positive and negative symptoms of schizophrenia reflect impairments in the perception and initiation of action', *Psychological medicine* 17(03), 631-648.
- Fromkin, V. (1980), *Errors in linguistic performance: Slips of the tongue, ear, pen, and hand*, Academic Press.
- Fromkin, V. A. (1971), 'The Non-Anomalous Nature of Anomalous Utterances', *Language* 47(1), 27-52.
- Fromkin, V. A.; Rodman, R. & Hyams, N. M. (1983), *An introduction to language*, Holt.
- Gallagher, S. (2000), 'Philosophical conceptions of the self: implications for cognitive science', *Trends in cognitive sciences* 4(1), 14-21.
- Gallagher, S. (2004), 'Neurocognitive models of schizophrenia: a neurophenomenological critique', *Psychopathology* 37(1), 8-19.
- Gallup, G. G. (1977), 'Self recognition in primates: A comparative approach to the bidirectional properties of consciousness.', *American psychologist* 32(5), 329.
- Gardiner, A. H. (1927), *Egyptian Grammar, being an Introduction to the study of Hieroglyphs*, Oxford, Clarendon Press.
- Garnier, M.; Dohen, M.; Loevenbruck, H.; Welby, P. & Bailly, L. (2006), 'The Lombard Effect: a physiological reflex or a controlled intelligibility enhancement?', in '7th International Seminar on Speech Production', 255-262.
- Garrett, M. F. (1980), 'The limits of accommodation: Arguments for independent processing levels in sentence production', *Errors in linguistic performance: Slips of the tongue, ear, pen, and hand*, 263-271.
- Gauker, C. (2002), 'No conceptual thought without language', *Behavioral and Brain Sciences* 25(6), 687-687.
- Gauker, C. (2005), 'On the evidence for prelinguistic concepts', *THEORIA. Revista de Teoria, Historia y Fundamentos de la Ciencia* 20(3), 287-297.
- Gauker, C. (2011), *Words and images: An essay on the origin of ideas*, OUP Oxford.
- Gernsbacher, M. A. & Kaschak, M. P. (2003), 'Neuroimaging studies of language production and comprehension', *Annual review of psychology* 54(1), 91-114.
- Geva, S.; Bennett, S.; Warburton, E. A. & Patterson, K. (2011). 'Discrepancy between inner and overt speech: Implications for post-stroke aphasia and normal language processing'. *Aphasiology* 25, 323-343.
- Gick, B. (2015), 'Speech production and articulatory phonetics', *Proceedings of the 18th ICPHS, Glasgow*.
- Glusker, P. (1987), Chelsea, in 'Orton Dyslexia Society Annual Meeting'.
- Gomi, H.; Shidara, M.; Takemura, A.; Inoue, Y.; Kawano, K. & Kawato, M. (1998), 'Temporal Firing Patterns of Purkinje Cells in the Cerebellar Ventral Paraflocculus During Ocular Following Responses in Monkeys I. Simple Spikes', *Journal of Neurophysiology* 80(2), 818-831.

- Gopnik, A.; Choi, S. & Baumberger, T. (1996), 'Cross-linguistic differences in early semantic and cognitive development', *Cognitive Development* 11(2), 197 - 225.
- Grafton, S. T.; Aziz-Zadeh, L. & Ivry, R. B. (2009), 'Relative hierarchies and the representation of action', *The cognitive neurosciences* 4, 641-652.
- Grafton, S. T. & de C. Hamilton, A. F. (2007), 'Evidence for a distributed hierarchy of action representation in the brain', *Human Movement Science* 26(4), 590 - 616.
- Grandchamp, R.; Rapin, L.; Perrone-Bertolotti, M.; Pichat, C.; Haldin, C.; Dohen, M.; Perrier, P. and Lachaux, J. P.; Garnier, M.; Baciú, M. & Lœvenbruck, H. (soumis), 'The Condialint Model: A neurocognitive account of varieties of inner speech along condensation, dialogicality and intentionality dimensions.', *Special Issue Frontiers in psychology, Cognitive Science*.
- Grandon, B.; Vilain, A.; Lœvenbruck, H.; Schmerber, S. & Truy, E. (2017), 'Realisation of voicing by French-speaking CI children after long-term implant use: An acoustic study', *Clinical Linguistics & Phonetics* 31(7-9), 598-611.
- Grappe, A.; Sarma, S. V.; Sacré, P.; González-Martínez, J.; Liégeois-Chauvel, C. & Alario, F.-X. (2018), 'An intracerebral exploration of functional connectivity during word production', *Journal of Computational Neuroscience*.
- Greenwald, A. G. (1980), 'The totalitarian ego: Fabrication and revision of personal history.', *American psychologist* 35(7), 603.
- Gregoromichelaki, E.; Kempson, R.; Purver, M.; Mills, G. J.; Cann, R.; Meyer-Viol, W. & Healey, P. G. T. (2011), 'Incrementality and intention-recognition in utterance processing.', *Dialogue & Discourse* 2(1), 199-233.
- Grush, R. (2004), 'The emulation theory of representation: motor control, imagery, and perception', *Behavioral and brain sciences* 27(3), 377-396.
- Guenther, F. H.; Ghosh, S. S. & Tourville, J. A. (2006), 'Neural modeling and imaging of the cortical interactions underlying syllable production', *Brain and language* 96(3), 280-301.
- Guenther, F. H. & Perkell, J. S. (2004), 'A neural model of speech production and its application to studies of the role of auditory feedback in speech', *Speech motor control in normal and disordered speech*, 29-49.
- Guenther, F. H.; Hampson, M. & Johnson, D. (1998), 'A theoretical investigation of reference frames for the planning of speech movements.', *Psychological review* 105(4), 611.
- Guenther, F. H. & Vladusich, T. (2012), 'A neural theory of speech acquisition and production', *Journal of Neurolinguistics* 25(5), 408 - 422.
- Hadamard, J. (1945), *An essay on the psychology of invention in the mathematical field*, Princeton University Press.
- Hage, S. R. & Nieder, A. (2016), 'Dual Neural Network Model for the Evolution of Speech and Language', *Trends in Neurosciences* 39(12), 813 - 829.
- Hagège, C. (1985), 'L'homme de paroles. Contribution linguistique aux sciences humaines', Paris, Fayard.
- Haggard, P.; Newman, C. & Magno, E. (1999), 'On the perceived time of voluntary actions', *British Journal of Psychology* 90(2), 291-303.
- Hagoort, P. (2003), 'How the brain solves the binding problem for language: a neurocomputational model of syntactic processing', *NeuroImage* 20, S18 - S29.
- Haller, S.; Radue, E. W.; Erb, M.; Grodd, W. & Kircher, T. (2005), 'Overt sentence production in event-related fMRI', *Neuropsychologia* 43(5), 807 - 814.
- Hampton, J. (2002), 'Languages role in enabling abstract, logical thought', *Behavioral and Brain Sciences* 25(6), 688.
- Hardy, J. (2006), 'Speaking clearly: A critical review of the self-talk literature', *Psychology of Sport and Exercise* 7(1), 81-97.
- Hartsuiker, R. J. (2013), 'Are forward models enough to explain self-monitoring? Insights from patients and eye movements', *Behavioral and Brain Sciences* 36(4), 357 - 358.
- Hartsuiker, R. J. & Kolk, H. H. J. (2001), 'Error Monitoring in Speech Production: A Computational Test of the Perceptual Loop Theory', *Cognitive Psychology* 42(2), 113 - 157.
- Haruno, M.; Wolpert, D. M. & Kawato, M. (2003), 'Hierarchical MOSAIC for movement generation', *International Congress Series* 1250, 575 - 590.
- Heavey, C. L. & Hurlburt, R. T. (2008), 'The phenomena of inner experience', *Consciousness and cognition* 17(3), 798-810.
- Hécaen, H. & Dubois, J. (1969), *La naissance de la neuropsychologie du langage (1825-1865)*, Flammarion, Paris.

- Hegel, G. W. F. (1817/1867), *Précis de l'encyclopédie des sciences philosophiques : la logique, la philosophie de la nature, la philosophie de l'esprit*, Traduction de A. Véra, Paris: Germer Baillière.
- Hickmann, M. (2001), 'Language and cognition in development', *Pragmatics. Quarterly Publication of the International Pragmatics Association (IPrA)* 11(2), 105-126.
- Hickok, G. (2009), 'The functional neuroanatomy of language', *Physics of Life Reviews* 6(3), 121 - 143.
- Hickok, G. (2012), 'Computational neuroanatomy of speech production', *Nature Reviews Neuroscience* 13(2), 135-145.
- Hickok, G.; Houde, J. & Rong, F. (2011), 'Sensorimotor Integration in Speech Processing: Computational Basis and Neural Organization', *Neuron* 69(3), 407-422.
- Hickok, G. & Poeppel, D. (2015), 'Chapter 8 - Neural basis of speech perception', in Michael J. Aminoff; François Boller & Dick F. Swaab (ed.), 'The Human Auditory System', Elsevier, 149 - 160.
- Hickok, G. & Poeppel, D. (2007), 'The cortical organization of speech processing', *Nature Reviews Neuroscience* 8(5), 393-402.
- Hirst, D. & Di Cristo, A. (1998), *Intonation systems: a survey of twenty languages*, Cambridge University Press.
- Hjelmslev, L. (1943/1971), *Prolégomènes a une théorie du langage*, Editions de Minuit, Paris.
- Hoffman, E. (1989), *Lost in translation: A life in a new language.*, New York: Penguin Books.
- von Holst, E. & Mittelstaedt, H. (1950), 'Das Reafferenzprinzip: Wechselwirkungen zwischen Zentralnervensystem und Peripherie', *Naturwissenschaften* 37, 464-476.
- Houde, J. F. & Nagarajan, S. S. (2011), 'Speech production as state feedback control', *Frontiers in Human Neuroscience* 5, 82.
- Houde, J. F. & Jordan, M. I. (1998), 'Sensorimotor adaptation in speech production', *Science* 279(5354), 1213-1216.
- Houde, J. F.; Nagarajan, S. S.; Sekihara, K. & Merzenich, M. M. (2002), 'Modulation of the Auditory Cortex during Speech: An MEG Study', *Journal of Cognitive Neuroscience* 14(8), 1125-1138.
- Huettig, F. & Hartsuiker, R. J. (2010), 'Listening to yourself is like listening to others: External, but not internal, verbal self-monitoring is based on speech perception', *Language and Cognitive Processes* 25(3), 347-374.
- Hume, D. (1739), *A Treatise of Human Nature, Book 1, Of the understanding*, London: John Noon.
- Hurlburt, R. T. (1990), *Sampling normal and schizophrenic inner experience*, Plenum press.
- Hurlburt, R. T. (2011), *Investigating pristine inner experience: Moments of truth*, Cambridge University Press.
- Hurlburt, R. T. & Heavey, C. L. (2015), 'Investigating pristine inner experience: Implications for experience sampling and questionnaires', *Consciousness and Cognition* 31(0), 148 - 159.
- Hurlburt, R. T.; Heavey, C. L. & Kelsey, J. M. (2013), 'Toward a phenomenology of inner speaking', *Consciousness and cognition* 22(4), 1477-1494.
- Imamizu, H. & Kawato, M. (2009), 'Brain mechanisms for predictive control by switching internal models: implications for higher-order cognitive functions', *Psychological Research PRPF* 73(4), 527-544.
- Indefrey, P.; Brown, C. M.; Hellwig, F.; Amunts, K.; Herzog, H.; Seitz, R. J. & Hagoort, P. (2001), 'A neural correlate of syntactic encoding during speech production', *Proceedings of the National Academy of Sciences* 98(10), 5933-5936.
- Indefrey, P. & Levelt, W. J. M. (2004), 'The spatial and temporal signatures of word production components', *Cognition* 92(1), 101-144.
- Ito, M. (2008), 'Control of mental activities by internal models in the cerebellum', *Nature Reviews Neuroscience* 9(4), 304.
- Ivry, R. B. & Spencer, R. M. C. (2004), 'The neural representation of time', *Current Opinion in Neurobiology* 14(2), 225 - 232.
- Ivry, R. B.; Spencer, R. M.; Zelaznik, H. N. & Diedrichsen, J. (2002), 'The Cerebellum and Event Timing', *Annals of the New York Academy of Sciences* 978(1), 302-317.
- Jackendoff, R. (1996), 'How language helps us think', *Pragmatics & Cognition* 4(1), 1-34.
- Jackendoff, R. (2002), *Foundations of language: brain, meaning, grammar, evolution*, Oxford, UK: Oxford University Press.
- Jackendoff, R. (2011), 'What is the human language faculty?: Two views', *Language* 87(3), 586-624.
- Jacob, A. (1967/1992), *Temps et Langage. Essai sur les structures du sujet parlant*, Paris : A. Colin.
- Jakobson, R. (1941), 'Kindersprache, Aphasie, und allgemeine Lautgesetze. Uppsala', *przedruk w Selected writings I, Translated by R. Keiler. Child Language, Aphasia, and Phonological Universals. The Hague: Mouton (1968)*, 328-401.

- Jakobson, R. (1963), *Essais de linguistique générale*, traduit et préfacé par Nicolas Ruwet, éditions de Minuit, Paris.
- Jeannerod, M. (1984), 'The Timing of Natural Prehension Movements', *Journal of Motor Behavior* 16(3), 235-254.
- Jeannerod, M. (2001), 'Neural simulation of action: a unifying mechanism for motor cognition', *Neuroimage* 14, S103-S109.
- Jeannerod, M. & Pacherie, E. (2004), 'Agency, simulation and self-identification', *Mind & language* 19(2), 113-146.
- Johnson, K. & Mullennix, J. W. (1997), *Talker variability in speech processing*, San Diego, CA: Academic Press.
- John-Steiner, V. (1985/1997), *Notebooks of the mind: Explorations of thinking*, Oxford University Press.
- Jones, S. R. (2010), 'Do we need multiple models of auditory verbal hallucinations? Examining the phenomenological fit of cognitive and neurological models', *Schizophrenia bulletin* 36(3), 566.
- Jones, S. R. & Fernyhough, C. (2007a), 'Thought as action: Inner speech, self-monitoring, and auditory verbal hallucinations', *Consciousness and cognition* 16(2), 391-399.
- Jones, S. R. & Fernyhough, C. (2007b), 'Neural correlates of inner speech and auditory verbal hallucinations: A critical review and theoretical integration', *Clinical psychology review* 27(2), 140-154.
- Jordan, M. I. & Rumelhart, D. E. (1992), 'Forward models: Supervised learning with a distal teacher', *Cognitive science* 16(3), 307-354.
- Jun, S. A. & Fougeron, C. (2000), 'A phonological model of French intonation', *Intonation: Analysis, modeling and technology*, 209-242.
- Jun, S. A. & Fougeron, C. (2002), 'The Realizations of the Accentual Phrase in French Intonation', *Probus, a special issue on Intonation in the Romance Languages* 14, 147-172.
- Junqua, J.-C. (1993), 'The Lombard reflex and its role on human listeners and automatic speech recognizers', *The Journal of the Acoustical Society of America* 93(1), 510-524.
- Kagan, J. (1981), 'The second year: The emergence of self-awareness', .
- Kahneman, D. (2011), *Thinking, fast and slow*, Farrar, Straus and Giroux: New York.
- Kahneman, D.; Paul, S. & Tversky, A. (1982), *Judgment under uncertainty: Heuristics and biases*, Cambridge university press.
- Kalman, R. E. (1960), 'A new approach to linear filtering and prediction problems', *Trans. ASME, Journal of Basic Engineering* 82(1), 35-45.
- Kapur, A.; Kapur, S. & Maes, P. (2018), 'AlterEgo: A Personalized Wearable Silent Speech Interface', *23rd International Conference on Intelligent User Interfaces*, 43-53.
- Kawato, M.; Furukawa, K. & Suzuki, R. (1987), 'A hierarchical neural-network model for control and learning of voluntary movement', *Biological cybernetics* 57(3), 169-185.
- Kawato, M.; Isobe, M.; Maeda, Y. & Suzuki, R. (1988), 'Coordinates transformation and learning control for visually-guided voluntary movement with iteration: A Newton-like method in a function space', *Biological Cybernetics* 59(3), 161-177.
- Kawato, M.; Maeda, Y.; Uno, Y. & Suzuki, R. (1990), 'Trajectory formation of arm movement by cascade neural network model based on minimum torque-change criterion', *Biological Cybernetics* 62(4), 275-288.
- Keele, S. W.; Cohen, A. & Ivry, R. (1990), 'Motor programs: Concepts and issues', *Attention and performance: Vol. 13. Motor representation and control*, M. Jeannerod (ed.), Lawrence Erlbaum, Hillsdale, NJ, 77-110.
- Kell, C. A.; Darquea, M.; Behrens, M.; Cordani, L.; Keller, C. & Fuchs, S. (2017), 'Phonetic detail and lateralization of reading-related inner speech and of auditory and somatosensory feedback processing during overt reading', *Human Brain Mapping* 38(1), 493-508.
- Kell, C. A.; Giraud, A.-L.; Morillon, B. & Kouneiher, F. (2011), 'Lateralization of Speech Production Starts in Sensory Cortices? A Possible Sensory Origin of Cerebral Left Dominance for Speech', *Cerebral Cortex* 21(4), 932-937.
- Keller, C. & Kell, C. A. (2016), 'Asymmetric intra- and interhemispheric interactions during covert and overt sentence reading', *Neuropsychologia* 93, 448 - 465.
- Keller, H. (1908), *The world I live in*, New York Review of Books.
- Kempen, G. & Hoenkamp, E. (1987), 'An Incremental Procedural Grammar for Sentence Formulation', *Cognitive Science* 11(2), 201-258.
- Kervajan, L. (2006), Problèmes de représentation de la Langue des Signes Française en vue du Traitement Automatique, in 'Verbum ex machina (TALN vol. 2): Actes de la 13e conférence sur le traitement automatique des langues naturelles', 670.

- Kim, H. S. (2002), 'We talk, therefore we think? A cultural analysis of the effect of talking on thinking.', *Journal of personality and social psychology* 83(4), 828.
- Klinger, E. & Cox, W. M. (1987), 'Dimensions of thought flow in everyday life', *Imagination, Cognition and Personality* 7(2), 105-128.
- Knoblich, G. & Flach, R. (2001), 'Predicting the Effects of Actions: Interactions of Perception and Action', *Psychological Science* 12(6), 467-472.
- Knoblich, G.; Seigerschmidt, E.; Flach, R. & Prinz, W. (2002), 'Authorship effects in the prediction of handwriting strokes: Evidence for action simulation during action perception', *The Quarterly Journal of Experimental Psychology Section A* 55(3), 1027-1046.
- Korba, R. J. (1990), 'The Rate of Inner Speech', *Perceptual and Motor Skills* 71(3), 1043-1052.
- Kosslyn, S. M. (2005), 'Mental images and the brain', *Cognitive Neuropsychology* 22(3-4), 333-347.
- von Kriegstein, K.; Dogan, Ş.; Grüter, M.; Giraud, A.-L.; Kell, C. A.; Grüter, T.; Kleinschmidt, A. & Kiebel, S. J. (2008), 'Simulation of talking faces in the human brain improves auditory speech recognition', *Proceedings of the National Academy of Sciences* 105(18), 6747-6752.
- Lacan, J. (1966), *Écrits*. Paris, Seuil.
- Ladefoged, P. (2003), 'Preserving the sounds of disappearing languages', <https://linguistics.ucla.edu/people/ladefoged/Preserving%20sounds.pdf>.
- Ladefoged, P. & Maddieson, I. (1996), *The sounds of the world's languages*, Vol. 1012, Blackwell Oxford.
- Lakoff, G. & Johnson, M. (1999), *Philosophy in the Flesh*, New York: Basic Books.
- Langacker, R. W. (1999), *Grammar and conceptualization*, Mouton de Gruyter.
- Langland-Hassan, P. (2008), 'Fractured phenomenologies: Thought insertion, inner speech, and the puzzle of extraneity', *Mind & Language* 23(4), 369-401.
- Langland-Hassan, P.; Gauker, C.; Richardson, M. J.; Dietz, A. & Faries, F. R. (2017), 'Metacognitive deficits in categorization tasks in a population with impaired inner speech', *Acta Psychologica* 181(Supplement C), 62 - 74.
- Larsen, S. F.; Schrauf, R. W.; Fromholt, P. & Rubin, D. C. (2002), 'Inner speech and bilingual autobiographical memory: A Polish-Danish cross-cultural study', *Memory* 10(1), 45-54.
- Larson, R. & Csikszentmihalyi, M. (1983), 'The experience sampling method', *New Directions for Methodology of Soc. and Behav. Sci.* 15, 41-56.
- Laurent, L.; Millot, J.-L.; Andrieu, P.; Camos, V.; Floccia, C. & Mathy, F. (2016), 'Inner speech sustains predictable task switching: direct evidence in adults', *Journal of Cognitive Psychology* 28(5), 585-592.
- Laver, J. D. M. (1980), 'Monitoring systems in the neurolinguistic control of speech production', *Errors in linguistic performance: Slips of the tongue, ear, pen, and hand*, 287-305.
- Lavigne, K. M.; Rapin, L. A.; Metzack, P. D.; Whitman, J. C.; Jung, K.; Dohen, M.; Løevenbruck, H. & Woodward, T. S. (2015), 'Left-Dominant Temporal-Frontal Hypercoupling in Schizophrenia Patients With Hallucinations During Speech Perception', *Schizophrenia bulletin* 41(1), 259-267.
- Lecours, A. & Joanette, Y. (1980), 'Linguistic and other psychological aspects of paroxysmal aphasia', *Brain and Language* 10(1), 1 - 23.
- Lécuyer, R.; Pécheux, M.-G. & Streri, A. (1996), *Le développement cognitif du nourrisson*, Nathan.
- Leibniz, G. W. (1765/1921), *Nouveaux Essais sur l'entendement humain*, Flammarion, Paris.
- Leiner, H. C.; Leiner, A. L. & Dow, R. S. (1986), 'Does the cerebellum contribute to mental skills?', *Behavioral neuroscience* 100(4), 443.
- Leroy-Gourhan, A. (1964), *Le Geste et la Parole*, Paris: Albin Michel.
- Leslie, A. M. (1994), 'ToMM, ToBy, and Agency: Core architecture and domain specificity'. In *Mapping the mind: Domain specificity in cognition and culture*. Gelman, L. H. & S. (ed.), 119-148.
- Levelt, W. J. M. (1983), 'Monitoring and self-repair in speech', *Cognition* 14(1), 41 - 104.
- Levelt, W. J. M. (1989), *Speaking: From intention to articulation*, Cambridge, MA: MIT Press.
- Levelt, W. J. M.; Roelofs, A. & Meyer, A. S. (1999), 'A theory of lexical access in speech production', *Behavioral and brain sciences* 22(01), 1-38.
- Levine, D. N.; Calvanio, R. & Popovics, A. (1982), 'Language in the absence of inner speech'. *Neuropsychologia* 20(4), 391-409.
- Lewis, M. & Brooks-Gunn, J. (1979), 'Toward a theory of social cognition: The development of self', *New directions for child and adolescent development* 1979(4), 1-20.
- Lieberman, A. M. & Mattingly, I. G. (1985), 'The motor theory of speech perception revised', *Cognition* 21(1), 1-36.
- Lieberman, A. M.; Cooper, F.; Harris, K. & MacNeilage, P. F. (1962), A motor theory of speech perception, in Stockholm, ed., 'Proc. Speech Communication Seminar'.

- Libet, B.; Gleason, C. A.; Wright, E. W. & Pearl, D. K. (1983), 'Time of conscious intention to act in relation to onset of cerebral activity (readiness-potential)', *Brain* 106(3), 623-642.
- Lindblom, B. (1988), *Phonetic invariance and the adaptive nature of speech*, London : Academic Press.
- Linden, D. E. J.; Thornton, K.; Kuswanto, C. N.; Johnston, S. J.; van de Ven, V. & Jackson, M. C. (2011), 'The Brains Voices: Comparing Nonclinical Auditory Hallucinations and Imagery', *Cerebral Cortex* 21(2), 330-337.
- Llorens, A.; Trébuchon, A.; Liegeois-Chauvel, C. & Alario, F.-X. (2011), 'Intra-Cranial Recordings of Brain Activity During Language Production', *Frontiers in Psychology* 2, 375.
- Locke, J. (1690), *Essai philosophique concernant l'entendement humain.*, Trad. Coste, Paris, Vrin, 1972.
- Locke, J. L. (1983), *Phonological acquisition and change*, Academic Press New York.
- Lœvenbruck, H. (1996), 'Pistes pour le contrôle dun robot parlant capable de réduction vocalique'.
- Lœvenbruck, H. (1999), 'An investigation of articulatory correlates of the accentual phrase in French, in 'Proceedings of the 14th International Congress of Phonetic Sciences', 667-670.
- Lœvenbruck, H. (2000), 'Effets articulatoires de l'emphase contrastive sur la Phrase Accentuelle en français', *Actes des XXIIIèmes Journées d'Etude sur la Parole, Aussois, France*, 165-168.
- Lœvenbruck, Henri (2018), *J'irai tuer pour vous*, Flammarion, Paris.
- Lœvenbruck, H. (2018), 'What the neurocognitive study of inner language reveals about our inner space'. In *Épistémocritique*, Vol. 18 : Langage intérieur / Espaces intérieur, Inner Speech / Inner Space. Smadja, S. & Patoine, P.-L. (ed.).
- Lœvenbruck, H.; Baciú, M.; Segebarth, C. & Abry, C. (2005), 'The left inferior frontal gyrus under focus: an fMRI study of the production of deixis via syntactic extraction and prosodic focus', *Journal of Neurolinguistics* 18(3), 237 - 258.
- Lœvenbruck, H.; Dohen, M. & Vilain, C. (2009), 'Pointing is special', *Some aspects of speech and the brain*, 211-258.
- Lœvenbruck, H.; Grandchamp, R.; Rapin, L.; Nalborczyk, L.; Dohen, M.; Perrier, P.; Baciú, M. & Perrone-Bertolotti, M. (2018), *A cognitive neuroscience view of inner language: to predict and to hear, see, feel*, Oxford University Press, Oxford, UK, 131-167.
- Lœvenbruck, H.; Jannet, M. A. B.; D'Imperio, M.; Spini, M. & Champagne-Lavau, M. (2013), 'Prosodic cues of sarcastic speech in French: slower, higher, wider, in '14th Annual Conference of the International Speech Communication Association (Interspeech 2013)', 3537-3541.
- Lœvenbruck, H.; Nalborczyk, L.; Perrone-Bertolotti, M.; Baeyens, C. & Grandchamp, R. (2017), *Un cas d'incarnation linguistique: implication du système moteur orofacial dans les ruminations mentales*, *Recherches sur la philosophie et le langage*, N° 33, Vrin, 297-324.
- Loewenstein, J. & Gentner, D. (2005), 'Relational language and the development of relational mapping', *Cognitive Psychology* 50(4), 315 - 353.
- Loy, D. (2012), *Nonduality: A study in comparative philosophy*, Prometheus Books.
- Luce, P. A. & Pisoni, D. B. (1998), 'Recognizing spoken words: The neighborhood activation model', *Ear and hearing* 19(1), 1-36.
- Lupyan, G. (2009), 'Extracommunicative functions of language: Verbal interference causes selective categorization impairments', *Psychonomic Bulletin & Review* 16(4), 711-718.
- Lupyan, G. & Mirman, D. (2013), 'Linking language and categorization: Evidence from aphasia', *Cortex* 49(5), 1187-1194.
- Mack, A. & Bachant, J. (1969), 'Perceived movement of the afterimage during eye movements', *Perception & Psychophysics* 6(6), 379-384.
- MacKay, D. G. (1990), 'Perception, action, and awareness: A three-body problem'. In *Relationships between perception and action: Current approaches*. Prinz, O. N. & W. (ed.), Springer-Verlag Berlin, 269-303.
- MacKay, D. G. (1992b), 'Constraints on theories of inner speech'. In *Auditory imagery*, Reisberg, D. (ed.), Lawrence Erlbaum, Hillsdale, N. J., 121-149.
- Maddieson, I. (2013), *Consonant Inventories*, Leipzig: Max Planck Institute for Evolutionary Anthropology, chapter <https://wals.info/chapter/1>.
- Maingueneau D. (2016), *Les termes clés de l'analyse du discours*. Editions du Seuil.
- Mani, N. & Plunkett, K. (2010), 'In the Infant's Mind's Ear Evidence for Implicit Naming in 18-Month-Olds', *Psychological Science* 21(7), 908-913.
- Marian, V. & Neisser, U. (2000), 'Language-dependent recall of autobiographical memories.', *Journal of Experimental Psychology: General* 129(3), 361.
- Mariën, P.; Ackermann, H.; Adamaszek, M.; Barwood, C. H. S.; Beaton, A.; Desmond, J.; De Witte, E.; Fawcett, A. J.; Hertrich, I.; Küper, M. & others (2014), 'Consensus paper: language and the cerebellum: an ongoing enigma', *The Cerebellum* 13(3), 386-410.

- Marshall, R. C.; Rappaport, B. Z. & Garcia-Bunuel, L. (1985), 'Self-monitoring behavior in a case of severe auditory agnosia with aphasia', *Brain and Language* 24(2), 297-313.
- Marslen-Wilson, W. & Tyler, L. K. (1980), 'The temporal structure of spoken language understanding', *Cognition* 8(1), 1 - 71.
- Marslen-Wilson, W. D. (1987), 'Functional parallelism in spoken word-recognition', *Cognition* 25(1), 71 - 102.
- Martin, G. B. & Clark, R. D. (1982), 'Distress crying in neonates: Species and peer specificity.', *Developmental psychology* 18(1), 3.
- Martin, P. (1987), 'Prosodic and rhythmic structures in French', *Linguistics* 25(5), 925-950.
- Martinet, A. (1960), *Éléments de linguistique générale*, A. Colin.
- Martinet, A. (1989), *Fonction et dynamique des langues*, A. Colin.
- Martínez-Manrique, F. & Vicente, A. (2010), 'What the...! The role of inner speech in conscious thought', *Journal of Consciousness Studies* 17(9-1), 141-167.
- Marvel, C. L. & Desmond, J. E. (2012), 'From storage to manipulation: How the neural correlates of verbal working memory reflect varying demands on inner speech', *Brain and language* 120(1), 42-51.
- Marx, K. & Engels, F. (1846), *Idéologie Allemande, Chapitre 3*.
- Marx, K. & Engels, F. (1857), *Grundrisse, Chapitre 3*.
- Maspero, G. (1875), *Histoire ancienne des peuples de l'Orient classique. Les origines : Egypte et Chaldée*, Paris, Hachette.
- McCarthy, J. J. (2007), 'What Is Optimality Theory?', *Language and Linguistics Compass* 1(4), 260-291.
- McCarthy-Jones, S. & Fernyhough, C. (2011), 'The varieties of inner speech: links between quality of inner speech and psychopathological variables in a sample of young adults', *Consciousness and cognition* 20(4), 1586-1593.
- McClelland, J. L. & Elman, J. L. (1986), 'The TRACE model of speech perception', *Cognitive Psychology* 18(1), 1 - 86.
- Meacham, J. A. (1979), 'The role of verbal activity in remembering the goals of actions', *The development of self-regulation through private speech* 6, 237-263.
- Meillier, C. (1982), 'La fonction thérapeutique de la musique et de la poésie dans le recueil des Bucoliques de Théocrite', *Bulletin de l'Association Guillaume Budé* 1(2), 164-186.
- Meltzoff, A. N. & Moore, M. K. (1977), 'Imitation of facial and manual gestures by human neonates', *Science* 198(4312), 75-78.
- Ménard, L.; Loevenbruck, H. & Savariaux, C. (2006), 'Articulatory and acoustic correlates of contrastive focus in French children and adults', *Speech production: Models, phonetic processes and techniques*, 227-251.
- Merleau-Ponty, M. (1945), *Phénoménologie de la Perception*, Gallimard.
- Merleau-Ponty, M. (2002), *Causeries, 1948*, Paris : Le Seuil.
- Mertens, P. (1993), International grouping, boundaries, and syntactic structure in French, in 'ESCA Workshop on Prosody'.
- Miall, R. C. & Wolpert, D. M. (1996), 'Forward models for physiological motor control', *Neural networks* 9(8), 1265-1279.
- Miall, R. C. (2003), 'Connecting mirror neurons and forward models', *Neuroreport* 14(17), 2135-2137.
- Miall, R. C.; Weir, D. J.; Wolpert, D. M. & Stein, J. F. (1993), 'Is the cerebellum a Smith Predictor?', *Journal of Motor Behaviour* 25, 203-216.
- Millikan, R. G. (1989), 'An ambiguity in the notion function', *Biology and Philosophy* 4(2), 172-176.
- Millikan, R. G. (2005), *Language: A biological model*, Oxford University Press on Demand.
- Mitchell, D. C. (1994), 'Sentence parsing', *Handbook of psycholinguistics*, 375-409.
- Mithen, S. (1996), *The prehistory of the mind*, London and New York: Thames & Hudson.
- Molinari, M.; Chiricozzi, F. R.; Clausi, S.; Tedesco, A. M.; De Lisa, M. & Leggio, M. G. (2008), 'Cerebellum and Detection of Sequences, from Perception to Cognition', *The Cerebellum* 7(4), 611-615.
- Monnin, J.; Loevenbruck, H. & Vilain, A. (2015), Consonant and syllable frequency effects in stop and fricative acquisition in preschool Drehu- and French-acquiring children, in '18th International Congress of Phonetic Sciences (ICPhS 2015)', 665.
- Monnin, P. & Grosjean, F. (1993), 'Les structures de performance en français: caractérisation et prédiction', *L'Année Psychologique* 93(1), 9-30.
- Mooneyham, B. W. & Schooler, J. W. (2013), 'The costs and benefits of mind-wandering: a review.', *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale* 67(1), 11.
- Morel, M.-A. & Danon-Boileau, L. (1998), *Grammaire de l'intonation l'exemple du français*, Editions Ophrys.

- Morin, A. (2005), 'Possible links between self-awareness and inner speech theoretical background, underlying mechanisms, and empirical evidence', *Journal of Consciousness Studies* 12(4-5), 4-5.
- Morin, A. (2009), 'Self-awareness deficits following loss of inner speech: Dr. Jill Bolte Taylor's case study', *Consciousness and Cognition* 18(2), 524 - 529.
- Morin, A. (2018), *The self-reflective functions of inner speech: thirteen years later*, Oxford University Press, Oxford, UK, 276-298.
- Morin, A. & Hamper, B. (2012), 'Self-Reflection and the Inner Voice: Activation of the Left Inferior Frontal Gyrus During Perceptual and Conceptual Self-Referential Thinking', *Open Neuroimaging Journal* 6, 78-89.
- Morin, A.; Uttl, B. & Hamper, B. (2011), 'Self-reported frequency, content, and functions of inner speech', *Procedia-Social and Behavioral Sciences* 30, 1714-1718.
- Moritz-Gasser, S. & Duffau, H. (2013), 'The anatomo-functional connectivity of word repetition: insights provided by awake brain tumor surgery', *Frontiers in Human Neuroscience* 7, 405.
- Moritz-Gasser, S.; Herbet, G. & Duffau, H. (2013), 'Mapping the connectivity underlying multimodal (verbal and non-verbal) semantic processing: A brain electrostimulation study', *Neuropsychologia* 51(10), 1814 - 1822.
- Moritz-Gasser, S.; Herbet, G. & Duffau, H. (2015), 'Integrating emotional valence and semantics in the human ventral stream: a hodological account', *Frontiers in Psychology* 6, 32.
- Morrison, C. M. & Conway, M. A. (2010), 'First words and first memories', *Cognition* 116(1), 23 - 32.
- Morton, J. (1969), 'Interaction of information in word recognition.', *Psychological review* 76(2), 165.
- Moulin-Frier, C.; Diard, J.; Schwartz, J.-L. & Bessière, P. (2015), 'COSMO ("Communicating about Objects using Sensory-Motor Operations"): A Bayesian modeling framework for studying speech communication and the emergence of phonological systems', *Journal of Phonetics* 53, 5 - 41.
- Mounin, G. (1968), *Clefs pour la linguistique*, Seghers.
- Munding, D.; Dubarry, A.-S. & Alario, F.-X. (2016), 'On the cortical dynamics of word production: a review of the MEG evidence', *Language, Cognition and Neuroscience* 31(4), 441-462.
- Munson, B.; Edwards, J. & Beckman, M. E. (2005), 'Phonological knowledge in typical and atypical speech-sound development', *Topics in Language Disorders* 25(3), 190.
- Murdoch, B. E. (2010), 'The cerebellum and language: Historical perspective and review', *Cortex* 46(7), 858 - 868.
- Nakajima, Y.; Kashioka, H.; Shikano, K. & Campbell, N. (2003), 'Non-audible murmur recognition', *Proceedings of Eurospeech 2003*, Nara, Japan.
- Nagel, E. (1961), 'The structure of science: Problems in the logic of scientific explanation', .
- Nalborczyk, L.; Perrone-Bertolotti, M.; Baeyens, C.; Grandchamp, R.; Polosan, M.; Spinelli, E.; Koster, E. H. W. & Lœvenbruck, H. (2017), 'Orofacial electromyographic correlates of induced verbal rumination', *Biological Psychology* 127, 53 - 63.
- Nayani, T. H. & David, A. S. (1996), 'The auditory hallucination: a phenomenological survey', *Psychological Medicine* 26(01), 177-189.
- Neisser, U. (1988), 'Five kinds of self-knowledge', *Philosophical Psychology* 1(1), 35-59.
- Nelson, K. (2005), 'Emerging levels of consciousness in early human development', , 116-141.
- Nelson, K. & Fivush, R. (2000), *Socialization of memory*, New York : Oxford University Press, 283-295.
- Nève, F.-X. (1996), *Essai de grammaire de la langue des signes française*, Bibliothèque de la Faculté de philosophie et lettres de l'Université de Liège, Genève, Droz.
- Newmeyer, F. J. (2016), 'On language not being at root a communication system: Some morphosyntactic considerations'. *Papers dedicated to Anne Reboul*. Dupuy, L.; Grabizna, A.; Foudon, N. & Saint-Germier, P. (eds.).
- Newton, A. M. & de Villiers, J. G. (2007), 'Thinking While Talking: Adults Fail Nonverbal False-Belief Reasoning', *Psychological Science* 18(7), 574-579.
- Ngon, C. & Peperkamp, S. (2016), 'What infants know about the unsaid: Phonological categorization in the absence of auditory input', *Cognition* 152, 53-60.
- Nickels, L. & Howard, D. (1995), 'Phonological Errors in Aphasic Naming: Comprehension, Monitoring and Lexicality', *Cortex* 31(2), 209 - 237.
- Nooteboom, S. G. (1980), 'Speaking and unspeaking: Detection and correction of phonological and lexical errors in spontaneous speech', .
- Norris, D. (1994), 'Shortlist: a connectionist model of continuous speech recognition', *Cognition* 52(3), 189-234.
- Norris, D. & McQueen, J. M. (2008), 'Shortlist B: a Bayesian model of continuous speech recognition.', *Psychological review* 115(2), 357.

- Norris, D.; McQueen, J. M. & Cutler, A. (2000), 'Merging information in speech recognition: Feedback is never necessary', *Behavioral and Brain Sciences* 23(3), 299-325.
- Norris, D.; McQueen, J. M. & Cutler, A. (2016), 'Prediction, Bayesian inference and feedback in speech recognition', *Language, Cognition and Neuroscience* 31(1), 4-18.
- Numminen, J. & Curio, G. (1999), 'Differential effects of overt, covert and replayed speech on vowel-evoked responses of the human auditory cortex', *Neuroscience Letters* 272(1), 29-32.
- Numminen, J.; Salmelin, R. & Hari, R. (1999), 'Subject's own speech reduces reactivity of the human auditory cortex', *Neuroscience letters* 265(2), 119-122.
- Obleser, J.; Zimmermann, J.; Van Meter, J. & Rauschecker, J. P. (2007), 'Multiple Stages of Auditory Speech Perception Reflected in Event-Related fMRI', *Cerebral Cortex* 17(10), 2251-2257.
- Ombredane, A. (1951), *L'aphasie et l'élaboration de la pensée explicite*, Paris, Presses Universitaires de France.
- Oomen, C. C. E. & Postma, A. (2001), 'Effects of increased speech rate on monitoring and self-repair', *Journal of Psycholinguistic Research* 30, 163-184.
- Oppenheim, G. M. & Dell, G. S. (2008), 'Inner speech slips exhibit lexical bias, but not the phonemic similarity effect', *Cognition* 106(1), 528-537.
- Oppenheim, G. M. & Dell, G. S. (2010), 'Motor movement matters: The flexible abstractness of inner speech', *Memory & cognition* 38(8), 1147-1160.
- Orenstein, B. (2018), 'Conférence à Grenoble Ecole de Management, 12 novembre 2018'.
- Pacherie, E. (2008), 'The phenomenology of action: A conceptual framework', *Cognition* 107(1), 179-217.
- Paivio, A. (1990), *Mental representations: A dual coding approach*, Oxford University Press.
- Panaccio, C. (2014), *Le Discours Intérieur. De Platon à Guillaume d'Ockham*, Le Seuil.
- Paradis, M. (1997), 'Représentation lexicale et conceptuelle chez les bilingues: deux langues trois systèmes', *Explorations du lexique*, 15-27.
- Parmentier, M. (2014), 'Relations linguistiques et mathématiques chez Leibniz', *Methodos [En ligne]* 14.
- Pascal, B. (1670), *Pensées*, Edition de Port-Royal (édition Brunschvig, 1897, Lafuma, 1935, Sellier, 1976).
- Patri, J.-F.; Diard, J. & Perrier, P. (2016), Modélisation bayésienne de la planification motrice des gestes de parole: Évaluation du rôle des différentes modalités sensorielles, in '31e Journées d'Études sur la Parole (JEP-TALN-RECITAL 2016)', 419-427.
- Patri, J.-F.; Perrier, P.; Schwartz, J.-L. & Diard, J. (2018), 'What drives the perceptual change resulting from speech motor adaptation? Evaluation of hypotheses in a Bayesian modeling framework', *PLoS Computational Biology* 14(1), 1-38.
- Pavlenko, A. (1999), 'New approaches to concepts in bilingual memory', *Bilingualism: Language and Cognition* 2(3), 209-230.
- Pavlenko, A. (2006), *Bilingual minds: Emotional experience, expression, and representation*, Multilingual Matters : Clevedon, UK.
- Pavlenko, A. (2014), *The bilingual mind: And what it tells us about language and thought*, Cambridge University Press.
- Peirce, C. S. (1868), 'Some consequences of four incapacities', *The Journal of Speculative Philosophy* 2(3), 140-157.
- Peirce, C. S. (1934), 'Pragmatism and Pragmaticism', *Collected Papers of Charles Sanders Peirce*, Vol. V, Cambridge, MA : Harvard University Press.
- Pelletier, J. (2004), 'Vers une philosophie cognitive du langage?' In *La philosophie cognitive*. Pacherie, E. & Proust, J. (ed.), Ophrys, 139-162.
- Perkell, J.; Matthies, M.; Lane, H.; Guenther, F.; Wilhelms-Tricarico, R.; Wozniak, J. & Guiod, P. (1997), 'Speech motor control: Acoustic goals, saturation effects, auditory feedback and internal models', *Speech Communication* 22(2), 227 - 250.
- Perkell, J. S. (2012), 'Movement goals and feedback and feedforward control mechanisms in speech production', *Journal of Neurolinguistics* 25(5), 382 - 407.
- Perrier, P.; Løevenbruck, H. & Payan, Y. (1996), 'Control of tongue movements in speech: The equilibrium point hypothesis perspective', *Journal of Phonetics* 24(1), 53-75.
- Perrone-Bertolotti, M.; Dohen, M.; Løevenbruck, H.; Sato, M.; Pichat, C. & Baciú, M. (2013), 'Neural correlates of the perception of contrastive prosodic focus in French: A functional magnetic resonance imaging study', *Human Brain Mapping* 34(10), 2574-2591.
- Perrone-Bertolotti, M.; Grandchamp, R.; Rapin, L.; Baciú, M.; Lachaux, J.-P. & Løevenbruck, H. (2016), *Langage intérieur*, Collection Neuropsychologie, De Boeck Supérieur : Louvain-la-Neuve, 109-123.

- Perrone-Bertolotti, M.; Rapin, L.; Lachaux, J.-P.; Baciú, M. & Loevenbruck, H. (2014), 'What is that little voice inside my head? Inner speech phenomenology, its role in cognitive performance, and its relation to self-monitoring', *Behavioural Brain Research* 261(0), 220 - 239.
- Petrilli, S. (2017), *The self as a sign, the world, and the other: Living semiotics*, New York: Routledge.
- Philippe, G. (2001), 'Le paradoxe énonciatif endophasique et ses premières solutions fictionnelles', in Bergounioux Gabriel (dir.), *Langue française*, n° 132 : *La Parole intérieure*, 2001, p. 96-105.
- Piaget, J. (1923), *Le Langage et la pensée chez l'enfant*, Paris : Delachaux et Niestlé.
- Picard, L.; Eustache, F. & Piolino, P. (2009), 'De la mémoire épisodique à la mémoire autobiographique: approche développementale', *L'Année psychologique* 109(2), 197-236.
- Pichon, S. & Kell, C. A. (2013), 'Affective and Sensorimotor Components of Emotional Prosody Generation', *Journal of Neuroscience* 33(4), 1640-1650.
- Pickering, M. J. & Garrod, S. (2007), 'Do people use language production to make predictions during comprehension?', *Trends in Cognitive Sciences* 11(3), 105-110.
- Pickering, M. J. & Clark, A. (2014), 'Getting ahead: forward models and their place in cognitive architecture', *Trends in Cognitive Sciences*, 18(9), 451 - 456.
- Pickering, M. J. & Garrod, S. (2013), 'An integrated theory of language production and comprehension', *Behavioral and Brain Sciences*, 36(4), 329-347.
- Pickering, M. J. & Garrod, S. (2014), 'Neural integration of language production and comprehension', *Proceedings of the National Academy of Sciences*, 111(43), 15291-15292.
- Pierrehumbert, J. B. (1980), 'The phonology and phonetics of English intonation', PhD thesis, Massachusetts Institute of Technology.
- Pinker, S. (1994), *The language instinct: How the mind creates language*, HarperCollins.
- Pinker, S. (1997), *How the mind works*.
- Pinker, S. & Bloom, P. (1990), 'Natural language and natural selection', *Behavioral and brain sciences* 13(4), 707-727.
- Platon (2008), *Œuvres complètes*, Trad. Luc Brisson, Flammarion, Paris.
- Poeppe, D. (2003), 'The analysis of speech in different temporal integration windows: cerebral lateralization as [] asymmetric sampling in time', *Speech Communication* 41(1), 245-255.
- Poeppe, D. & Monahan, P. J. (2011), 'Feedforward and feedback in speech perception: Revisiting analysis by synthesis', *Language and Cognitive Processes* 26(7), 935-951.
- Poppenk, J.; Walia, G.; McIntosh, A. R.; Joanisse, M. F.; Klein, D. & Köhler, S. (), 'Why is the meaning of a sentence better remembered than its form? An fMRI study on the role of novelty-encoding processes', *Hippocampus* 18(9), 909-918.
- Post, B. M. B. (2000), 'Tonal and phrasal structures in French intonation'.
- Postma, A. (2000), 'Detection of errors during speech production: A review of speech monitoring models', *Cognition* 77(2), 97-132.
- Prévost, A. (1731/1839), *Histoire de Manon lescaut et du Chevalier Des Grieux*, Ernest Bourdin et comp., Paris.
- Price, C. J. (2012), 'A review and synthesis of the first 20 years of PET and fMRI studies of heard speech, spoken language and reading', *Neuroimage* 62(2), 816-847.
- Prinz, W. (1997), 'Perception and Action Planning', *European Journal of Cognitive Psychology* 9(2), 129-154.
- Proust, J. & Pacherie, E. (2008), 'Neurosciences et compréhension d'autrui'. In *Philosophie et neurosciences*, Faucher, L. & Poirier, P. (ed.), Bellarmine.
- Raichle, M. E. (2015), 'The brain's default mode network', *Annual Review of Neuroscience*(38), 433-47.
- Raichle, M. E.; Fiez, J. A.; Videen, T. O.; MacLeod, A.-M. K.; Pardo, J. V.; Fox, P. T. & Petersen, S. E. (1994), 'Practice-related Changes in Human Brain Functional Anatomy during Nonmotor Learning', *Cerebral Cortex* 4(1), 8-26.
- Rapin, L.; Dohen, M. & Loevenbruck, H. (2016), *Les hallucinations auditives verbales*, Collection Neuropsychologie, De Boeck Supérieur, 347-370.
- Rapin, L. A.; Dohen, M.; Loevenbruck, H.; Whitman, J. C.; Metzack, P. D. & Woodward, T. S. (2012), 'Hyperintensity of functional networks involving voice-selective cortical regions during silent thought in schizophrenia', *Psychiatry Research: Neuroimaging*.
- Rapin, L.; Dohen, M.; Polosan, M.; Perrier, P. & Loevenbruck, H. (2013), 'An EMG study of the lip muscles during covert auditory verbal hallucinations in schizophrenia', *Journal of Speech, Language, and Hearing Research* 56(6), S1882-S1893.
- Rathbone, C. J.; Moulin, C. J. A. & Conway, M. A. (2008), 'Self-centered memories: The reminiscence bump and the self', *Memory & Cognition* 36(8), 1403-1414.

- Rauschecker, J. P. & Scott, S. K. (2009), 'Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing', *Nature neuroscience* 12(6), 718-724.
- Reboul, A. (2017), *Cognition and Communication in the Evolution of Language*, Vol. 5, Oxford University Press, Oxford, UK.
- Reboul, A. C. (2015), 'Why language really is not a communication system: a cognitive view of language evolution', *Frontiers in psychology* 6.
- Reisberg, D.; Smith, J. D.; Baxter, D. A. & Sonenshine, M. (1989), "'Enacted" auditory images are ambiguous;"pure" auditory images are not', *The Quarterly Journal of Experimental Psychology* 41(3), 619-641.
- Repp, B. H. & Knoblich, G. (2004), 'Perceiving Action Identity', *Psychological Science* 15(9), 604-609.
- Rico, C. (2008), 'Figure et théorie du signe : les solutions de Saint Jérôme', *Modèles linguistiques* 29(58), 79-98.
- Ricœur, P. (1983), *Temps et récit, vol. I*, Paris : Seuil.
- Ricœur, P. (1990), *Soi-même comme un autre*, Paris : Seuil.
- Riès, S.; Janssen, N.; Burle, B. & Alario, F.-X. (2013), 'Response-Locked Brain Dynamics of Word Production', *PLOS ONE* 8(3), 1-14.
- Riès, S.; Janssen, N.; Dufau, S.; Alario, F.-X. & Burle, B. (2011), 'General-Purpose Monitoring during Speech Production', *Journal of Cognitive Neuroscience* 23(6), 1419-1436.
- Rimbaud, A. (1991), 'Lettre à Paul Demeny, 15 mai 1871'. In *Œuvre- Vie*, Edition du centenaire établie par Alain Borer, arléa : Paris, 187.
- Rinpoche, K. (1987), *Foundation of Buddhist meditation*, New Delhi, India: Indraprastha Press.
- Rizzolatti, G.; Fadiga, L.; Gallese, V. & Fogassi, L. (1996), 'Premotor cortex and the recognition of motor actions', *Cognitive Brain Research* 3(2), 131 - 141.
- Robinet, I. (1995), *Méditation Taoïste*, Albin Michel: Paris.
- Roelofs, A. (2014), 'A dorsal-pathway account of aphasic language production: The WEAVER++/ARC model', *Cortex* 59, 33 - 48.
- Rolland, G. & Lœvenbruck, H. (2002), Characteristics of the accentual phrase in French: An acoustic, articulatory and perceptual study, in 'Speech Prosody 2002, International Conference'.
- Rolland, G. & Lœvenbruck, H. (2003), 'Pertinence physiologique et perceptive du Syntagme Accentuel en français', *Journées Prosodie 2001*, 61-66.
- Rosenthal, V. (2012), 'La voix de l'intérieur', *Intellectica*, 2 (58), 53-89.
- Rossi, M. (1985), 'L'intonation et l'organisation de l'énoncé', *Phonetica* 42(2-3), 135-153.
- Rossi, M. (1999), *L'intonation: le système du français: description et modélisation*, Editions Ophrys.
- Rossi, M. (2001), 'Les lapsus et la production de la parole', *Psychologie française* 46(1), 27-41.
- Roulet, E. (1984), 'Speech acts, discourse structure, and pragmatic connectives', *Journal of Pragmatics* 8(1), 31 - 47.
- Roulet, E. & Green, C. W. (2006), 'French-English Glossary of linguistic terms', Retrieved from <https://feglossary.sil.org/sites/feglossary/files/amden.pdf>.
- Rousseau, J.-J. (1781), *Essai sur l'origine des langues ou, il est parlé de la mélodie et de limitation musicale*, Œuvres posthumes, Du Peyrou, Genève.
- Roy, A. C.; Craighero, L.; Fabbri-Destro, M. & Fadiga, L. (2008), 'Phonological and lexical motor facilitation during speech listening: A transcranial magnetic stimulation study', *Journal of Physiology-Paris* 102(1), 101 - 105.
- Ruby, P. & Decety, J. (2001), 'Effect of subjective perspective taking during simulation of action: a PET investigation of agency', *Nature neuroscience* 4(5), 546.
- Ruiter, D. & Peter, J. (2007), 'Postcards from the mind: The relationship between speech, imagistic gesture, and thought', *Gesture* 7(1), 21-38.
- de Ruiter, J. P. & Cummins, C. (2013), 'Forward modelling requires intention recognition and non-impooverished predictions', *Behavioral and Brain Sciences* 36(4), 351-351.
- Russell, B. (1989), 'Ecrits de Logique Philosophique', .
- Ryle, G. (1972), 'Thinking and saying', *Rice Institute Pamphlet-Rice University Studies* 58(3).
- Sachs, J. S. (1974), 'Memory in reading and listening to discourse', *Memory & Cognition* 2(1), 95-100.
- Saint-Paul, G. (1892), *Essais sur le langage intérieur*, A. Storck, Lyon.
- Saltzman, E. L. & Munhall, K. G. (1989), 'A dynamical approach to gestural patterning in speech production', *Ecological psychology* 1(4), 333-382.
- Sato, M.; Baciú, M.; Lœvenbruck, H.; Schwartz, J. L.; Cathiard, M. A.; Segebarth, C. & Abry, C. (2004), 'Multistable representation of speech forms: A functional MRI study of verbal transformations', *NeuroImage* 23(3), 1143-1151.

- Schaller, S. (1991/2012), *A man without words*, Second edition. Berkeley, CA: University of California Press.
- Schmahmann, J. D. (2004), 'Disorders of the Cerebellum: Ataxia, Dysmetria of Thought, and the Cerebellar Cognitive Affective Syndrome', *The Journal of Neuropsychiatry and Clinical Neurosciences* 16(3), 367-378.
- Schooler, J. W. (2002), 'Re-representing consciousness: dissociations between experience and meta-consciousness', *Trends in Cognitive Sciences* 6(8), 339 - 344.
- Schooler, J. W.; Mrazek, M. D.; Franklin, M. S.; Baird, B.; Mooneyham, B. W.; Zedelius, C. & Broadway, J. M. (2014), Chapter One - The Middle Way: Finding the Balance between Mindfulness and Mind-Wandering, in Brian H. Ross, ed., , Academic Press, , 1 - 33.
- Schooler, J. W.; Smallwood, J.; Christoff, K.; Handy, T. C.; Reichle, E. D. & Sayette, M. A. (2011), 'Meta-awareness, perceptual decoupling and the wandering mind', *Trends in cognitive sciences* 15(7), 319-326.
- Schwartz, B. L. & Evans, S. (2001), 'Episodic memory in primates', *American Journal of Primatology* 55(2), 71-85.
- Schwartz, J. L.; Basirat, A.; Ménard, L. & Sato, M. (2012), 'The Perception-for-Action-Control Theory (PACT): A perceptuo-motor theory of speech perception', *Journal of Neurolinguistics* 25(5), 336-354.
- Schwartz, J.-L.; Abry, C.; Boë, L.-J. & Cathiard, M. A. (2002), 'Phonology in a theory of perception-for-action-control'. In *Phonetics, Phonology, and Cognition*. Laks, J. D. & B. (ed.), Oxford University Press, Oxford, 244-280.
- Schwartz, J.-L.; Boë, L.-J. & Abry, C. (2007), 'Linking dispersion-focalization theory and the maximum utilization of the available distinctive features principle in a perception-for-action-control theory'. In *Experimental approaches to phonology*. Solé, M.; Beddor, P. & Ohala, M. (ed.), Oxford University Press, Oxford, 104-124.
- Schwartz, J.-L. & Lœvenbruck, H. (2012), 'La parole, entre objets vocaux et objets du monde'. In *La psychologie cognitive*. Denis, M. (ed.), Editions de la Maison des Sciences de l'Homme.
- Schwartz, J.-L.; Sato, M. & Fadiga, L. (2008), 'The common language of speech perception and action: a neurocognitive perspective', *Revue française de linguistique appliquée* 13(2), 9-22.
- Scott, M.; Yeung, H. H.; Gick, B. & Werker, J. F. (2013), 'Inner speech captures the perception of external speech', *The Journal of the Acoustical Society of America* 133(4), EL286-EL292.
- Scott-Phillips, T. (2015), *Speaking Our Minds: Why human communication is different, and how language evolved to make it special*, Basingstoke: Palgrave Macmillan.
- Seal, M. L.; Aleman, A. & McGuire, P. K. (2004), 'Compelling imagery, unanticipated speech and deceptive memory: Neurocognitive models of auditory verbal hallucinations in schizophrenia', *Cognitive Neuropsychiatry* 9(1), 43-72.
- Searle, J. R. (1972), *Chomskys revolution in linguistics*, Vol. 18, New York Review of Books.
- Sebanz, N. & Knoblich, G. (2009), 'Prediction in Joint Action: What, When, and Where', *Topics in Cognitive Science* 1(2), 353-367.
- Selkirk, E. O. (1986), *Phonology and syntax: the relationship between sound and structure*, MIT press.
- Shah, P. R. (1983), 'Tantra its therapeutic aspect', .
- Shergill, S. S.; Brammer, M. J.; Fukuda, R.; Bullmore, E.; Amaro Jr, E.; Murray, R. M. & McGuire, P. K. (2002), 'Modulation of activity in temporal cortex during generation of inner speech', *Human brain mapping* 16(4), 219-227.
- Shergill, S. S.; Bullmore, E.; Simmons, A.; Murray, R. & McGuire, P. (2000), 'Functional anatomy of auditory verbal imagery in schizophrenic patients with auditory hallucinations', *American Journal of Psychiatry* 157(10), 1691-1693.
- Silverman, K. & Pierrehumbert, J. (1990), 'The timing of prenuclear high accents in English', *Papers in laboratory phonology I*, 72-106.
- Simcock, G. & Hayne, H. (2003), 'Age-related changes in verbal and nonverbal memory during early childhood.', *Developmental Psychology* 39(5), 805-814.
- Skeide, M. A. & Friederici, A. D. (2016), 'The ontogeny of the cortical language network', *Nature Reviews Neuroscience* 17(5), 323.
- Skinner, B. F. (1957), *Verbal behavior*, New York: Appleton-Century-Crofts..
- Skipper, J. I.; Van Wassenhove, V.; Nusbaum, H. C. & Small, S. L. (2007), 'Hearing lips and seeing voices: how cortical areas supporting speech production mediate audiovisual speech perception', *Cerebral Cortex* 17(10), 2387-2399.
- Skipper, J. I.; Devlin, J. T. & Lametti, D. R. (2017), 'The hearing ear is always found close to the speaking tongue: Review of the role of the motor system in speech perception', *Brain and Language* 164, 77 - 105.

- Smadja, S. (2018), 'Le langage intérieur : un nouveau protocole d'enquête. Fait linguistique et fait endophasique / Investigating Inner speech : a new protocole. Linguistic and inner speech fact. In *Épistémocritique*, Vol. 18 : Langage intérieur / Espaces intérieur, Inner Speech / Inner Space. Smadja, S. & Patoine, P.-L. (ed.)
- Smallwood, J. & Andrews-Hanna, J. (2013), 'Not all minds that wander are lost: the importance of a balanced perspective on the mind-wandering state', *Frontiers in Psychology* 4, 441.
- Smallwood, J. & Schooler, J. W. (2015), 'The Science of Mind Wandering: Empirically Navigating the Stream of Consciousness', *Annual Review of Psychology* 66(1), 487-518.
- De Smet, H. J. D.; Paquier, P.; Verhoeven, J. & Mariën, P. (2013), 'The cerebellum: Its role in language and related cognitive and affective functions', *Brain and Language* 127(3), 334 - 342.
- Smith, B. L.; Hillenbrand, J.; Wasowicz, J. & Preston, J. (1986), 'Durational characteristics of vocal and subvocal speech-implications concerning phonological organization and articulatory difficulty', *Journal of Phonetics* 14(2), 265-281.
- Smith, E. E.; Shoben, E. J. & Rips, L. J. (1974), 'Structure and process in semantic memory: A featural model for semantic decisions.', *Psychological review* 81(3), 214.
- Smith, J. D.; Wilson, M. & Reisberg, D. (1995), 'The role of subvocalization in auditory imagery', *Neuropsychologia* 33(11), 1433-1454.
- Sokolov, A. N. (1972), *Inner speech and thought*, Plenum Press New York.
- Sokolov, A. A.; Miall, R. C. & Ivry, R. B. (2017), 'The Cerebellum: Adaptive Prediction for Movement and Cognition', *Trends in Cognitive Sciences* 21(5), 313 - 332.
- Sommer, I. E. C.; Diederer, K. M. J.; Blom, J. D.; Willems, A.; Kushan, L.; Slotema, K.; Boks, M. P. M.; Daalman, K.; Hoek, H. W.; Neggers, S. F. W. & others (2008), 'Auditory verbal hallucinations predominantly activate the right inferior frontal area', *Brain* 131(12), 3169-3177.
- Sperry, R. W. (1950), 'Neural basis of the spontaneous optokinetic response produced by visual inversion.', *Journal of comparative and physiological psychology* 43(6), 482.
- Stackhouse, J.; Vance, M.; Pascoe, M. & Wells, B. (2007), *Compendium of auditory and speech tasks*, John Wiley & Sons.
- Stackhouse, J. & Wells, B. (1997), *Childrens speech and literacy difficulties: A psycholinguistic framework*, Wiley.
- Stanovich, K. E. & West, R. F. (2000), 'Individual differences in reasoning: Implications for the rationality debate?', *Behavioral and Brain Sciences* 23(5), 645-665.
- Stemberger, J. P. (1982), 'Syntactic errors in speech', *Journal of Psycholinguistic Research* 11(4), 313-345.
- Stephens, G. L. & Graham, G. (2000), *When self-consciousness breaks: Alien voices and inserted thoughts.*, The MIT press.
- Stevens, K. N. (1989), 'On the quantal nature of speech', *Journal of Phonetics* 17, 3-45.
- Stevens, K. N. & Halle, M. (1967), 'Remarks on analysis by synthesis and distinctive features', , 88-102.
- Stokoe, W. C. (1960), 'Sign language structure: an outline of the visual communication.', .
- Stoodley, C. J.; Valera, E. M. & Schmahmann, J. D. (2012), 'Functional topography of the cerebellum for motor and cognitive tasks: An fMRI study', *NeuroImage* 59(2), 1560 - 1570.
- Stricker, S. (1885), *Du langage et de la musique*, Traduit de l'allemand par Frédéric Schwiedland, Alcan, Paris.
- Strijkers, K.; Runnqvist, E.; Costa, A. & Holcomb, P. (2013), 'The poor helping the rich: How can incomplete representations monitor complete ones?', *Behavioral and Brain Sciences* 36(4), 374-375.
- Suddendorf, T. S. & Corballis, M. C. (1997), 'Mental time travel and the evolution of the human mind', *Genet. Soc. Gen. Psychol. Monogr.* 123, 133-167.
- Sun, Y. & Peperkamp, S. (2016), 'The role of speech production in phonological decoding during visual word recognition: evidence from phonotactic repair', *Language, Cognition and Neuroscience* 31(3), 391-403.
- Swiney, L. & Sousa, P. (2014), 'A new comparator account of auditory verbal hallucinations: how motor prediction can plausibly contribute to the sense of agency for inner speech', *Frontiers in human neuroscience* 8, 72-86.
- Talmy, L. (2000), *Toward a cognitive semantics*, MIT press.
- Tamada, T.; Miyauchi, S.; Imamizu, H.; Yoshioka, T. & Kawato, M. (1999), 'Cerebro-cerebellar functional connectivity revealed by the laterality index in tool-use learning', *Neuroreport* 10(2), 325-331.
- Tanenhaus, M. K. & Trueswell, J. C. (1995), *Sentence comprehension.*, Academic Press, San Diego, 217-262.
- Tian, X. & Poeppel, D. (2012), 'Mental imagery of speech: linking motor and perceptual systems through internal simulation and estimation', *Frontiers in human neuroscience* 6, 152-162.

- Tian, X. & Poeppel, D. (2013), 'The Effect of Imagination on Stimulation: The Functional Specificity of Efference Copies in Speech Processing', *Journal of Cognitive Neuroscience* 25(7), 1020-1036.
- Tian, X. & Poeppel, D. (2015), 'Dynamics of Self-monitoring and Error Detection in Speech Production: Evidence from Mental Imagery and MEG', *Journal of Cognitive Neuroscience* 27(2), 352-364.
- Toda, T. & Shikano, K. (2005), 'NAM-to-speech conversion with Gaussian mixture models', *Proceedings of Interspeech 2005, September 4-8, Lisbon, Portugal*, 1957-1960.
- Tomasello, M. (2003), *Constructing a language: A usage-based theory of language acquisition*, Cambridge, MA: Harvard University Press.
- Tomasello, M.; Carpenter, M. & Liszkowski, U. (2007), 'A new look at infant pointing', *Child Development* 78(3), 705-722.
- Touati, P. (1989), 'Structures prosodiques du suédois et du français. Profils temporels et configurations tonales.', *Lund Working Papers* 21.
- Touratier, C. (2000), *La sémantique*, Armand Colin : Paris.
- Tourville, J. A.; Reilly, K. J. & Guenther, F. H. (2008), 'Neural mechanisms underlying auditory feedback control of speech', *Neuroimage* 39(3), 1429-1443.
- Tran, V.-A.; Bailly, G.; Lœvenbruck, H. & Toda, T. (2010), 'Improvement to a NAM-captured whisper-to-speech system', *Speech Communication* 52 (4), 314-326.
- Tremblay, P. & Dick, A. S. (2016), 'Broca and Wernicke are dead, or moving past the classic model of language neurobiology', *Brain and Language* 162, 60 - 71.
- Tremblay, S.; Shiller, D. M.; Ostry, D. J. & others (2003), 'Somatosensory basis of speech production', *Nature* 423(6942), 866-869.
- Tsakiris, M.; Schütz-Bosbach, S. & Gallagher, S. (2007), 'On agency and body-ownership: Phenomenological and neurocognitive reflections', *Consciousness and cognition* 16(3), 645-660.
- Tulving, E. (1972), *Episodic and semantic memory*, New York: Academic, 381-403.
- Tulving, E. (1983), *Elements of episodic memory*, Clarendon Press, Oxford.
- Tulving, E. (2002), 'Episodic Memory: From Mind to Brain', *Annual Review of Psychology* 53(1), 1-25.
- Turgeon, C.; Prémont, A.; Trudeau-Fiset, P. & Ménard, L. (2015), 'Exploring consequences of short- and long-term deafness on speech production: A lip-tube perturbation study', *Clinical Linguistics & Phonetics* 29(5), 378-400.
- Ueno, T.; Saito, S.; Rogers, T. & Lambon Ralph, M. (2011), 'Lichtheim 2: Synthesizing Aphasia and the Neural Basis of Language in a Neurocomputational Model of the Dual Dorsal-Ventral Language Pathways', *Neuron* 72(2), 385 - 396.
- Vaissière, J. (1991), *Rhythm, accentuation and final lengthening in French*, Springer, 108-120.
- Verluyten, S. P. (1982), *Recherches sur la prosodie et la métrique du français*, PhD thesis, Universitaire Instelling Antwerpen, Belgium.
- Vicente, A. & Martínez-Manrique, F. (2016), 'The nature of unsymbolized thinking', *Philosophical Explorations* 19(2), 173-187.
- Vihman, M. & Croft, W. (2007), 'Phonological development: Toward a 'radical' templatic phonology', *Linguistics* 45(4), 683-725.
- Vihman, M. M. (1996), *Phonological development: The origins of language in the child.*, Blackwell Publishing.
- Vilain, A.; Dole, M.; Lœvenbruck, H.; Pascalis, O. & Schwartz, J.-L. (sous presse), 'The role of production abilities in the perception of consonant category in infants', *Developmental Science*.
- Volterra, V.; Caselli, M. C.; Capirci, O. & Pizzuto, E. (2005), 'Gesture and the emergence and development of language', *Beyond nature-nurture: Essays in honor of Elizabeth Bates*, 3-40.
- Vuilleumier, P.; Armony, J. L.; Driver, J. & Dolan, R. J. (2001), 'Effects of Attention and Emotion on Face Processing in the Human Brain: An Event-Related fMRI Study', *Neuron* 30(3), 829 - 841.
- Vygotski, L. S. (1934/1997), *Pensée et langage*, Trad. française Françoise Sève. La Dispute, Paris.
- Warren, P.; Grabe, E. & Nolan, F. (1995), 'Prosody, phonology and parsing in closure ambiguities', *Language and Cognitive Processes* 10(5), 457-486.
- Watt Smith, T. (2015), *The Book of Human Emotions: An Encyclopedia of Feeling from Anger to Wanderlust*, Profile Books.
- Weber, A. & Scharenborg, O. (2012), 'Models of spoken-word recognition', *Wiley Interdisciplinary Reviews: Cognitive Science* 3(3), 387-401.
- Weiskrantz, L. (1985), *Animal intelligence: Proceedings of a Royal Society discussion meeting held on 6 and 7 June 1984*, Oxford, UK : Clarendon.
- Welby, P. & Lœvenbruck, H. (2006), 'Anchored down in Anchorage: Syllable structure and segmental anchoring in French', *Italian Journal of Linguistics/Rivista di linguistica* 18, 74-124.

- Welby, P. S. (2003), 'The slaying of Lady Mondegreen, being a study of French tonal association and alignment and their role in speech segmentation', PhD thesis, The Ohio State University.
- West, P. (2000), Long-distance coarticulatory effects of British English/l/and/r: An EMA, EPG and acoustic study, in 'Proceedings of the 5th Seminar on Speech Production: Model and Data', 105-108.
- Whiten, A. (2013), 'Humans are not alone in computing how others see the world', *Animal Behaviour* 86(2), 213 - 221.
- Whorf, B. L. (1956), *Language, thought, and reality: selected writings of Benjamin Lee Whorf*, Carroll, J. B. (ed.), Technology Press of MIT.
- Wiener, N. (1948), *Cybernetics: Control and communication in the animal and the machine*, (New York: John Wiley & Sons.
- van Wijk, C. & Kempen, G. (1987), 'A dual system for producing self-repairs in spontaneous speech: Evidence from experimentally elicited corrections', *Cognitive Psychology* 19(4), 403 - 440.
- Wiley, N. (2006), 'Pragmatism and the dialogical self', *International Journal for Dialogical Science* 1(1), 5-21.
- Wiley, N. (2006), 'Inner speech as a language: A Saussurean inquiry', *Journal for the Theory of Social Behaviour* 36(3), 319-341.
- Wiley, N. (2014), 'Chomskys Anomaly: Inner Speech', *International Journal for Dialogical Science* 8(1), 1-11.
- Wilkinson, S. & Fernyhough, C. (2017). 'Auditory verbal hallucinations and inner speech: A predictive processing perspective', *Before consciousness: in search of the fundamentals of mind*. Exeter: Imprint Academic, 285-304.
- Wilson, A. & Ross, M. (2003), 'The identity function of autobiographical memory: Time is on our side', *Memory* 11(2), 137-149.
- Wilson, M. (2002), 'Six views of embodied cognition', *Psychonomic bulletin & review* 9(4), 625-636.
- Wilson, M. & Knoblich, G. (2005), 'The case for motor involvement in perceiving conspecifics.', *Psychological bulletin* 131(3), 460.
- Wise, R. J. S.; Greene, J.; Büchel, C. & Scott, S. K. (1999), 'Brain regions involved in articulation', *The Lancet* 353(9158), 1057 - 1061.
- Wittgenstein, L. (1953), *Philosophische Untersuchungen (Philosophical Investigations)*. Transl. G.E.M. Anscombe, New York: MacMillan Publishing Company. Traduction française : Recherches philosophiques, Paris, Gallimard, 2004.
- Wolff, F. (1997), *Dire le monde*, Paris, Presses Universitaires de France, 2ème édition 2004.
- Wolpert, D. M. (1997), 'Computational approaches to motor control', *Trends in cognitive sciences* 1(6), 209-216.
- Wolpert, D. M. & Kawato, M. (1998), 'Multiple paired forward and inverse models for motor control', *Neural Networks* 11(7-8), 1317-1329.
- Wolpert, D. M.; Doya, K. & Kawato, M. (2003), 'A unifying computational framework for motor control and social interaction', *Philosophical Transactions of the Royal Society of London B: Biological Sciences* 358(1431), 593-602.
- Wolpert, D. M.; Ghahramani, Z. & Jordan, M. I. (1995), 'An internal model for sensorimotor integration', *Science-AAAS-Weekly Paper Edition* 269(5232), 1880-1882.
- Yamaguchi, N. (2008), Markedness, Frequency: Can We Predict the Order of Acquisition of Consonants?, in 'LingO 2007, the second Oxford Postgraduate Conference in Linguistics', 236-243.

ANNEXES

CURRICULUM VITÆ DÉTAILLÉ

Hélène LÖEVENBRUCK

Née le 9 mars 1968 à Paris XII – Nationalité française – 2 enfants

Chargée de Recherche 1^{ère} classe (CR1) CNRS

Institut des Sciences Humaines et Sociales (InSHS), section 34 (Sciences du Langage)

Helene.Loevenbruck@univ-grenoble-alpes.fr

<http://lpnc.univ-grenoble-alpes.fr/Helene-Loevenbruck>

Affiliation :

D'octobre 1998 à août 2012 :

Institut de la Communication Parlée, devenu Département Parole & Cognition de GIPSA-lab, CNRS UMR 5216, Grenoble INP, Université Grenoble Alpes

Depuis septembre 2012 (avec rattachement officiel en octobre 2013) :

Laboratoire de Psychologie et NeuroCognition, CNRS UMR 5105, Université Grenoble Alpes

FORMATION

1992	Ingénieure ENSERG, <i>section Architecture des Systèmes</i>	ENSERG/ENSIMAG, INP Grenoble
1992	DEA SCIENCES COGNITIVES	INP Grenoble
1994	Stage de 2 mois au <i>Laboratoire de Contrôle Moteur</i>	<i>Université McGill, Montréal, Canada</i>
1996	Doctorat SCIENCES COGNITIVES (Bénéficiaire bourse Ministère E.S.R., 1992-95)	INP Grenoble
1997	DEA SCIENCES du LANGAGE, <i>Option Phonétique et Parole</i>	Univ. Stendhal, Grenoble
1997-98	Post-doctorat : bénéficiaire bourse <i>Fyssen</i> <i>Relations articulation/structure prosodique</i>	Dept of linguistics, <i>the Ohio State Univ., USA</i>

PRIX ET DISTINCTIONS

2006	Médaille de bronze CNRS 2006, section 45 (cognition, langage)
2007	Finaliste du Prix Excellencia

ENSEIGNEMENT

1993-95	Vacataire TD Mathématiques (50 h/an), TP Informatique (64 h/an)	IUT 1, Univ. J. Fourier, Grenoble 1
1995-97	ATER Langage C (18h/an), Algo. (30h/an), Génie Logiciel (228h/an)	ENSERG (INPG), CNU 61
1999-01	Tutrice TD Licence Sc. du langage, Phonétique expérimentale. (22h/an)	Univ. Stendhal, Grenoble 3
2006--	Master 2R IC2A, Sc. Cognitives : Intro. Sciences Cognitives (4h/an)	INP Grenoble
2008-13	Master 2R IC2A, Sc. Cognitives : Communication Parlée (9h/an)	INP Grenoble
2009--	Master 2R IC2A, Sc. Cognitives : Compétences précoces (9h/an)	INP Grenoble
2013-14	Master 1R Psychologie : Psycholinguistique – Sc. du langage (15h)	UPMF, Grenoble 2
2015-16	Master 1R Sciences du langage : Phonétique et Cognition (12h)	Univ. Stendhal, Grenoble 3
2017-18	License ETC Phonétique et Phonologie (24h)	Université Grenoble Alpes

ENCADREMENT

Post-doctorants	Co-encadrement de 3 post-doctorants (Pauline Welby 2003-06, Francesca Carota 2004-06, Romain Grandchamp 2014-18)
Doctorants	Co-direction de 12 thèses de doctorat, dont 5 en cours et 5 soutenues (Marc Sato, 2004, chercheur au CNRS depuis 2007 ; Marion Dohen, 2005, Maître de conférences à Grenoble-INP depuis 2006 ; Viet-Anh Tran, 2010, ingénieur R&D chez Deezer depuis 2015 ; Julia Monnin, 2010, orthophoniste, Nouméa ; Lucile Rapin, 2011, <i>Clinical Research Coordinator</i> , MNI, McGill University, Canada depuis 2017)
Stages M2	Co-direction de 21 étudiants en Master 2 ^{ème} année ou équivalent
Stages M1	Co-direction de 12 étudiants en Master 1 ^{ère} année ou équivalent

ACTIVITÉS EDITORIALES

- **Rédactrice en chef** de la revue de Sc. Cognitives *In Cognito – Cahiers Romains de Sciences Cognitives* - <http://www.in-cognito.net>.
- **Relectures pour revues internationales** : Gesture, Italian Journal of Linguistics, International J. of Behav. Development, Interaction Studies, In Cognito, Journal of Phonetics, Language and Speech, Revue Canadienne de Linguistique, Schizophrenia Bull., Speech Communication
- **Relectures pour congrès et colloques** : Audio Visual Speech Processing, Australasian International Conference on Speech Science and Technology, International Congress of Phonetic Sciences, Conference on Laboratory Phonology, International Seminar on Speech Production, Interspeech, Journées d'Etude sur la Parole, Journées de Phonétique Clinique, Rencontres des Jeunes Chercheurs en Parole
- **Relectures pour ouvrage** : Neuropsycholinguistic perspectives on language cognition - Essays in honour of Jean-Luc Nespoulous

EXPERTISES, JURYS

- **Agence Nationale de la Recherche**, SHS, 2006, 2010
- **Agence d'évaluation de la recherche et de l'enseignement supérieur (AERES)** – Sciences du Langage, 2008
- **A*Midex** : appel à projets « Pépinière d'excellence » de l'index de l'Université d'Aix-Marseille, 2018
- **Comité Français d'Évaluation de la Coopération Universitaire et Scientifique avec le Brésil**, 2007
- **Centre National de la Recherche Scientifique** – PEPS, 2009; Concours IR, 2008 ; PIR Neuroinformatique et neurosciences computationnelles (NEURO IC) 2010
- **Délégation Régionale à la Recherche Clinique** du CHU de Grenoble (CPRB), 2007, 2011
- **Comités de sélection** – Univ. Toulouse Le Mirail, linguistique, section 07, 2011 ; Univ. Claude Bernard Lyon 1, sections 69/16, 2015
- **Jury de thèse en phonétique**, Université Sorbonne Nouvelle Paris 3, 2014
- **Jury de thèse en sciences du langage**, Université de Toulouse - Jean Jaurès, 2017, 2019
- **Comité de suivi de thèse en neurosciences cognitives**, Université Claude Bernard, Lyon, 2016
- **Jurys du Master National Recherche de Neuropsychologie** (2010-11, 2013, 2016, 2018), **Master Psychologie Cognitive et sociale** de Grenoble (2016, 2017), du **Master de Neuropsychologie de l'enfant** de Grenoble (2018)

ACTIVITÉS COLLECTIVES

- **Membre élue du Conseil de Laboratoire** de GIPSA-lab (jusqu'en 2012) et du **LPNC** depuis février 2016
- **Co-responsable de l'équipe** Parole Cerveau Multimodalité Développement de GIPSA-lab, 2007-11

- **Responsable de l'équipe** Langage du LPNC depuis octobre 2017
- **Responsable adjointe du master** (IC2A spécialité) **Sciences Cognitives** depuis 2006
- **Membre du conseil scientifique du PRASC** (Pôle Rhône-Alpes de Sciences Cognitives), 2003-10
- **Membre du comité de pilotage de l'ARC 2 Rhône-Alpes** Qualité de vie et vieillissement, axe 1b, 2011-17

PARTICIPATIONS A DES CONTRATS DE RECHERCHE

- **Projet CEE ESPRIT/BRA 1992-1995, No. 6975**
Speech Maps Sound-to-gesture inversion in speech
Membre participante
- **Projet APN CNRS 1999-2001, No. 7904**
De la prosodie à la syntaxe : Recherche de patrons d'activation cérébrale par IRM fonctionnelle
Responsable
- **Projet ACI Jeunes Chercheurs 2002, No. 0220244**
Caractérisation physiologique et perceptive de l'alignement tonal
Membre participante
- **Projet International Jacques Cartier 2002, No. AB / CB / 216/02**
Les effets articulatoires et acoustiques de l'accent de focalisation : une étude développementale
Co-responsable
- **Projet International Paidologos NIDCD 2003-présent**
Paidologos : cross-language investigation of phonological development
(<http://www.ling.ohio-state.edu/~edwards/>)
Membre participante
- **Contrat Fédération ELESA 2005**
Décodage de la prosodie en parole silencieuse
Responsable
- **Projet BQR INPG 2006**
LA VOIX DU SILENCE : Prosodie de la parole silencieuse et du murmure inaudible -- EMG du larynx et stéthoscopie maxillo-faciale
Co-responsable
- **Projet PHC Sakura 2008-2009**
CASSIS : Computer-Assisted communication and Silent Speech InterfaceS
Membre participante
- **Projet Région Rhône-Alpes, Cluster 11 Handicap, Vieillesse, Neurosciences 2008-2010**
Réorganisation fonctionnelle du langage et de la mémoire chez les patients épileptiques, avant et après intervention chirurgicale
Membre participante
- **Projet Région Rhône-Alpes, Cluster 11 Handicap, Vieillesse, Neurosciences 2009-2011**
Traces orofaciales des hallucinations auditives chez les schizophrènes : électromyographie et électroglottographie
Responsable
- **Projet International Jacques Cartier Grant 2009 – France – Canada**
Étude des mécanismes cognitifs des hallucinations auditives verbales chez les schizophrènes
Responsable
- **Projet ANR - DFG 2011-2013**
CROSSMOD: Représentation inter-modale de la langue et du visage chez des nourrissons allemands et français
Membre participante
- **Projets SFR Santé-Société Grenoble 2011-12, 2012-13**
Étude de la compensation potentielle des déficits de langage par l'utilisation de gestes manuels communicatifs chez les enfants
Membre participante

- **Projet ARC2 Région Rhône-Alpes –2013-2016**

Troubles de la parole chez les enfants sourds porteurs d'implants cochléaires : évaluation, diagnostic et pistes pour la remédiation

Membre participante

- **Projet ANR –2014-2018 (48 mois) : ANR-13-BSH2-0003-01**

INNERSPEECH : Corrélats neuraux de la parole intérieure

Responsable

- **Projet PEPS – Université de Grenoble –2014-2015**

Embodied emotion and speech Monitoring Grenoble (EMG)

Membre participante

- **Projets CMIRA-COOPERA –2014-2016**

Unités représentationnelles en parole : Psycholinguistique, développement et modélisation

Membre participante

- **Projet ERC Speech Unit(e)s –2014-2019**

The multisensory-motor unity of speech: Understanding how speech unites the sensory and motor streams, and how speech units emerge from perceptuo-motor interactions

Membre participante

- **Projet SFR Santé-Société Grenoble 2015**

Evaluation et dépistage des troubles du développement des sons de parole

Responsable

- **Projet IUF Monica Baciu – 2015-20**

Language Recovery and Plasticity in patients with Non-fluent Aphasia

Membre participante

- **Projet ARC2 région Rhône-Alpes –2016-2019**

Évaluation et dépistage des troubles du développement des sons de parole

Responsable

- **Projet ANR –2017-2021 (48 mois) : ANR-17-CE28-0006-05**

HearCog : Restauration des fonctions auditives et cognitives chez l'enfant sourd implanté cochléaire

Membre participante

- **Projet RSC IDEX Université Grenoble Alpes 2018 (12 mois)**

Les Langues de Babylab

Responsable

- **Projet SFR Santé-Société Grenoble 2018**

EULALIES : évaluation et dépistage des Troubles du Développement des Sons de Parole (TDSP) chez les enfants francophones

Membre participante

- **Projet CNRS-INS2I PICS (projet international de coopération scientifique) 2018-20**

EULALIES: Evaluation des Troubles du Développement des Sons de Parole

Membre participante

- **Projet ANR - DFG 2018-2021 (36 mois) : ANR-17-FRAL-CE28-0006-05**

BabyFaceSpeech : Regarde qui parle : interaction entre le traitement des visages et la perception de la parole au cours de la première année de vie chez des nourrissons mono- et bilingues Allemands et Français.

Membre participante

- **Projet IRS IDEX Université Grenoble Alpes 2018-19 (12 mois)**

CL3i : Le contexte linguistique : modèles théoriques, fondements cognitifs, opérativité informatique

Membre participante

- **Projet IRS IDEX Université Grenoble Alpes 2018-21 (36 mois)**

BEATBOX : Le Human Beatbox, un langage musical au service de la rééducation orthophonique

Membre participante

- **Projet RSC IDEX Université Grenoble Alpes 2019 (12 mois)**

LMBabylab : Langues et Musique de Babylab

Responsable

ORGANISATION DE COLLOQUES, SEMINAIRES, JOURNEES D'ETUDE, ECOLES D'ETE

- **Journées Prosodie 2001** : Organisées à Grenoble par le CNRS, Octobre 2001
- **Séminaires** de l'ICP et du Département Parole et Cognition (GIPSA-lab), 2002-2009.
- **Colloque du PRASC** : « Mais qu'est-ce qu'on a dans la tête ? Les sciences de la cognition face à la question de l'esprit ». Organisé à Grenoble by le PRASC (Pôle Rhône-Alpes de Sciences Cognitives), Mai 2003
- **Colloque du PRASC « Cognition en Rhône-Alpes »**: Organisé à Grenoble par le PRASC (Pôle Rhône-Alpes de Sciences Cognitives), Décembre 2004
- **VOCOID Workshop** : VOcalisation, COmmunication, Imitation and Deixis in infant and adult human and non-human primates. Organisé par GIPSA-lab et le Centre PsyCLE (Université de Provence), à Grenoble, Mai 2007
- **Brain and Speech Summerschool 2007**: "Cognitive and physical models of speech production, perception and perception-production interaction. Part II: Brain and Speech", Organisé par GIPSA-lab et ZAS de Berlin (Humboldt Univ.), à Autrans, France, Sept. 2007
- **Journée d'étude** sur les méthodes émergentes de traitements des données IRM, organisée par le groupe grenoblois NICOSIA (NeuroImaging COgnition Signal and Image Analysis), janvier 2011.
- **Célébration des 20 ans du Master de Sciences Cognitives** de Grenoble, organisé par le Master, Grenoble, juillet 2012.
- **Séminaires de Sciences Cognitives** organisés conjointement par le master IC2A spécialité Sciences Cognitives et le Pôle Grenoble Cognition depuis l'automne 2012.
- **11ème Colloque de la Société des Neurosciences**, Lyon-Grenoble, 21-24 mai 2013
- **Workshop international "Inner Speech"**, Grenoble, 7 juin 2016
- **Journée Scientifique du LPNC** : « Inner language forms, functions and dysfunctions », Grenoble, 23 juin 2017
- **Ateliers de formation et Journée Scientifique internationale** : « Développement de l'articulation : Du développement typique à la Dyspraxie Verbale », Grenoble, 22-25 novembre 2017

COLLABORATIONS

• Locales

Centre référent des troubles du langage et des apprentissages scolaires, CHU Grenoble
GIPSA-lab, Département Parole & Cognition, UMR CNRS 5216, Grenoble
Grenoble Institut des Neurosciences, Unité INSERM U836
Laboratoire Inter-universitaire de Psychologie, Personnalité, Cognition, Changement Social, Grenoble
Pôle de Neurologie et de Psychiatrie – CHU Grenoble
Service d'OtoRhinoLaryngologie – CHU Grenoble
UMS IRMaGe, Grenoble

• Nationales

Centre de Neuroscience Cognitive, Lyon, Bron, UMR 5229
Centre de Recherche Cerveau & Cognition, Toulouse, UMR CNRS 5549
INSERM U1028, Brain dynamics and cognition, Lyon
Laboratoire Dynamique Du Langage, Lyon, UMR CNRS 5596
Laboratoire Parole et Langage, Aix-en-Provence, UMR CNRS 6057
Laboratoire Psychologie de la Perception, Paris
NICOSIA, Grenoble working group on NeuroImaging COgnition Signal and Image Analysis
Transcultures, EA 3327, Université de la Nouvelle-Calédonie, Nouméa

• Internationales

Cognitive Neuropsychiatry of Schizophrenia Lab, Vancouver, Canada
Department of Linguistics, the Ohio State University, Columbus, OH, USA

Department of Linguistics, Memorial University of Newfoundland, St John, Canada

Justus-Liebig University Giessen, Germany

Laboratoire de Phonétique, Université du Québec à Montréal, Canada

Laboratoire d'acquisition bilingue, Université de Montréal, Canada

NeuroLab, Ferrara University, Italie

Universidade Federal de Parana, Brésil

Université de Liège, Belgique

SELECTION DE PUBLICATIONS

181 articles (34 articles de revues avec comité de lecture, 7 chapitres d'ouvrages, 3 commentaires, 137 articles ou résumés de congrès)

Articles de recherche dans revues avec comité de lecture et contributions à des ouvrages de recherche :

Sous presse

Haldin, C., Acher, A., Kauffmann, L., Hueber, T., Cousin, E., Badin, P., Perrier, P., Fabre, D., Perennou, D., Detante, O., Jaillard, A., Løevenbruck, H. & Baciú, M. (sous presse). Effet de la rééducation perceptivo-motrice sur la récupération de la parole chez deux patientes avec aphasie non fluente chronique post-AVC. *Revue de Neuropsychologie*.

Nalborczyk, L., Batailler, C., Løevenbruck, H., Vilain, A. & Bürkner, P.-C. (sous presse). An Introduction to Bayesian Multilevel Models Using brms: A Case Study of Gender Effects on Vowel Variability in Standard Indonesian. *Journal of Speech, Language, and Hearing Research*.

2018

Kubicek, C., Gervain, J., Løevenbruck, H., Pascalis, O., & Schwarzer, G. (2018). Goldilock vs. Goldlöckchen: Visual speech preference for same-rhythm-class languages in 6-month-old infants. *Infant and Child Development*. doi: 10.1002/icd.2084 - **hal-01741631**

Løevenbruck, H. (2018). What the neurocognitive study of inner language reveals about our inner space. *Épistémocritique, n° 18 : Langage intérieur - Espaces intérieurs / Inner Speech - Inner Space*, Stéphanie Smadja, Pierre-Louis Patoine (eds.) [<http://epistemocritique.org/what-the-neurocognitive-study-of-inner-language-reveals-about-our-inner-space/>] - **hal-02039667**

Løevenbruck, H., Grandchamp, R., Rapin, L., Nalborczyk, L., Dohen, M., Perrier, P., Baciú, M. & Perrone-Bertolotti, M. (2018). A cognitive neuroscience view of inner language: to predict, to hear, to see and to feel. In *Inner Speech: New Voices*. Peter Langland-Hassan & Agustín Vicente (eds.), Oxford University Press, 131-167 - **hal-01898992**

2017

Grandon, B., Vilain, A., Løevenbruck, H., Schmerber, S. & Truy, E. (2017). Realisation of voicing by French-speaking CI children after long-term implant use: An acoustic study. *Clinical Linguistics & Phonetics*, 31, 1-14. doi: 10.1080/02699206.2017.1302511 - [PMID: 28362227](#) - **halshs-01500526-1**

Haldin, C., Acher, A., Kauffmann, L., Hueber, T., Cousin, E., Badin, P., Perrier, P., Fabre, D., Perennou, D., Detante, O., Jaillard, A., Løevenbruck, H. & Baciú, M. (2017). Speech recovery and language plasticity can be facilitated by Sensori-Motor Fusion (SMF) training in chronic non-fluent aphasia. A case report study. *Clinical Linguistics & Phonetics*, Taylor & Francis, 2017, pp.1 - 27. <10.1080/02699206.2017.1402090> - **hal-01651198**

Løevenbruck, H., Nalborczyk, L., Perrone-Bertolotti, M., Baeyens, C., Grandchamp, R. (2017). Un cas d'incarnation linguistique : implication du système moteur orofacial dans les ruminations mentales. *La cognition incarnée*, J. Dokic et D. Perrin (dir.), Recherches sur la philosophie et le langage, N° 33, Vrin, Paris, 297-324, ISBN : 978-2-7116-8423-6. - **hal-01887636**

Nalborczyk L., Perrone-Bertolotti M., Baeyens C., Grandchamp R., Polosan M., Spinelli E., Koster E. H. W. & Løevenbruck H. (2017). Orofacial electromyographic correlates of induced verbal rumination. *Biological Psychology*, 127, 53-63. - **hal-01910991**

Richoz, A.-R., Quinn, P. C., Hillairet de Boisferon, A., Berger, C., Løevenbruck, H., Lewkowicz, D. J., Lee, K., Dole, M., Caldara, R. and Pascalis, O. (2017). Audio-Visual Perception of Gender by Infants Emerges Earlier for Adult-Directed Speech. *PLOS ONE*, 12, 1-15. doi:10.1371/ journal.pone.0169325 - [PMCS218491](#) - **hal-01519114**

2016

- Acher, A., Fabre, D., Hueber, T., Badin, P., Detante, O., Cousin, E., Pichat, C., Løevenbruck, H., Haldin, C. & Baciú, M. (2016). Retour visuel en aphasiologie : résultats comportementaux, acoustiques et en neuroimagerie. Nathalie Joyeux, Silvia Topouzkhianian. XVIèmes Rencontres Internationales d'Orthophonie. Orthophonie et technologies innovantes, Décembre 2016, Paris, France. *Ortho Edition*, 227-260. - **hal-01419292**
- Perrone-Bertolotti M., Grandchamp R., Rapin L., Baciú M., Lachaux J.P. & Løevenbruck H. (2016). Langage intérieur. In Pinto, S. & Sato, M. (eds.). *Traité de Neurolinguistique : du cerveau au langage*. Collection Neuropsychologie, De Boeck Supérieur, 109-123. - **hal-01467107**
- Rapin L., Dohen M. & Løevenbruck H. (2016). Les hallucinations auditives verbales. In Pinto, S. & Sato, M. (eds.). *Traité de Neurolinguistique Neurolinguistique : du cerveau au langage*. Collection Neuropsychologie, De Boeck Supérieur, 347-370. - **hal-01442463**

2015

- Hillairet de Boisferon, A., Dupierrix, E., Quinn, P.C., Løevenbruck, H., Lewkowicz, D.J., Lee, K., & Pascalis, O. (2015). Perception of Multisensory Gender Coherence in 6 and 9-month-old Infants. *Infancy*, 20 (6), 661-674. doi: 10.1111/inf.12088 - **hal-01203120**
- Lavigne K. M., Rapin L. A. ; Metzack P. D., Whitman J. C., Jung K., Dohen M., Løevenbruck H., Woodward T. S. (2015). Left-dominant temporal-frontal hypercoupling in schizophrenia patients with hallucinations during speech perception. *Schizophrenia Bulletin*, 41 (1), 259-267. doi:10.1093/schbul/sbu004 - **hal-00966620**

2014

- Kubicek, C., Gervain, J., Hillairet de Boisferon, A., Pascalis, O., Løevenbruck, H., & Schwarzer, G. (2014). The influence of infant-directed speech on 12-month-olds' intersensory perception of fluent speech. *Infant Behavior & Development*, 37(4), 644-651. doi:10.1016/j.infbeh.2014.08.010. [PMID: 25238663](#) - **hal-1099560**
- Kubicek C., Hillairet de Boisferon A., Dupierrix E., Pascalis O., Løevenbruck H., Gervain J. & Schwarzer G. (2014). Cross-modal matching of audiovisual German and French fluent speech in infancy. *PLoS ONE* 9(2): e89275. doi:10.1371/journal.pone.0089275 - **hal-00964518**
- Pascalis O., Løevenbruck H., Quinn P. C., Kandel S., Tanaka J. W. & Lee K. (2014). On the Links Among Face Processing, Language Processing, and Narrowing during Development. *Child Development Perspectives*. 8 (2), 65-70. DOI: 10.1111/cdep.12064 - **hal-00964721**
- Perrone-Bertolotti, M., Rapin L., Lachaux J.-P., Baciú M. & Løevenbruck H. (2014). What is that little voice inside my head? Inner speech phenomenology, its role in cognitive performance and its relation to self-monitoring. *Behavioural Brain Research*, 261, 220-239. *Epub: 10.1016/j.bbr.2013.12.034*. - **hal-00967730**

2013

- Rapin L., Dohen M., Polosan M., Perrier P., Løevenbruck H. (2013). An EMG study of the lip muscles during covert auditory verbal hallucinations in schizophrenia. *Journal of Speech, Language, and Hearing Research*, 56, S1882–S1893. - **hal-00926630**
- Perrone-Bertolotti, M., Dohen, M., Løevenbruck, H., Sato, M., Pichat, C., Baciú, M. (2013). Neural correlates of the perception of contrastive prosodic focus in French. An fMRI study. *Human Brain Mapping*, 34 (10), 2574-2591. - **hal-00762363**, **PMID: 22488985**
- Kubicek C., Hillairet de Boisferon A., Dupierrix E., Løevenbruck H., Gervain J., and Schwarzer G. (2013). Face-scanning behavior to silently talking faces in 12-month-old infants: The impact of pre-exposed auditory speech. *International Journal of Behavioral Development*, 37 (2), 106-110. doi:10.1177/0165025412473016 - **hal-00864344**

2012

- Cathiard M.-A., Vilain A., Laboissière R., Løevenbruck H., Savariaux C. & Schwartz J.-L. (2012). Bimodal Perception within the natural time-course of speech production. In Bailly G., Perrier P. & Vatikiotis-Bateson E. (Eds.), *Audiovisual Speech Processing*. Cambridge University Press, 121-158. - **hal-00821829**
- Rapin, L., Metzack, P. D., Whitman, J. C., Dohen, M., Løevenbruck, H., & Woodward, T. S. (2012). Hyperintensity of functional networks involving voice-selective cortical regions during silent thought in schizophrenia. *Psychiatry Research: Neuroimaging*, 102 (2), 110-117. - **hal-00734043**, **PMID: 22703623**
- Schwartz J.-L. & Løevenbruck H. (2012). La parole, entre objets vocaux et objets du monde. In *La psychologie cognitive*. Collection Cogniprisme. M. Denis (ed.). Editions de la Maison des Sciences de l'Homme, ISBN 978-2-7351-1509-9, 165-199. - **hal-00782274**

2011

- Dohen M., Løevenbruck H., Roustan B., Vilain C. (2011). Pointage manuel et vocal : corrélats sensorimoteurs et neurocognitifs, coordination geste manuel / parole. *Faits de Langue*, 37, 97-116. - **hal-00784592**
- 2010**
- Tran V.-A., Bailly G., Løevenbruck H. & Toda T. (2010). Improvement to a NAM-captured whisper-to-speech system. *Speech Communication*. 52 (4), 314-326. - **hal-00459973**
- 2009**
- Dohen M. & Løevenbruck H. (2009). Interaction of audition and vision for the perception of prosodic contrastive focus. *Language and Speech*, 52 (3), 177-206. - **hal-00371180, PMID: 19624029**
- Dohen M., Løevenbruck H. & Hill H. (2009). Recognizing prosody from the lips: is it possible to extract prosodic focus from lip features? In Alan Wee-Chung Liew & Shilin Wang (Eds.), *Visual Speech Recognition: Lip Segmentation and Mapping*, Medical Information Science Reference, Hershey, New York, ISBN 978-1-60566-186-5, 416-438. - **hal-00360765**
- Heracleous P., Beautemps D., Tran V.-A., Løevenbruck H. & Bailly G. (2009). Exploiting visual information for NAM recognition. *IEICE Electronics Express*, 6, 2, 77-82. - **hal-00357985**
- Løevenbruck H., Dohen M. & Vilain C. (2009). Pointing is 'special'. In Fuchs S., Løevenbruck H. Pape D. & Perrier P. (eds.), *Some Aspects of Speech and the Brain*, Peter Lang, 211-258, ISBN 978-3-631-57630-4. - **hal-00360758**
- 2008**
- Løevenbruck H., Vilain C. & Dohen M. (2008). From gestural pointing to vocal pointing in the brain. *Revue Française de Linguistique Appliquée*, Vol. XIII, 2, 23-33. - **hal-00360753**
- 2007**
- D'Imperio M., Espesser R., Løevenbruck H., Menezes C., Nguyen N. & Welby P. (2007). Are tones aligned to articulatory events? Evidence from Italian and French. Cole, Jennifer and José I. Hualde (eds.), *Laboratory Phonology 9*, Berlin: Mouton de Gruyter, 577-608. - **hal-00244489**
- 2006**
- Ménard L., Løevenbruck H. & Savariaux C. (2006). Articulatory and acoustic correlates of contrastive focus in French: a developmental study, in Harrington, J. & Tabain, M. (eds), *Speech Production: Models, Phonetic Processes and Techniques*, Psychology Press : New York, 227-251. - **hal-00371878**
- Sato M., Schwartz J.-L., Abry C., Cathiard M.-A., & Løevenbruck H. (2006). Multistable syllables as enacted percepts: A source of an asymmetric bias in the verbal transformation effect. *Perception & Psychophysics*, 68 (3), pp. 458-474. - **hal-00186572, PMID: 16900837**
- Welby P. & Løevenbruck 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.). - **hal-00371892**
- 2005**
- Løevenbruck H., Baciú M., Segebarth C. & Abry C. (2005). The left inferior frontal gyrus under focus: an fMRI study of the production of deixis via syntactic extraction and prosodic focus. *Journal of Neurolinguistics*, 18, 237-258. - **hal-00371865**
- 2004**
- Dohen M., Løevenbruck H., Cathiard M.-A. & Schwartz J.-L. (2004). Visual perception of contrastive focus in reiterant French speech. *Speech Communication*, 44, 155-172. - **hal-00370938**
- Sato M., Baciú M., Løevenbruck H., Schwartz J.-L., Segebarth C., Cathiard M.-A. & Abry C. (2004). Multistable representation of speech forms: a functional MRI study of verbal transformations. *NeuroImage*, 23, 1143-1151. **PMID: 15528113**
- 1999**
- Løevenbruck H., Collins M.J., Beckman M.E., Krishnamurthy A.K. & Ahalt S.C. (1999). Temporal coordination of articulatory gestures in consonant clusters and sequences of consonants. *Proceedings of Linguistics Phonetics 1998, Item Order in Language and Speech*. Fujimura O., Joseph B. D. & Palek B. (eds.), Charles University, Prague, The Karolinum Press. Vol II: 547-573. - **hal-00371717**
- 1996**
- Perrier P., Løevenbruck H. & Payan Y. (1996). Control of the tongue movements in speech: The Equilibrium point Hypothesis perspective. *Journal of Phonetics*, 24, 53-75. - **hal-00080443**

INVITATION A DES CONGRES, TABLES RONDES OU CONFERENCES A L'ETRANGER

En France :

Aix-en-Provence (2004, 2015), Grenoble (1998, 2001, 2007, 2008, 2009, 2011, 2012, 2013, 2015, 2016, 2017, 2018), Lille (2018), Lyon (2004, 2005, 2008, 2014), Paris (2013, 2014, 2015, 2016, 2018)

A l'étranger :

- Løevenbruck, H. (1998). A motor control approach to the role of prosody in 'on-line alien-monitoring'. Département de linguistique de l'Université de l'Ohio, Columbus, OH, USA.
- Løevenbruck, H. (2009). Some articulatory and cerebral correlates of prosodic focus in French. Workshop on "Experimental Studies on Intonation: Phonetic, Phonological and Psycholinguistic Aspects of Sentence Prosody". 6 January 2009, Potsdam, Allemagne.
- Løevenbruck, H. (2012). Advantages and disadvantages of the picture-prompted repetition task. Participation in the panel discussion on "Using corpora to study size effects » within the symposium « Ohio State University Research Collective on Sound Size Effect », Columbus, OH, USA
- Løevenbruck, H. (2015). Neurocognitive recordings of inner language. Workshop mind-brain project - University of Navarra, Pamplona, Spain, November 13th, 2015.
- Løevenbruck, H. (2018). Cognitive neuroscience tapping inner language: prediction and perception - Public lecture organised by the Hearing the Voice project, Durham University, England, 22 March 2018 (<https://hearingthevoice.org/2018/03/01/cognitive-neuroscience-tapping-inner-language-prediction-and-perception-22-march-2018/>)

VALORISATION SOCIALE ET CULTURELLE DES ACTIVITES DE RECHERCHE

- **Conférences grand public** : Grenoble et sa région (2007, 2012, 2014, 2015, 2018), Paris (2015, 2018), Strasbourg (2018), Vincennes (2005)
- **Projets Art-Science** :
 - Du Pointage au Langage (2012) : une des Histoires Courtes de JF Dars & A. Papillault (<https://vimeo.com/328843449>)
 - Participation à l'élaboration du spectacle « Fromage de tête », sur les mécanismes de la pensée, avec les n+1, les Ateliers du Spectacle, 2012-14
 - Des voix dans la tête : Impromptu scientifique sur la parole intérieure, avec les n+1, les Ateliers du Spectacle, et représentations de 2014 à 2018 à Apt, Orsay, Paris et Istres.
 - Mise en scène du « Lorenzo circuit » de C. Fernyhough, Cerveaurium de R. Grandchamp, juin 2017
 - Le langage et son développement – Présentation du projet art-science Les Langues de Babylab et des premiers pas du spectacle Soliloc' avec Aurélie Maisonneuve à l'Espace Scénique Transdisciplinaire (EST), Université Grenoble Alpes
- **Interventions dans la Presse, Radio, Télévision**
 - Presse** : Journal du CNRS (2007), Dauphiné Libéré (2007), La Presse de Montréal (2014), Sciences et Vie Junior (2016), Sciences et Avenir (2017), Le Monde (2017)
 - Radio** : France Inter (2007, 2018), Radio Canada (2015)
 - Télévision** : TéléGrenoble (2007, 2009), Radio Canada TV (tournage 2016, diffusion 2017 : <http://ici.radio-canada.ca/tele/decouverte/2016-2017/segments/reportage/13785/voix>)
- **Participation à des animations grand public**
 - Stands ou animations à la **Fête de la Science** : ICP puis Département Parole et Cognition, puis LPNC, quasi-annuellement
 - Stands annuels au **Forum de Sciences Cognitives** : pour le Master de sciences cognitives de Grenoble, les laboratoires GIPSA-lab et LPNC, le Pôle Grenoble Cognition et la revue In Cognito
 - Membre du comité d'Organisation de la **Semaine du Cerveau** à Grenoble, depuis 2008

Sélection de Publications

Løevenbruck H., Baciú M., Segebarth C. & Abry C. (2005). The left inferior frontal gyrus under focus: an fMRI study of the production of deixis via syntactic extraction and prosodic focus. *Journal of Neurolinguistics*, 18, 237-258. doi:10.1016/j.jneuroling.2004.12.002 - hal-00371865

Løevenbruck H., Dohen M. & Vilain C. (2009). Pointing is 'special'. In Fuchs S., Løevenbruck H. Pape D. & Perrier P. (eds.), *Some Aspects of Speech and the Brain*, Peter Lang, 211-258, ISBN 978-3-631-57630-4. - hal-00360758

Rapin L., Dohen M., Polosan M., Perrier P., Løevenbruck H. (2013). An EMG study of the lip muscles during covert auditory verbal hallucinations in schizophrenia. *Journal of Speech, Language, and Hearing Research*, 56, S1882–S1893. PMID : 24687444 - hal-00926630

Løevenbruck, H., Nalborczyk, L., Perrone-Bertolotti, M., Baeyens, C., Grandchamp, R. (2017). Un cas d'incarnation linguistique : implication du système moteur orofacial dans les ruminations mentales. *La cognition incarnée*, J. Dokic et D. Perrin (dir.), *Recherches sur la philosophie et le langage*, N° 33, Vrin, Paris, 297-324, ISBN : 978-2-7116-8423-6. - hal-01887636

Løevenbruck, H., Grandchamp, R., Rapin, L., Nalborczyk, L., Dohen, M., Perrier, P., Baciú M. & Perrone-Bertolotti, M. (2018). A cognitive neuroscience view of inner language: to predict, to hear, to see and to feel. In *Inner Speech: New Voices*. Peter Langland-Hassan & Agustín Vicente (eds.), Oxford University Press, 131-167 - hal-01898992

THE LEFT INFERIOR FRONTAL GYRUS UNDER FOCUS:

AN FMRI STUDY OF THE PRODUCTION OF DEIXIS VIA SYNTACTIC EXTRACTION AND PROSODIC FOCUS

Hélène Løevenbruck^{a*}, Monica Baciú^b, Christoph Segebarth^c, Christian Abry^a

^aInstitut de la Communication Parlée, UMR CNRS 5009 / INPG / Université Stendhal, 46 av. Félix Viallet, 38031 Grenoble Cedex 01, France,

^bLaboratoire de Psychologie et Neurocognition / UMR CNRS 5105, Université Pierre Mendès-France, BP 47, 38040 Grenoble Cedex 09, France,

^cUnité Mixte INSERM / Université Joseph Fourier, U594 Neuroimagerie Fonctionnelle et Métabolique, LRC CEA 30V, Centre Hospitalier Universitaire, Pavillon B, BP 217, 38043 Grenoble Cedex 09, France.

Running title:

An fMRI study of the production of deixis

* Corresponding author:

Tel: + 33 4 76 57 47 14

Fax: + 33 4 76 57 47 10

e-mail: loeven@icp.inpg.fr

ABSTRACT

The left inferior frontal gyrus (LIFG, BA 44, 45, 47) has been associated with linguistic processing (from sentence- to syllable- parsing) as well as action analysis. We hypothesize that the function of the LIFG may be the monitoring of action, a function well adapted to agent deixis (verbal pointing at the agent of an action). The aim of this fMRI study was therefore to test the hypothesis that the LIFG is involved during the production of agent deixis.

We performed an experiment whereby three kinds of deictic sentences were pronounced, involving prosodic focus, syntactic extraction and prosodic focus with syntactic extraction.

A common pattern of activation was found for the three deixis conditions in the LIFG (BA 45 and/or 47), the left insula and the bilateral premotor (BA 6) cortex. Prosodic deixis additionally activated the left anterior cingulate gyrus (BA 24, 32), the left supramarginal gyrus (LSMG, BA 40) and Wernicke's area (BA 22).

Our results suggest that the LIFG is involved during agent deixis, through either prosody or syntax, and that the LSMG and Wernicke's area are additionally required in prosody-driven deixis. Once grammaticalized, deixis would be handled solely by the LIFG, without the LSMG and Wernicke's area.

KEYWORDS

Speech production, prosodic focus, syntactic extraction, agent deixis, Left Inferior Frontal Gyrus.

INTRODUCTION

The left inferior frontal gyrus (LIFG; Brodmann Areas (BA) 44, 45 and 47) has been the focus of numerous studies in different fields, ranging from neuropsychology to formal linguistics. Various interpretations as to the roles of the LIFG have thus arisen.

Broca's original hypothesis that the LIFG must be dedicated to the production of fluent, articulate speech (Broca, 1861) has been reconsidered in the 1960s when it was noticed that Broca's aphasics, in addition to having problems with speech production, also present difficulties in speech comprehension (Caramazza & Zurif, 1976). This observation has led the LIFG to be considered as the locus for syntax – both in speech production and in speech comprehension (Zurif, 1980). Broca's aphasics may remain capable of some syntactic processing, however, and they may fail only for specific syntactic constructions (such as for object cleft sentences, center-embedded relatives or for passive constructions; see for instance Grodzinsky, Piñango, Zurif & Draï, 1999 for a review). It was therefore concluded that the LIFG could not be considered as the locus for all aspects of syntax (Linebarger, Schwartz & Saffran, 1983).

Various functional roles have been assigned to the LIFG since then, on the basis of neuroimaging data and further observations on Broca's aphasics. Grodzinsky (Grodzinsky, 2000) has argued that the LIFG must have a highly restricted role, handling intrasentential dependency relations only (tracking of moved phrasal constituents in grammatical decoding and building of full-fledged syntactic trees in grammatical encoding). Other, less specific, roles in syntactic processing have been assigned to the LIFG. For instance, it has been considered to be involved in complex syntactic processing when thematic-role monitoring is required, i.e. the processing of "who-does-what-to-whom" (Just, Carpenter, Keller, Eddy & Thulborn, 1996; Caplan, Alpert, Waters & Olivieri, 2000). In agreement with this analysis, Friederici hypothesizes that the LIFG must be associated with the procedural syntactic knowledge, which depends on temporal parameters and on sequencing (Friederici, 1990; see also Friederici, 2002). This hypothesis is in line with the results from studies on sequence processing. LIFG activation has been observed during phoneme detection requiring sequence processing (Démonet, Price, Wise & Frackowiak, 1994). Certain data suggest that the LIFG is also involved in the processing of non-linguistic sequences (Penhune, Zatorre & Evans, 1998; Dominey & Lelekov, 2000), while other data, such as those on procedural learning in Broca's aphasia, seem to support the hypothesis of a linguistic specificity of the LIFG (Goschke, Friederici, Kotz & van Kampen, 2001).

The involvement of the LIFG in complex syntactic processing has been questioned by Stowe and colleagues (Stowe, Broere, Paans, Wijers, Mulder, Vaalburg *et al.*, 1998, Stowe, 2000; see also Hickok, 2000) who attribute the LIFG activations associated with sentential complexity to processes in working memory. They argue that the LIFG only supports temporary storage of verbal information (including information on the phrasal syntactic and semantic structure) during sentence processing.

In an *a priori* unrelated domain, the LIFG has been associated with the observation and mental imagery of actions such as object grasping (Grafton, Arbib, Fadiga & Rizzolatti, 1996) or finger motion (Iacoboni, Woods, Brass, Bekkering, Mazziotta & Rizzolatti, 1999; Binkofski, Amunts, Stephan, Posse, Schormann, Freund, *et al.*, 2000). Rizzolatti, Fogassi & Gallese (1997) have suggested that the LIFG must be linking recognition of motor action (understanding actions performed by others) and production of motor action, a prerequisite in the phylogenetic development of communication and speech. Grèzes & Decety (2001) argue however that some tasks having been considered to activate the LIFG (such as implicit motor imagery, complex manipulation of objects) seem rather to involve the premotor cortex. Also, Grèzes and colleagues (Grèzes, Costes & Decety, 1998) were unable to replicate Rizzolatti *et al.*'s activation of the LIFG during "observation-in-order-to-imitate", while they obtained its involvement during the observation of meaningful actions. They therefore conclude that the activations of the LIFG must rather reflect silent verbalization, and that the LIFG must be associated solely with speech processing. However, Buccino, Binkofski, Fink, Fadiga, Fogassi, Gallese *et al.* (2001) observed a somatotopic organization of the activations in the premotor cortex (including the LIFG) during the observation of motor actions performed with different effectors. The LIFG was activated during the observation of mouth and hand actions but not during the observation of foot actions. These results are not in line with the verbalization hypothesis, which entails that the LIFG should be activated during observation of the action, whatever the effector used.

On the basis of this abundant literature, we hypothesize that the role of the LIFG is that of an attentional parser of action, which is well adapted to linguistic processing. This action parser supports the attentional monitoring of thematic roles handled in morphosyntactic analysis and is involved in the spatial and temporal indexing of predicates (actions) and their arguments (patient, agent). If the LIFG is involved in linguistic action parsing, then deixis on the agent of an action – in the sense of verbal pointing at the agent (Levinson, 1983) – which requires thematic-role monitoring, should activate the LIFG. The aim of the present fMRI study was therefore to explore the cerebral activations due to the production of deictic sentences. Three types of deictic sentences were tested, involving agent deixis through either prosody or syntax, or through a combination of both [insert footnote 1].

MATERIALS AND METHODS

Subjects

Sixteen healthy, male, right-handed (Edinburgh Handedness Inventory, Oldfield, 1971) native speakers of French were examined. All subjects gave their informed consent for the fMRI examination. The study was performed in accordance with the institutional review board regulations.

Stimuli

The stimuli consisted of short sentences in French, visually presented in the middle of a projection screen. The following four isosyllabic sentences were presented, one for each condition:

(1) Baseline condition: “Madeleine m’amena”

(/ma.də.lən.ma.mə.na/, Madeleine brought me around),

(2) Prosodic deixis condition: “MADELEINE m’amena”

(MADELEINE brought me around),

(3) Syntactic deixis condition: “C’est Mad’leine qui m’am’na”

(/sɛ.ma.dlən.ki.ma.mna/, It’s Mad’leine who brought me ’round),

(4) Combined deixis condition (prosodic and syntactic deixis): “C’est MAD’LEINE qui m’am’na” (It’s MAD’LEINE who brought me ’round).

When the first name “Madeleine” was presented in capital letters, the subjects were instructed to use contrastive focus, as if they were correcting a wrong information communicated to them, such as: “Jennifer brought you around” (rather than Madeleine). In addition, it was made clear to the subjects that, when they were requested to produce the syntactic extraction construction, the latter was meant to point at the agent “Madeleine”, excluding all other possible agents (such as “Jennifer”).

The number of syllables in the sentence was maintained equal to 6, using schwa deletion. Each sentence was presented for 3 seconds at the beginning of the corresponding condition. Then a fixation mark, alternating every 3 seconds between a ‘+’ and a ‘x’ sign, appeared in the middle of the screen. This alternation was aimed at triggering the silent repetition (14 times per condition) of the sentence presented. The stimuli were generated by means of Psyscope V.1.1 (Carnegie Mellon Department of Psychology) running on a Macintosh computer (Power Macintosh 9600). They were transmitted to the subjects by means of a video projector (Eiki LC 6000), a projection screen situated behind the magnet and a mirror centered above the subject’s eyes.

Paradigm and tasks

A day before the experiment, the subjects were extensively trained with the 4 sentences listed above (over and over rehearsal of the 4 conditions). The subjects were positioned in front of a computer screen, instructed and trained to execute the tasks, first in an overt speech production mode, then in a covert mode.

Pre- and post-scan audio DAT recordings were carried out to estimate the subjects’ task performance during the fMRI scan. Subjects were prompted by exactly the same script as during the scans. They produced each of the sentences 4 times (instead of 14 times per condition in the actual experiment). For the pre-scan recording, they were instructed to speak aloud, at a comfortable speaking rate. For the post-scan recording, the instruction was to speak aloud and to produce the same intonation patterns they had mentally produced during the scans.

Pre- and post-scan acoustic recordings were assessed using the *Praat* software (Boersma, 2001). Duration and fundamental frequency (F0) measurements were semi-automatically carried out on each of the 256 utterances (16 subjects, 4 repetitions of each of the 4 conditions). For the duration analysis, the beginning and end of each utterance were detected from the spectrogram using classical phonetic criteria (onset of voicing for /ma/, onset of noise for /ε/, offset of voicing for /na/). For the intonation analysis, peak F0 values were automatically measured using a peak-detection algorithm on the F0 traces provided by the *Praat* software.

Three functional scans were performed during each fMRI session. A block paradigm was used. A scan comprised eight epochs (each condition was repeated once) of 42 seconds each. The order of presentation of the four conditions was alternated between scans and between subjects. Subjects were instructed to silently read the sentence presented at the beginning of each condition and to repeat it, using inner speech, at each alternation of the fixation cross. Thus, during each epoch, the specific sentence was repeated 14 times.

MR acquisition

Functional MR imaging was performed on a 1.5 Tesla MR imager (Philips NT) with echo-planar (EPI) acquisition. Twenty-five adjacent, axial, slices (5 mm thickness each) were imaged 10 times during each epoch. The imaging volume was oriented parallel to the bi-commissural plane. It encompassed the whole brain and the upper part of the cerebellum. An EPI MR pulse sequence was used for the functional scans. The major MR parameters of this sequence were: TR = 3700 ms, TE = 45 ms, pulse angle = 90°, acquisition matrix = 64x64, reconstruction matrix = 128x128, field-of-view = 256x256 mm². Between the first and the second functional MR scans, a high-resolution 3D anatomical MR scan was obtained from the volume functionally examined.

Data processing

Data analysis was performed using the SPM-99 software (Statistical Parametric Mapping - Wellcome Department of Cognitive Neurology, London, UK) running on a Unix workstation under the MATLAB environment (Mathworks, Sherbon, USA).

Functional MR images were subjected to the following pre-processing steps. In a first step, motion correction was applied. All images within a functional scan were realigned by means of a rigid body transformation. Then, the anatomical volume was spatially normalized into a reference space using as template a representative brain from the Montreal Neurological Institute series. The spatial normalization parameters were subsequently applied to the set of functional images. Finally, to conform to the assumption underlying SPM that the data are normally distributed, and to allow for some inter-subject variability during group analysis, the functional images were spatially smoothed.

Statistics

Contrasts between conditions were determined pixelwise using the General Linear Model. Statistical significance threshold for individual pixels was established at $p = 0.001$. Clusters of activated pixels were then identified, based on the intensity of the individual responses and the spatial extent of the clusters. Finally, a significance threshold of $p = 0.05$ (corrected for multiple comparisons) was applied for identification of the activated clusters. The results of the fixed effect group and random effects analyses are reported here.

RESULTS

Audio results

An example of analysis of the acoustic recordings with the *Praat* software is given in Figure 1, which displays the acoustic waveform (top panel), the spectrogram with the superimposed fundamental frequency (F0) curve (middle panel) and the duration and intonation labelling (bottom panels).

Insert Figure 1 about here

The mean utterance duration was 1139 ms (standard deviation: 159 ms) before and 1108 ms (standard deviation: 129 ms) after the scans. The mean sentence durations for each condition are shown in Table 1.

Insert Table 1 about here

An example of intonation analysis is provided in Figure 1 for one repetition of the prosodic deixis sentence by speaker DB. The rise towards the high pitch accent [insert footnote 2], corresponding to the F0 peak on the first syllable /ma/, is labelled as LHF (Low High sequence with focus). The post-focal F0 trace falls to a flat floor (the fall is labelled as two L%, i.e. two low Accentual Phrase boundary tones), which is a typical feature of French focus realizations (see Jun & Fougeron, 2000 for a model of focus intonation in French). Different subjects made different choices as to the syllable bearing the high pitch accent in the prosodic deixis and combined deixis conditions, but their choices were maintained between measurements. To be more specific, in the prosodic deixis condition, the focused constituent being /ma.d↔.lɛn/, three syllables were possible slots for the high pitch accents. In the combined deixis condition, 2 slots were available (/ma.dlɛn/). Overall, although inter-speaker variability was observed, no significant intra-speaker variability was detected between recordings. For instance, in the 4 repetitions of the prosodic deixis condition, 5 subjects put a high pitch accent on the first syllable (/ma/), 5 subjects promoted the second syllable (/də/), 4 subjects promoted the last syllable (/lɛn/) and 2 subjects alternated between the second and third syllables. But, for each subject, the association between pitch accent and syllable did not vary between recordings.

Overall, the subjects' performances as measured by the audio recordings indicated that there was no intra-speaker variability neither in rhythm nor in intonation between recordings.

FMRI Results

Table 2 represents the peaks of activations and their corresponding stereotaxic Talairach coordinates, provided by the fixed effect analysis. Figures 2-4 represent the functional activations obtained for the main effects with the fixed effect analysis.

Insert Table 2 about here

The pattern of activations common to the three deixis conditions (each compared to the baseline) included parts of the LIFG (BA 45, 47), the left insula and the premotor cortex (BA 6) bilaterally. Prosodic deixis additionally activated the left anterior cingulate gyrus (BA 24, 32), the left supramarginal gyrus (LSMG, BA 40) and the left postero-superior temporal gyrus (Wernicke's area, BA 22).

The (prosodic deixis - syntactic deixis) contrast yielded significant activation in the left posterosuperior temporal gyrus and the LSMG.

Insert Figures 2-4 about here

The results of the random effect analysis, using the same statistical significance threshold ($p = 0.001$ corrected), for the same contrasts did not provide significant activations. With a less stringent significance threshold however ($p=0.05$ non corrected), the contrasts provide a similar pattern of activations as the one obtained with the fixed effect analysis.

DISCUSSION

Baseline condition

No "resting" condition was included in the paradigm. The adequacy of using the "resting state" in fMRI studies on cognition is a matter of debate. While it is generally accepted that the "resting" state involves activity within many brain regions, some authors (Mazoyer, Zago, Mellet, Bricogne, Etard, Houdé *et al.*, 2001; Raichle, MacLeod, Snyder, Powers, Gusnard & Shulman, 2001) consider that it nevertheless may constitute an adequate control condition in functional imaging studies on cognition. The hypotheses underlying this opinion are that the "resting" state corresponds to a well-defined baseline cognitive state presenting specific electroencephalographic and metabolic signatures and involving a specific network of cortical areas, and that the

brain activity specific to the “resting” state is interrupted and temporarily suspended during the performance of cognitive tasks.

The concept of a “stationary level of activity” maintained during the “resting” state is however debated, as by Laufs, Krakow, Sterzer, Eger, Beyerle, Salek-Haddadi *et al.* (2003), asserting that “*instead of globally stabilizing at a homogeneous baseline level, brain activity fluctuates within and between different modes that imply different segregated functional networks and have distinct EEG signatures*”.

The mental processes taking place during the “resting” state thus seem by no means well-defined. Furthermore, since certain studies have indicated the involvement of semantic processes during the “resting” state (e.g. Binder, Frost, Hammeke, Bellgowan, Rao & Cox, 1999) and since the monitoring of thematic roles may be related to semantic processing, we have preferred not to use the “resting” state as baseline condition. The baseline condition used in this study therefore involved covert production – without prosodic deixis – of the sentence used in the prosodic deixis condition. With this particular choice, the activations obtained in the three (deixis – baseline) contrasts are specific of deixis production rather than of deixis in combination with covert speech production.

Thematic role monitoring

The production of the three deictic sentences (conditions 2, 3 and 4) should involve thematic role monitoring, or the tracking of “who-did-what-to-whom”, since they operate a contrastive pointing at one specific agent of the action (Madeleine). The prosodic deixis condition (2) involves no phrasal constituent movement [insert footnote 3], whereas both the syntactic (3) and the combined (4) deixis conditions involve presentational sentences which can be analysed (in a Chomskyan framework) as cleft-sentences requiring a transformational movement [insert footnote 4]. Other researchers however analyze sentences such as “It’s the N that VERB PHRASE” (like “C’est Madeleine qui m’amena”) as presentational relative constructions which do not involve an embedding of one clause into another and therefore do not require movement (see Tomasello, 2003, following Lambrecht, 1988 and Fox & Thompson, 1990). According to this last analysis, none of the deixis conditions we studied involved more movement than the baseline condition.

In summary, whatever the framework of analysis, we claim that the prosodic deixis condition involves thematic role monitoring using on-line vocal pointing, whereas the syntactic deixis uses a formalized presentational construction.

Habituation effects

Only four sentences were used over and over again. The subjects had been extensively trained to task performance the day before the experiment, with the same four sentences. In doing so, we ensured correct task performance during the fMRI experiment (the pre- and post-scan audio recordings used the same four sentences and thus allowed assessment of the subjects’ performance during the scan). We also ensured that task performance was practically effortless for the subjects. Thus, habituation effects may have taken place. Subjects needed to be skilled, however. The study therefore addresses the production of deictic sentences in a practiced mode.

Audio results

The baseline and prosodic conditions involved the same three words (Madeleine m’amena, /ma.də.lən.ma.mə.na/) and the syntactic and combined conditions used the same five words (c’est Mad’leine qui m’am’na, /sɛ.ma.dlən.ki.ma.mna/). Although schwa deletion was imposed to maintain a same number of syllables (six) in the two sets, a slight duration difference is observed (Table 1). The mean durations for the three-word set are lower than those for the five-word set of sentences, by at most 205ms before and 184 ms after the scans. This difference is of the order of a typical syllable duration in French (about 200 ms) and is due to the presence of two longer syllables in the five-word set (/dlən/ and /mna/).

Left Inferior Frontal Gyrus

This fMRI study shows activation in the LIFG for all the deixis conditions compared with the baseline. The LIFG was therefore activated during verbal pointing at the agent of the action, through either prosody or syntax.

In an attempt to relate the deficits of Broca's aphasics to current linguistic theoretical frameworks, Grodzinsky has concluded that the LIFG must have a highly specific, double role (Grodzinsky, 2000). In speech comprehension, it would process the transformational movement in syntax (Chomsky, 1981). The receptive deficit of Broca's aphasics would be one of trace deletion, and hence of thematic role misinterpretation. In speech production, the LIFG would be involved in the construction of the upper parts in the hierarchical structure (or tree) of sentences. The productive deficit of Broca's aphasics would be due to syntactic-tree pruning, whereby the syntactic tree would remain intact up to the tense node and be pruned from this node and up [insert footnote 5]. This interpretation may explain why Broca's aphasics often make tense errors ("six months ago my mother pass away") while producing correct agreements ("the boy stands").

Our results are difficult to reconcile with Grodzinsky's claim that the LIFG is devoted to trace maintenance in thematic-role interpretation and full-fledged syntactic tree construction (i.e. with a preserved tense node) in speech production and that "processes underlying these highly structured syntactic abilities, and only these, are located in the anterior language areas" (Grodzinsky, 2000). They show, in contrast, that the LIFG is not devoted to trace maintenance and tense processing only. The prosodic deixis condition and the baseline condition used in our fMRI study had exactly the same syntactic constituents, in the same order, and exactly the same tense. They only differed in the presence or absence of a prosodic focus on the agent of the action. The prosodic deixis condition therefore did not involve more trace maintenance or more tense processing than the baseline condition. Yet, the (prosodic deixis – baseline) contrast revealed activation in the LIFG.

Certain studies have suggested that the LIFG is involved in subvocal rehearsal, one of the components of the phonological/articulatory loop of verbal working memory (Paulesu, Frith & Frackowiak, 1993; Poeppel, 1996; Schumacher, Lauber, Awh, Jonides, Smith & Koeppe, 1996; Dronkers, 2000). This led to the hypothesis, formulated by Stowe and colleagues (Stowe, Broere, Paans, Wijers, Mulder, Vaalburg *et al.*, 1998; Stowe, 2000; see also Kaan & Swaab, 2002), that the LIFG primarily supports temporary storage of verbal (including structural) information.

The involvement of the LIFG in the subvocal rehearsal component has however not always been observed (Lassen & Larsen, 1980). Also, subvocal rehearsal was purposely limited in Caplan *et al.*'s (2000) PET study, yet LIFG activation was observed. While subjects performed a plausibility judgment task on syntactically complex constructions, a task requiring intricate thematic-role tracking, they were concurrently engaged in a repetitive simple articulation task. The increase of LIFG activity with the complexity of syntactic structures observed in the latter study must therefore most likely be ascribed to the parsing of the thematic roles rather than to subvocal rehearsal of more complex sentences. Finally, our own results do not support the hypothesis that the LIFG is simply involved in storage. The prosodic deixis condition does not involve more storage than the baseline condition, at least in terms of the number of phonemes or words, yet the (prosodic deixis – baseline) contrast shows LIFG activation.

The activation of the LIFG observed during verbal pointing at the agent of an action is consistent with functional neuroimaging studies on complex syntactic processing (Just *et al.*, 1996; Caplan *et al.*, 2000). As mentioned before, these studies have shown the involvement of the LIFG in plausibility judgments about syntactically complex constructions (with cleft-object sentences, or sentences with center-embedded clauses), the latter requiring intricate tracking of thematic roles.

Further evidence on the role of the LIFG in the tracking of thematic roles comes from lesion studies. The study by Caplan & Hanna (1998) was devoted to the production of utterances that convey thematic roles in normal and agrammatic subjects. On the basis of previous results in the literature (Saffran, Schwartz & Marin, 1980, Ostrin & Schwartz, 1986) and of a new study of sixty aphasic patients, Caplan & Hanna have suggested that aphasic patients have difficulty producing thematic roles. Other studies of the productions of agrammatic patients provide converging results (e.g. Whitworth, 1995; Collina, Marangolo & Tabossi, 2001; Webster, Franklin & Howard, 2001). Studies of the comprehension deficits in aphasic patients also suggest that the LIFG is involved in thematic role processing (e.g. Caplan, Baker & Dehaut, 1985; Schwartz, Linebarger, Saffran & Pate, 1987; Friederici & Gorell, 1998; Rigalleau, Baudiffier & Caplan, 2004).

Our findings are also in line with studies on the observation and mental imagery of action which show LIFG activation in action tracking (Grafton *et al.*, 1996; Rizzolatti *et al.*, 1997; Iacoboni *et al.*, 1999; Binkofski *et al.*,

2000). According to Rizzolatti and colleagues, the role of the LIFG in speech would have evolved from a “basic mechanism originally not related to communication: the capacity to recognize actions” (Rizzolatti & Arbib, 1998).

Several functional imaging studies as well as lesion studies therefore support our claim that the role of the LIFG is that of an action-structure parser, which, in morphosyntactic encoding and decoding, handles the parsing of the predicate and its arguments, or, in other terms, the attentional monitoring of “who-does-what-to-whom”.

Functional dissociation within the LIFG

Peak activation was observed in the anterior portion of the LIFG, i.e. in BA 45 and/or BA 47, not in BA44 – the portion traditionally associated with syntactic processing. Although most brain areas (Brett, Johnsrude & Owen, 2002), and BA 44 and 45 more specifically (Amunts, Schleicher, Bürgel, Mohlberg, Uylings & Zilles, 1999), cannot be precisely delineated from functional imaging data, this finding deserves comment. Several functional separations have been suggested to exist within the LIFG. A number of studies have suggested an anterior-posterior functional dissociation within the LIFG. Posterior LIFG (*pars triangularis* and *opercularis*, BA 44/45) has been associated with phonological or syntactic processing while anterior ventral LIFG (*pars orbitalis*, BA 47) has been associated with semantic processing. Very recent studies suggest however that portions BA 44, 45 and 47 are all activated for both semantic and syntactic tasks (Ni, Constable, Mencl, Pugh, Fulbright, Shaywitz *et al.*, 2000). Furthermore action monitoring (and especially lip reading) seems also to recruit all three portions (Calvert & Campbell, 2003; Buccino, Lui, Canessa, Patteri, Lagravinese, Benuzzi *et al.*, 2004).

In her neurocognitive model of sentence comprehension, Friederici (2002) proposes that thematic role assignment, which, according to her, involves lexical-semantic and morpho-syntactic processes, recruits both the anterior (BA 45/47) and the posterior (BA 44/45) portions of the LIFG. The peak activation detected in the anterior part of the LIFG is therefore not contrary to other findings and could provide additional data to the debate on functional segregation. The impossibility to precisely localize the activations with respect to the cytoarchitectonic areas within the LIFG and the recent findings on the involvement of the 3 portions of the LIFG in semantic as well as syntactic processing lead us however to express a comment of caution not to overinterpret this observation.

Left Insula

The left insula was also found activated in all the (deixis – baseline) contrasts. The involvement of the left precentral gyrus of the insula in articulatory planning during speech has already been shown (Dronkers, 1996). Prosody has both acoustic (variations in the fundamental frequency) and articulatory correlates (Beckman, Edwards & Fletcher, 1992; Fougeron & Keating, 1997; Løevenbruck, 1999). The production of prosodic focus may require more precise laryngeal control as well as more accurate articulatory planning of the movements of the tongue and jaw, which could underly why the prosodic deixis condition yields significant activation of the left insula when compared with the baseline condition (same words to articulate, but a more stringent prosody). Similarly, the syntactic deixis condition compared with the baseline condition likely requires more accurate articulatory planning, given the larger number of consonant clusters involved (due to the schwa deletions imposed to keep the number of syllables constant).

Wernicke’s area, Left Supramarginal Gyrus, and Left Inferior Frontal Gyrus

The activation of the left supramarginal gyrus and of Wernicke’s area in the prosodic deixis condition but not in the other two deixis conditions suggests that, when deixis is already encoded by syntax, no additional recruitment of Wernicke’s area and of the posterior parietal lobule is necessary. The posterior parietal lobule is often considered an association area, part of a network for spatial awareness, that integrates distributed multimodal sensory signals (somatosensory, visual, auditory) to form an interactive representation of space (Andersen, 1997; Mesulam 1981, 1999). Among the extensive body of research on the role of the posterior/inferior parietal cortex, two sets of findings are of particular interest in the present study. The first deals with pointing and the second with verbal working memory or rather with the temporo-parieto-frontal network.

Role of the Inferior Parietal Lobule in manual pointing

The first set of observations deals with the role of the posterior parietal regions of both hemispheres in linguistic and non linguistic manual pointing tasks. It has been suggested that the spatial representations formed in the

posterior parietal and premotor frontal regions could provide “perceptual – premotor interfaces for the organization of movements (e.g. pointing, locomotion) directed towards targets in personal and extrapersonal space” (Vallar, 1997, our underlining). This hypothesis is supported by the observation that patients with lesions in the right inferior posterior parietal region (and more specifically in the supramarginal gyrus, Vallar & Perani, 1986) often show motor impairment, in addition to the typical perceptual spatial hemineglect. Furthermore, in a study on two right brain-damaged patients with left visuospatial hemineglect, Vallar and colleagues showed that the sensory stimulations (moving luminous dots) that modulate the severity of the left somatosensory deficits similarly modulate the left motor disorders (improve muscle strength) associated with this syndrome (Vallar, Guariglia, Nico & Pizzamiglio, 1997). Also, in agreement with this hypothesis, patients with left unilateral neglect have been shown to present deficits in pointing tasks (e.g. Edwards & Humphreys, 1999) while PET studies on normal subjects show activation within the left and/or right inferior parietal lobule (IPL) during pointing tasks (e.g. Lacquaniti, Perani, Guigon, Bettinardi, Carrozzo, Grassi *et al.*, 1997; Kertzman, Schwarz, Zeffiro & Hallett, 1997).

Role of the IPL in linguistic manual pointing

The role of the right and left posterior/inferior parietal regions in pointing tasks may further be related to data on brain-damaged deaf signers. Bellugi and colleagues presented a study of deaf signers of American Sign Language (ASL), two of which presented lateralized parietal lesions, one in the right hemisphere and the other in the left (Bellugi, Poizner & Klima, 1989). Space in ASL is handled in two ways. The first is topographic: in the description of the layout of objects in space, spatial relations among signs reproduce the actual spatial relations among the objects. The second way is deictic: space is used for referential indexing. Noun phrases, for instance, may be associated with loci in space. Reference to a previously mentioned noun is performed by pointing again to its specific locus. Interestingly, the right-lesioned signer had difficulty in the use of space for topography: room description was distorted spatially, with left side of signing space neglected. In the use of space for syntax, however, the entire signing space (including the left) was covered and consistent reference to spatial loci was preserved. By contrast, the left-lesioned signer produced room descriptions without spatial distortions but made errors in the deictic use of space.

Differential lateralization of the IPL in pointing

These results could suggest a differential lateralization of IPL activity, with the right IPL involved in non linguistic pointing tasks and the left IPL involved in linguistic manual pointing (in sign language). Conflicting data have been reported by some recent studies in this respect. As to non linguistic pointing, Lacquaniti *et al.* (1997) have reported left IPL activation for immediate pointing to a target and bilateral IPL activation for pointing to a memorized target. Their interpretation is that the decoding of a memorized location in a body-centered frame (to direct the pointing movement) is handled in the right IPL. A left lateralization of the IPL activation is also found by Astafiev, Shulman, Stanley, Snyder, Van Essen & Corbetta (2003) for non linguistic pointing as well as pointing preparation. In a recent study on (non verbal) action monitoring, Chaminade & Decety (2002) have suggested that the hemispheric asymmetry observed for the IPL may be related to agency. They demonstrated that the left IPL is more activated when subjects imitate the actions by others (subject in a follower role), while the right IPL is more activated when the self is imitated (subject in a leader role). The asymmetry is also found in linguistic processing of spatial relations. Recent data on IPL lateralization during the production or comprehension of topographical relations in sign language are discussed by Campbell & Woll (2003). Thus, while it appears that the left IPL is involved in linguistic manual pointing (sign language), its implication in some non linguistic pointing and in the monitoring of the other suggests the need for further clarification of its role in pointing.

Role of the Inferior Parietal Lobule in working memory? The Temporo-Parieto-Frontal network

The second set of findings concerns the role of the left inferior parietal cortex in verbal working memory. Some neuroimaging studies of verbal working memory have shown the implication of the left inferior parietal cortex in short-term storage of phonologically coded verbal material (see e.g. Paulesu *et al.*, 1993; Jonides, Schumacher, Smith, Koeppe, Awh, Reuter-Lorenz *et al.*, 1998). Hickok and Poeppel offer a hypothesis which can account for these results by analogy with the visual – motor interface system presented above. According to these authors, “inferior parietal cortex is not the site of storage of phonemic representations per se (...) but rather serves to interface sound-based representations of speech in auditory cortex with articulatory-based representations of

speech in frontal cortex" (Hickok & Poeppel, 2000). In analogy with the dorsal pathway hypothesized in audition and vision, the left inferior parietal cortex would therefore play a role within a temporo-parieto-frontal network functioning as an interface system between auditory and articulatory processes. Involvement of this temporo-parieto-frontal circuit has also been described in action imitation (Iacoboni, Koski, Brass, Bekkering, Woods, Dubeau *et al.*, 2001). Iacoboni and colleagues conjecture that the description of the actions to be imitated is handled by the superior temporal sulcus (STS) and sent to the posterior parietal cortex, where it is combined with additional somatosensory information. This completed description would then be sent to the inferior frontal cortex where the goal of the actions to be imitated would be coded. Reafferent copies of the imitated actions would be sent back to the STS for action monitoring.

In summary, all these studies suggest that the inferior parietal regions in both hemispheres function as sensory integrators to form representations necessary in the organization of motor actions, such as (linguistic or non-linguistic) pointing at targets. The left hemisphere would have a linguistic predominance. A left temporo-parieto-frontal network might be recruited in the organization of verbal motor actions from auditory representations.

Our results, i.e. the activations of the left inferior parietal lobule together with the LIFG and Wernicke's area during prosodic deixis, are in line with this hypothesis. Prosodic deixis, i.e. expressive orofacial (manual and facial for sign language) deixis may be considered in continuity with manual pointing. In analogy with visually-guided manual pointing, prosodic pointing may need integrated representations (auditory and articulatory) to be formed via the superior temporal and inferior parietal regions in order to organize articulation and phonation in an adequate prosodic pattern.

We therefore hypothesize that non-grammaticalized verbal pointing recruits the temporo-parieto-frontal network and that grammaticalized deixis (syntactic deixis with or without supplementary prosody) is handled solely by the left IFG. Further experiments based on gradual grammaticalization tasks are needed to clarify matters.

CONCLUSION

The earliest report of a neurological disorder of prosody was published by Monrad-Krohn (1947). Monrad-Krohn coined the term "aprosodia" to describe the foreign accent syndrome in a patient who had suffered a lesion in Broca's area and suggested that intact left frontal lobe was essential for prosody production.

Prosody has traditionally been considered to be processed by the right hemisphere, however, a view reflecting the traditional conception of prosody as a well adapted subordinate to syntax and semantics, which recruit the left hemisphere (Dronkers, Pinker & Damasio, 2000). Several lesion studies (Ross, 1981; Weintraub, Mesulam & Kramer, 1981; Klouda, Robin, Graff-Radford & Cooper, 1988; Twist, Squires, Spielholz & Silverglide, 1991; Brådvik, Dravins, Holtås, Rosén, Ryding & Ingvar, 1991) provide data supporting this view. Recent neuroimaging studies provide converging results. When aspects of prosody associated with melody processing are studied, activation in the right hemisphere is found indeed (see e.g. Zatorre, Evans, Meyer & Gjedde, 1992; Tzourio, El Massioui, Crivello, Joliot, Renault & Mazoyer, 1997; Meyer, Alter, Friederici, Lohmann & von Cramon, 2002).

Recent linguistic studies have shown that prosody is itself a "complex grammatical (phonological) structure that must be parsed in its own right" (Beckman, 1996), however. Prosody, therefore, should recruit the left hemisphere, similarly as syntax and semantics. Interestingly, a recent review of the literature (Baum & Pell, 1999) shows that the processing (in production and perception) of prosody in general (affective and linguistic) is not strictly localizable to the right hemisphere. More specifically, this review quotes lesion studies on the production and on the perception of emphatic stress (which is related to prosodic focus) showing that left-damaged patients (most often Broca's aphasics) can be as strongly impaired as right-damaged patients or even more (see Bryan, 1989; Baum *et al.*, 1982; Ouellette and Baum, 1994; Pell 1998; see also Avrutin, Lubarsky & Greene, 1999; Geigenberger & Ziegler, 2001). Recent neuroimaging studies also provide converging results. Astésano and colleagues (Astésano, Besson & Alter, 2004) present electrophysiological evidence that attention to prosody (detection of prosodic mismatch) primarily recruits the left hemisphere. Incidentally, they also report that the electrophysiological response P800 has a larger amplitude at temporo-parietal electrodes, in accordance with our own results showing activations of the LSMG and Wernicke's area in the prosodic task. A number of fMRI studies on the receptive processing of several aspects of prosody also show activation in the left IFG (Dapretto,

Hariri, Bialik & Bookheimer, 1999). In addition, Mayer and colleagues's fMRI study on the production of prosodic features at the syllable and phrase levels has also revealed left IFG activation (Mayer, Wildgruber, Riecker, Dogil, Ackermann & Grodd, 2002). We consider that these results are consistent with our two conjectures: the LIFG being a parser of action structure, particularly well adapted to agent deixis and the left temporo-parieto-frontal network functioning as an interface between auditory and articulation/phonation processes, required in prosody-driven deixis.

ACKNOWLEDGEMENTS

We are very grateful to Véronique Aubergé and Pascal Perrier for valuable comments and advice and to Chantal Delon-Martin for her assistance with the statistical analysis of the fMRI data. We also thank Cécile Bricchet, Christophe Savariaux, Chafika Rabehi & Albert Rilliard for discussion. This research was supported by the grant "Aide à Projets Nouveaux" of the SHS department of the CNRS, project # 29031.

REFERENCES

- Amunts K, Schleicher A., Bürgel U., Mohlberg H., Uylings H. B. M., Zilles K. (1999). Broca's region revisited: cytoarchitecture and intersubject variability. *J. Comp. Neurol.*, 412, 319-341.
- Andersen R. A. (1997). Multimodal integration for the representation of space in the posterior parietal cortex. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.*, 352(1360), 1421-1428.
- Anderson R. S. (1972). How to get even. *Language*, 48 (4), 893-906.
- Astafiev S. V., Shulman G. L., Stanley C. M., Snyder A. Z., Van Essen D. & Corbetta M. (2003). Functional organization of human intraparietal and frontal cortex for attending, looking and pointing. *J. Neuroscience*, 23 (11), 4689-4699.
- Astésano C., Besson M. & Alter K. (2004). Brain potentials during semantic and prosodic processing in French. *Cognitive Brain Research*, 18, 172-184.
- Avrutin S., Lubarsky S. & Greene J. (1999). Comprehension of contrastive stress by Broca's aphasics. *Brain and Language*, 70 (2), 163-186.
- Baum S. R., Daniloff J. K., Daniloff R. & Lewis J. (1982). Sentence comprehension by Broca's aphasics: effects of some suprasegmental variables. *Brain and Language*, 17 (2), 261-271.
- Baum S. R. & Pell M. D. (1999). The neural bases of prosody: insights from lesion studies and neuroimaging. *Aphasiology*, 13 (8), 581-608.
- Beckman M. E. (1996). The parsing of prosody. *Language and Cognitive Processes*, 11 (1/2), 17-67.
- Beckman M. E., Edwards J. & Fletcher J. (1992). Prosodic structure and tempo in a sonority model of articulatory dynamics. *Papers in Laboratory Phonology II; Gesture, segment, prosody*. G. J. Docherty & D. R. Ladd (eds.), Cambridge University Press. 68-86.
- Bellugi U., Poizner H. & Klima E. S. (1989). Language, modality and the brain. *Trends in Neurosciences*, 12 (10), 380-388.
- Berthoud A.-C. (1990). Deixis, thématization et détermination. *La deixis. Colloque en Sorbonne*. M.-A. Morel & L. Danon-Boileau (dir.), PUF, 527-542.
- Binder J. R., Frost J. A., Hammeke T. A., Bellgowan P. S., Rao S. M. & Cox R. W. (1999). Conceptual processing during conscious resting state. A functional MRI study. *J. Cognitive Neuroscience*, 11 (1), 80-95.
- Binkofski F., Amunts K., Stephan K. M., Posse S., Schormann T., Freund H.-J., Zilles K. & Seitz R. (2000). Broca's region subserves imagery of motion: a combined cytoarchitectonic and fMRI study. *Hum. Brain Mapp.*, 11, 273-285.
- Boersma P. (2001). PRAAT, a system for doing phonetics by computer. *Glott International*, 5(9/10), 341-345.
- Brådvik B., Dravins C., Holtås S., Rosén I., Ryding E. & Ingvar D. (1991). Disturbances of speech prosody following right hemisphere infarcts. *Acta Neurologica Scandinavia*, 84, 114-126.
- Brett M., Johnsrude I. S. & Owen A. M. (2002). The problem of functional localization in the human brain. *Nature Reviews Neuroscience*, 3 (3), 243-249.
- Broca P. (1861). Remarques sur le siège de la faculté du langage articulé, suivies d'une observation d'aphémie (perte de la parole). *Bulletin de la Société d'Anatomie de Paris*, 6, 330-357.
- Bródy M. (1990). Some remarks on the focus field in Hungarian. *UCL Working Papers in Linguistics*, 2, 201-225.
- Bryan K. (1989). Language prosody and the right hemisphere. *Aphasiology*, 3, 285-299.

- Buccino G., Binkofski F., Fink G. R., Fadiga L., Fogassi L., Gallese V., Seitz R. J., Zilles K., Rizzolatti G. & Freund H.-J. (2001). Action observation activates premotor and parietal areas in a somatotopic manner: an fMRI study. *European Journal of Neuroscience*, 13, 400-404.
- Buccino G., Lui F., Canessa N., Patteri I., Lagravinese G., Benuzzi F., Porro C. A., Rizzolatti G. (2004). Neural circuits involved in the recognition of actions performed by nonconspecifics: an fMRI study. *J. Cogn. Neurosci.*, 16(1), 114-216.
- Calvert G. A. & Campbell R. (2003). Reading speech from still and moving faces: the neural substrates of visible speech. *J. Cognitive Neuroscience*, 15 (1), 57-70.
- Campbell R. & Woll B. (2003). Space is special in sign. *Trends in Cognitive Sciences*, 17 (1), 5-7.
- Caplan D., Baker C. & Dehaut F. (1985). Syntactic determinants of sentence comprehension in aphasia. *Cognition*, 21, 117-175.
- Caplan D. & Hanna J. E. (1998). Sentence production by aphasic patients in a constrained task. *Brain and Language*, 63 (2), 184-218.
- Caplan D., Alpert N., Waters G. & Olivieri A. (2000). Activation of Broca's area by syntactic processing under conditions of concurrent articulation. *Hum. Brain Mapp.*, 9, 65-71.
- Caramazza A. & Zurif E. B. (1976). Dissociation of algorithmic and heuristic processes in language comprehension: evidence from aphasia. *Brain and Language*, 3 (4), 572-582.
- Chaminade T. & Decety J. (2002). Leader of follower? Involvement of the inferior parietal lobule in agency. *Neuroreport*, 13 (15), 1975-1978.
- Chomsky N. (1976). Conditions on Rules of Grammar. *Linguistic Analysis*, 2, 303-351.
- Chomsky N. (1981). *Lectures on government and binding*. Foris.
- Collina S., Marangolo P. & Tabossi P. (2001). The role of argument structure in the production of nouns and verbs. *Neuropsychologia*, 39 (11), 1125-1137.
- Costa J. (1998). *Word Order Variation: A Constraint-Based Approach*. The Hague: Holland Academic Graphics.
- Dapretto M., Hariri A., Bialik M. & Bookheimer S. (1999). Cortical correlates of affective vs. linguistic prosody: an fMRI study. *NeuroImage*, 9 (6), S1054.
- Démonet J.-F., Price C., Wise R. & Frackowiak R. S. J. (1994). A PET study of cognitive strategies in normal subjects during language tasks: Influence of phonetic ambiguity and sequence processing on phoneme monitoring. *Brain*, 117, 671-682.
- Dominey P. F. & Lelekov T. (2000). Nonlinguistic transformation processing in agrammatic aphasia. *Commentary/Grodzinsky: Neurology of syntax. Behavioral and Brain Sciences*, 23, 30.
- Dronkers N. F. (1996). A new brain region for coordinating speech articulation. *Nature*, 384, 159-161.
- Dronkers N. F. (2000). The gratuitous relationship between Broca's aphasia and Broca's area. *Commentary/Grodzinsky: Neurology of syntax. Behavioral and Brain Sciences*, 23 (1), 30-31.
- Dronkers N. F., Pinker S. & Damasio A. (2000). Language and the Aphasias. *Principles of neural science*, E. R. Kandel, J. H. Schwartz & T. M. Jessell (eds.), McGraw-Hill. 1169-1187.
- Edwards M. G. & Humphreys G. W. (1999). Pointing and grasping in unilateral visual neglect: effect of on-line visual feedback in grasping. *Neuropsychologia*, 37 (8), 959-973.
- Fougeron C. & Keating P. A. (1997). Articulatory strengthening at edges of prosodic domains. *J. Acoust. Soc. Am.* 101 (6), 3728-3740.
- Fox B. & Thompson S. (1990). A discourse explanation of the "grammar" of relative clauses in English conversation. *Language*, 66, 297-316.
- Friederici A. D. (1990). On the properties of cognitive modules. *Psychol. Research*, 52, 175-180.
- Friederici A. D. (2002). Towards a neural basis of auditory sentence processing. *Trends in Cognitive Sciences*, 6 (2), 78-84.
- Friederici A. D. & Gorrell P. (1998). Structural prominence and agrammatic theta-role assignment: a reconsideration of linear strategies. *Brain and language*, 65 (2), 253-275.
- Geigenberger A. & Ziegler W. (2001). Receptive prosodic processing in aphasia. *Aphasiology*, 15 (12), 1169-1187.
- Goschke T., Friederici A. D., Kotz S. A. & van Kampen A. (2001). Procedural Learning in Broca's Aphasia: Dissociation Between the Implicit Acquisition of Spatio-Motor and Phoneme Sequences. *J. Cogn. Neurosci.*, 13, 370-388.

- Grafton S. T., Arbib M. A., Fadiga L. & Rizzolatti G. (1996). Localization of grasp representations in humans by positron emission tomography. 2. Observation compared with imagination. *Exp. Brain Res.*, 112 (1), 103-111.
- Grèzes J. & Decety J. (2001). Functional anatomy of execution, mental simulation, observation, and verb generation of actions: A meta-analysis. *Hum. Brain Mapp.*, 12, 1-19.
- Grèzes J., Costes N. & Decety J. (1998). Top down effect of strategy on the perception of human biological motion: a PET investigation. *Cogn. Neuropsychol.*, 15, 553-582.
- Grodzinsky Y. (2000). The neurology of syntax: Language use without Broca's area. *Behavioral and Brain Sciences*, 23 (1), 1-21.
- Grodzinsky Y., Piñango M. M., Zurif E. & Drai D. (1999). The critical role of group studies in neuropsychology: comprehension regularities in Broca's aphasia. *Brain and language*, 67 (2), 134-147.
- Hickok G. (2000). The left frontal convolution plays no special role in syntactic comprehension. Commentary/Grodzinsky: Neurology of syntax. *Behavioral and Brain Sciences*, 23, 35-36.
- Hickok G. & Poeppel D. (2000). Towards a functional neuroanatomy of speech perception. *Trends in Cognitive Sciences*, 4 (4), 131-138.
- Horvath J. (1986). *Focus in the theory of grammar and the structure of Hungarian*. Dordrecht: Foris.
- Iacoboni M., Woods R. P., Brass M., Bekkering H., Mazziotta J. C. & Rizzolatti G. (1999). Cortical mechanisms of human imitation. *Science*, 286, 2526-2528.
- Iacoboni M., Koski L. M., Brass M., Bekkering H., Woods R.P, Dubeau M.-C., Mazziotta J. C. & Rizzolatti G. (2001). Reafferent copies of imitated actions in the right superior temporal cortex. *Proc. Natl. Acad. Sci. USA*, 98 (24), 13995-13999.
- Jackendoff R. (2002). *Foundations of Language - Brain, Meaning, Grammar, Evolution*. Oxford University Press.
- Jonides J., Schumacher E. H., Smith E. E., Koeppe R. A., Awh E., Reuter-Lorenz P. A., Marshuetz C., Willis C. R. (1998). The role of parietal cortex in verbal working memory. *J. Neurosci.*, 18 (13), 5026-5034.
- Jun S.-A. & Fougeron C. (2000). A phonological model of French intonation. *Intonation: Analysis, modeling and technology*. A. Botinis (ed.) Dordrecht : Kluwer Academic Publishers. 209-242.
- Just M. A., Carpenter P. A., Keller T. A., Eddy W. F. & Thulborn K. R. (1996). Brain activation modulated by sentence comprehension. *Science*, 274, 114-116.
- Kaan E. & Swaab T. Y. (2002). The brain circuitry of syntactic comprehension. *Trends in Cognitive Sciences*, 6 (8), 350-356.
- Kertzman C., Schwarz U., Zeffiro T. A. & Hallett M. (1997). The role of posterior parietal cortex in visually guided reaching movements in humans. *Exp. Brain Res.*, 114 (1), 170-183.
- Klouda G. V., Robin D. A., Graff-Radford N. R. & Cooper W. E. (1988). The role of callosal connections in speech prosody. *Brain Lang.*, 35 (1), 154-171.
- Lacquaniti F., Perani D., Guigon E., Bettinardi V., Carrozzo M., Grassi F., Rossetti Y. & Fazio F. (1997). Visuomotor transformations for reaching to memorized targets: a PET study. *Neuroimage*, 5 (2), 129-146.
- Lambrecht K. (1988). Presentational cleft construction in spoken French. *Clause combining in grammar and discourse*. J. Haiman & S. A. Thompson (eds.), Amsterdam: Benjamins.
- Lassen N. A. & Larsen B. (1980). Cortical activity in the left and right hemispheres during language-related brain functions. *Phonetica*, 37, 27-37.
- Laufs H., Krakow K., Sterzer P., Eger E., Beyerle A., Salek-Haddadi A. & Kleinschmidt A. (2003). Electroencephalographic signatures of attentional and cognitive default modes in spontaneous brain activity fluctuations at rest. *Proc. Natl. Acad. Sci. USA*, 100 (19), 11053-11058.
- Levinson S. (1983). *Pragmatics*, Cambridge University Press, Cambridge, England.
- Linebarger M. C., Schwartz M. F. & Saffran E. M. (1983). Sensitivity to grammatical structure in so-called agrammatic aphasics. *Cognition*, 13 (3), 361-392.
- Lœvenbruck H. (1999). An investigation of articulatory correlates of the accentual phrase in French. *Proceedings of the XIVth International Congress of Phonetics Sciences*, San Francisco, CA, August 1999, 1, 667-670.
- Mayer J., Wildgruber D., Riecker A., Dogil G., Ackermann H. & Grodd W. (2002). Prosody production and comprehension: converging evidence from fMRI studies. *Proceedings of Speech Prosody 2002*, Aix-en-Provence, France, 11-13 April 2002, 487-490.

- Mazoyer B., Zago L., Mellet E., Bricogne S., Etard O., Houdé O., Crivello F., Joliot M., Petit L. & Tzourio-Mazoyer N. (2001). Cortical networks for working memory and executive functions sustain the conscious resting state in man. *Brain Res. Bull.*, 54 (3), 287-298.
- Mesulam M. M. (1981). A cortical network for directed attention and unilateral neglect. *Ann. Neurol.*, 10 (4), 309-325.
- Mesulam M. M. (1999). Spatial attention and neglect: parietal, frontal and cingulate contributions to the mental representation and attentional targeting of salient extrapersonal events. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.*, 354 (1387), 1325-1346.
- Meyer M., Alter K., Friederici A. D., Lohmann G. & von Cramon D. Y. (2002). fMRI reveals brain regions mediating slow prosodic modulations in spoken sentences. *Hum. Brain Mapp.*, 17, 73-88.
- Monrad-Krohn G. H. (1947). Dysprosody or altered "melody of language." *Brain*, 70, 405-415.
- Ni W., Constable R. T., Mencl W. E., Pugh K. R., Fulbright R. K., Shaywitz S. E., Shaywitz B. A., Gore J. C. & Shankweiler D. (2000). An event-related neuroimaging study distinguishing form and content in sentence processing. *J. Cognitive Neuroscience*, 12 (1), 120-133.
- Oldfield R. C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia*, 9 (1), 97-113.
- Ostrin R. K. & Schwartz M. F. (1986). Reconstructing from a degraded trace: a study of sentence repetition in agrammatism. *Brain and Language*, 28 (2), 328-345.
- Ouellette G. & Baum S. (1994). Acoustic analysis of prosodic cues in left and right-hemisphere-damaged patients. *Aphasiology*, 8, 257-283.
- Paulesu E., Frith C. D. & Frackowiak R. S. J. (1993). The neural correlates of the verbal component of working memory. *Nature*, 362, 342-344.
- Pell M. D. (1998). Recognition of prosody following unilateral brain lesion: influence of functional and structural attributes of prosodic contours. *Neuropsychologia*, 36(8), 701-715.
- Penhune V. B., Zatorre R. J. & Evans A. C. (1998). Cerebellar contributions to motor timing: A PET study of auditory and visual rhythm reproduction. *J. Cogn. Neurosci.*, 10, 752-765.
- Poeppl D. (1996). A critical review of PET studies of phonological processing. *Brain and Language*, 55(3), 317-351.
- Pollock J.-Y. (1989). Verb movement, universal grammar and the structure of IP. *Linguistic Inquiry*, 20, 365-424.
- Raichle M. E., MacLeod A. M., Snyder A. Z., Powers W. J., Gusnard D. A. & Shulman G. L. (2001). A default mode of brain function. *Proc. Natl. Acad. Sci.*, 98 (2), 676-682.
- Rigalleau F., Baudiffier V. & Caplan D. (2004). Comprehension of sentences with stylistic inversion by French aphasic patients. *Brain and Language*, 89 (1), 142-156.
- Rizzi L. (1997). The fine structure of the left periphery. *Elements of Grammar: Handbook of Generative Syntax*. L. Haegeman (ed.), Kluwer Academic Publishers, Dordrecht.
- Rizzolatti G., Fogassi L., Gallese V. (1997). Parietal cortex: from sight to action. *Curr. Opin. Neurobiol.*, 7 (4), 562-567.
- Rizzolatti G. & Arbib M. (1998). Language within our grasp. *Trends in Neurosciences*, 21, 188-194.
- Rooth M. (1985). Association with Focus. Ph.D. diss., University of Massachusetts at Amherst: GLSA.
- Ross E. D. (1981). The aprosodias: functional-anatomic organization of the affective components of language in the right hemisphere. *Arch. Neurol.* 38, 561-569.
- Saffran E. M., Schwartz M. F., Marin O. S. (1980). The word order problem in agrammatism. II. Production. *Brain and Language*, 10 (2), 263-280.
- Schumacher E. H., Lauber E., Awh E., Jonides J., Smith E. E. & Koeppel R. A. (1996). PET evidence for an amodal verbal working memory system. *NeuroImage*, 3 (2), 79-88.
- Schwartz M. F., Linebarger M. C., Saffran E. M. & Pate D. S. (1987). Syntactic transparency and sentence interpretation in aphasia. *Language and Cognitive Processes*, 2, 85-113.
- Stowe L. A. (2000). Sentence comprehension and the left inferior frontal gyrus: Storage, not computation. *Commentary/Grodzinsky: Neurology of syntax. Behavioral and Brain Sciences*, 23 (1), 51.
- Stowe L. A., Broere C. A., Paans A. M., Wijers A. A., Mulder G., Vaalburg W. & Zwarts F. (1998). Localizing components of a complex task: sentence processing and working memory. *NeuroReport*, 9 (13), 2995-2999.

- Tomasello M. (2003). *Constructing a language. A usage-based theory of language acquisition*. Harvard University Press, Cambridge, MA.
- Twist D., Squires N., Spielholz N. & Silverglide R. (1991). Event-related potentials in disorders of prosodic and semantic linguistic processing. *Neuropsychiatry, Neurosurgery, and Behavioral Neurology*, 4, 281-304.
- Tzourio N., El Massioui F., Crivello F., Joliot M., Renault B. & Mazoyer B. (1997). Functional anatomy of human auditory attention studied with PET. *NeuroImage*, 5, 63-77.
- Vallar G. (1997). Spatial frames of reference and somatosensory processing: a neuropsychological perspective. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.*, 352(1360), 1401-1409.
- Vallar G. & Perani D. (1986). The anatomy of unilateral neglect after right-hemisphere stroke lesions. A clinical/CT-scan correlation study in man. *Neuropsychologia*, 24 (5), 609-622.
- Vallar G., Guariglia C., Nico D. & Pizzamiglio L. (1997). Motor deficits and optokinetic stimulation in patients with left hemineglect. *Neurology*, 49 (5), 1364-1370.
- Webster J., Franklin S. & Howard D. (2001). An investigation of the interaction between thematic and phrasal structure in nonfluent agrammatic subjects. *Brain and Language*, 78 (2), 197-211.
- Weintraub S., Mesulam M. M. & Kramer L. (1981). Disturbances in prosody. A right-hemisphere contribution to language. *Arch. Neurol.*, 38(12), 742-744.
- Whitworth A. (1995). Characterising thematic role assignment in aphasic sentence production: procedures for elicited and spontaneous output. *Eur. J. Disord. Commun.*, 30 (3), 384-399.
- Zatorre R. J., Evans A. C., Meyer E. & Gjedde A. (1992). Lateralization of phonetic and pitch discrimination in speech processing. *Science*, 256, 846-849.
- Zurif E. B. (1980). Language mechanisms: a neuropsychological perspective. *American Scientist*, 68(3), 305-311.

FIGURE CAPTIONS

Figure 1

Acoustic analysis of one repetition of the prosodic deixis sentence by speaker DB. Top panel: acoustic waveform (top panel). Middle panel: spectrogram with superimposed F0 trace. Bottom panels: syllable duration and prosodic tiers. The high F0 peak on the first syllable /ma/ is labelled as LHF (Low High sequence with focus). The post-focal F0 trace falls to reach a flat floor (the fall is labelled as two low Accentual Phrase boundary tones, L%). The same F0 and duration pattern is observed across the 8 repetitions by this speaker.

Figure 2

(Prosodic deixis – baseline) contrast. Projection of the activation foci onto the right and left lateral surfaces of a standard brain (MNI template).

Figure 3

(Syntactic deixis – baseline) contrast. Projection of the activation foci onto the right and left lateral surfaces of a standard brain (MNI template).

Figure 4

(Combined deixis – baseline) contrast. Projection of the activation foci onto the right and left lateral surfaces of a standard brain (MNI template).

Table 1

Mean sentence duration and standard deviation (in ms) for the 4 conditions, before and after the scans.

	Baseline		Prosodic deixis		Syntactic deixis		Combined deixis	
	Mean	Standard deviation	Mean	Standard deviation	Mean	Standard deviation	Mean	Standard deviation
Pre-scan	1049	116	1052	114	1203	91	1254	155
Post-scan	1011	108	1046	79	1189	83	1195	80

Table 2

Talairach coordinates and Z-scores of activated regions in the deictic tasks.

Region	Talairach coordinates (mm)			Z-scores
	x	y	z	
Prosodic deixis - baseline				
Left insula	-32	4	9	> 8
Left insula	-36	-11	10	6.5
Left anterior cingulate (BA 24)	-12	5	18	5.1
LIFG (BA 47)	-32	20	-5	> 8
Wernicke's area (BA 22)	-48	-16	10	6.5
Left medial frontal gyrus (BA 6)	-28	2	46	7.3
Left anterior cingulate gyrus (BA 32)	-20	10	41	7.2
Left SMG (BA 40)	-44	-37	34	6.6
Right SMA (BA 6)	4	-9	56	6.4
Right SMA (BA 6)	16	-9	56	6.1
Syntactic deixis – baseline				
Left insula	-32	8	9	7.0
LIFG (BA 45)	-28	20	8	5.2
LIFG (BA 47)	-40	16	-1	4.9
Left superior frontal gyrus (BA 8)	-20	14	50	6.4
Left medial frontal gyrus (BA 6)	-24	3	51	6.0
Left medial frontal gyrus (BA 6)	-24	-9	56	5.2
Combined deixis – baseline				
Left insula	-36	4	9	6.1
LIFG (BA 45)	-28	16	8	6.0
Left SMA (BA 6)	-20	10	50	6.1
Left medial frontal gyrus (BA 6)	-20	-16	61	5.6
Left medial frontal gyrus (BA 8)	-20	14	41	5.5
Prosodic deixis - syntactic deixis				
Left posterosuperior temporal gyrus (BA 22)	-40	-35	6	7.0
Left SMG (BA 40)	-40	-33	29	6.4
Left SMG (BA 40)	-40	-33	48	6.2

Note. Corresponding Brodmann's areas are given in parentheses; LIFG, Left Inferior Frontal Gyrus; SMA, Supplementary Motor Area; SMG, Supramarginal Gyrus.

Figure 1

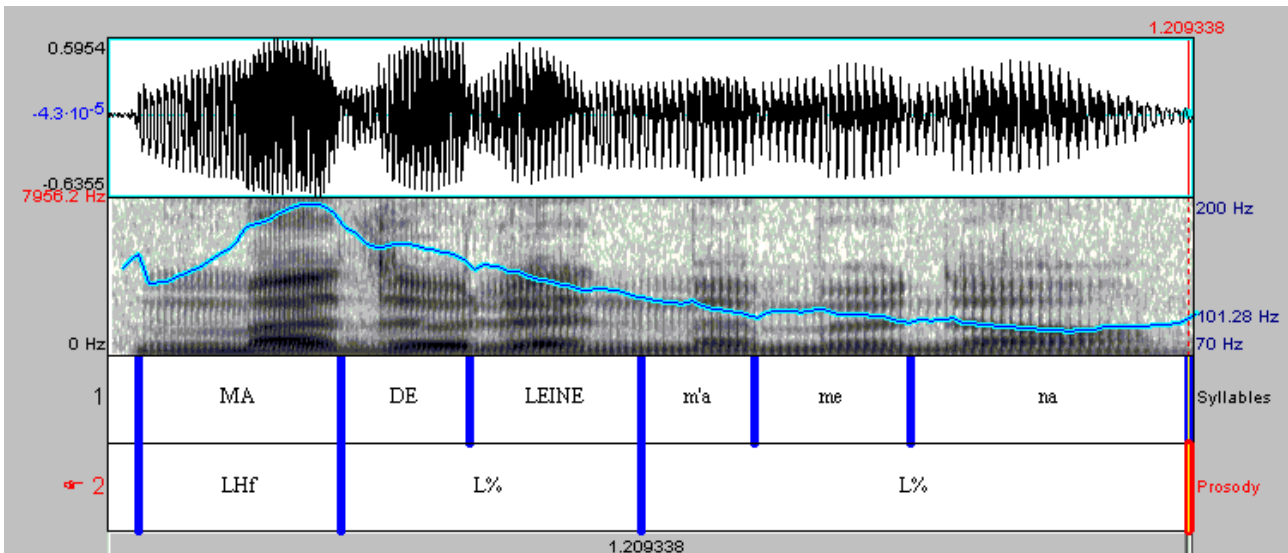


Figure 2

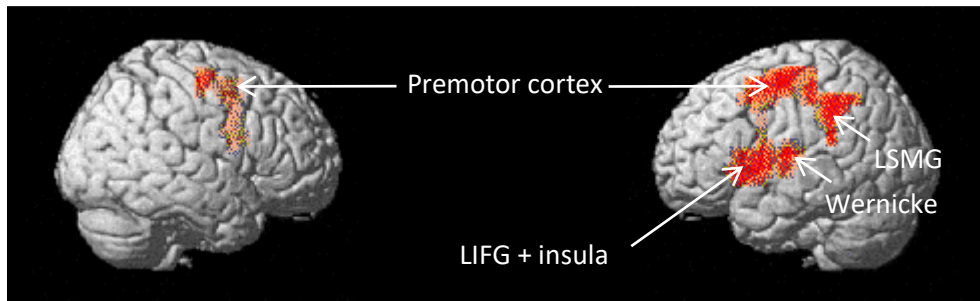


Figure 3

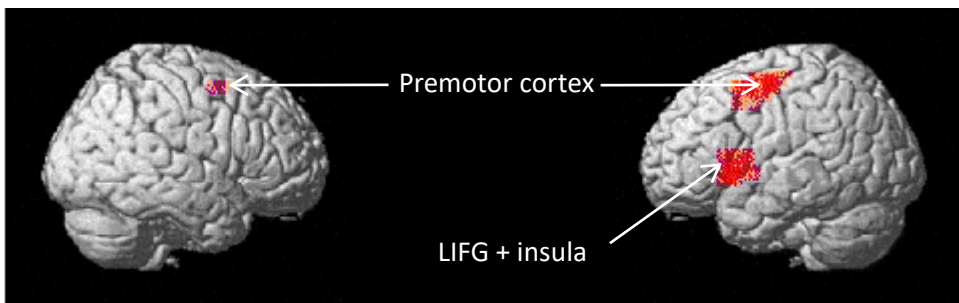
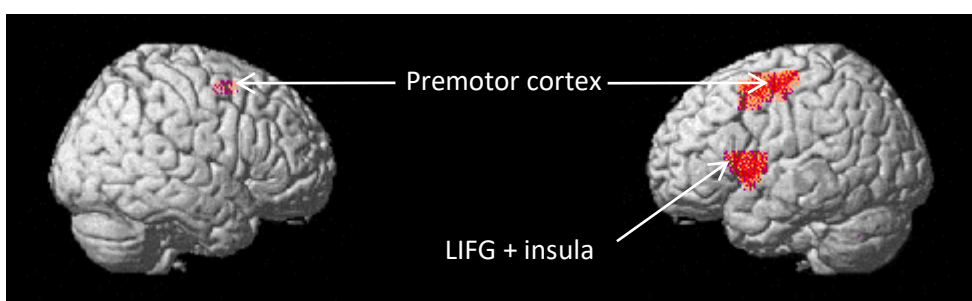


Figure 4



FOOTNOTES

Footnote 1

In French as well as in English, agent deixis can be conveyed by syntax or by prosody (Berthoud, 1990). For instance, when conveying “Madeleine brought me around”, one can specifically point at the agent “Madeleine” by using a deictic presentation form, either with the syntactic extraction “it’s Madeleine who brought me around”, or by producing an intonational contour bearing a focus on “Madeleine”. See also Jackendoff (2002) for the use of stress and various syntactic constructions in conveying information structure – the partitioning of the message into presupposition vs. focus.

Footnote 2

A high pitch accent is a local rising pitch movement which lends perceptual prominence. In French, when a constituent is prosodically focused, a pitch accent can be observed on one of the syllables of the constituent. This pitch-accented syllable is also usually longer. Inter-speaker variability may be observed in the syllable bearing the pitch prominence.

Footnote 3

Some linguists (e.g. Chomsky, 1976; Horvath, 1986; Brody, 1990; Rizzi, 1997), argue for the existence of focus movement at LF (Logical Form) even if focus appears in-situ in the S-structure, like it does in French. In the movement approach to focus, a sentence like “Mary loves JOHN” should be analyzed as [JOHN_i; [Mary loves t_i]] with the focused constituent moved at the left periphery, and establishing a relation with its trace t_i. Other linguists (see Anderson, 1972; Roots, 1985; Costa, 1998), however, have arguments for an in-situ interpretation of focus, where no movement is required. In the present case – where focus is on the subject, and therefore on the left part of the sentence, (“MADELEINE m’amena”) – even if a movement analysis were correct, the leftward focus movement would not provide a different structure from the one in the baseline condition [MADELEINE [m’amena]].

Footnote 4

A grammatical transformation over a sentence involves the movement of a constituent from one position to another in the sentence. The position abandoned by the constituent is known as the “trace” [t], and is “bound” by that constituent (as in the transformation “The girl pushed the boy” → “The boy who the girl pushed [t].”). It is through the link between the trace and the constituent that thematic roles (agent, theme, goal, source, experiencer, etc.) are transmitted.

Footnote 5

According to some theories of syntax (Pollock, 1989), tense and agreement are located at distinct functional levels in the structure tree.

Preliminary version produced by the authors.

In Lævenbruck H., Baciú M., Segebarth C. & Abry C. (2005). *Journal of Neurolinguistics*, 18, 237-258.
doi:10.1016/j.jneuroling.2004.12.002 - **hal-00371865**

9

Pointing is ‘special’

HÉLÈNE LŒVENBRUCK
MARION DOHEN
CORIANDRE VILAIN

Abstract: Deixis, or pointing, is the ability to draw the viewer/listener’s attention to an object, a person, a direction or an event. Pointing is involved at different stages of human communication development, in multiple modalities: first with the eyes, then with the finger, then with intonation and finally with syntax. It is ubiquitous and probably universal in human interactions. The ‘*special*’ role of index-finger pointing in language acquisition suggests that all pointing modalities may share a common cerebral network. This chapter aims at better grounding linguistic pointing in somatosensory as well as cerebral domains and at suggesting that it shares features with other pointing modalities. It is shown that manual and ocular pointings seem to recruit left posterior parietal and frontal cortices. Then vocal pointing is presented in details, for both production and perception. It is suggested that integrated multisensory representations may be needed in order to produce and perceive prosodic pointing and that these representations require the activation of associative cerebral areas. The results of a previous study support this hypothesis by showing that prosodic pointing seems to recruit a left temporo-parieto-frontal network whereas grammaticalized syntactic pointing mainly involves frontal regions. The involvement of the left parietal lobe is also apparent in our preliminary fMRI study of the perception of prosodic pointing. Finally, the exploratory results of a new study of multimodal pointing (digital, ocular, prosodic and syntactic) are presented. The common left parietal activation in ocular, digital and prosodic pointing is discussed in the framework of the link between gesture and language.

1. INTRODUCTION

1.1. WHAT IS SO SPECIAL ABOUT POINTING?

Recent research works suggest that deixis, ostension or pointing is at the root of human communication (e.g. Bates *et al.*, 1975; Bruner, 1975; Corballis, 1991; Hewes, 1981; Kita, 2003; Rolfe, 1996; Tomasello *et al.*, 2007). Pointing is the ability to draw the viewer/listener's attention to an object, a person, a direction or an event. Pointing is a communicative device which orients the attention of another person so that an object/person/direction/event becomes the shared focus of attention. It serves to single out, to individuate what will become the referent. In linguistics, deixis is defined as the way language expresses reference to points in time, space or events (Fillmore, 1997). In its narrow sense, it refers to the contextual meaning of deictic words, which include pronouns (me, you), place deictics (here, there), time deictics (now, then) and demonstratives (this, that). In its broad sense, deixis is a referential operation, i.e. it provides a mean to highlight relevant elements in the discourse, to designate, identify or even select an element. It is therefore tightly linked with topicalization, focus and extraction (see e.g. Berthoud, 1990; Jackendoff, 2002). Therefore, pointing, in its broad sense, can be conveyed in several ways: gesturally, with specific hand gestures, or verbally, with specific phonatory and articulatory movements (prosodic deixis) or through the use of a specific syntactic construction (syntactic deixis).

Pointing is 'special'¹ because it is...

... ubiquitous,

A pointing gesture is most often performed with the index finger and arm extended in the direction of the interesting object and with the other fingers curled inside the hand (Butterworth, 2003). The 'canonical' index finger pointing is *ubiquitous* in everyday interactions, in most cultures of the world. It is observed in oral as well as signed communication. In some sign languages, index pointing is used linguistically for referential indexing. Noun phrases, for instance, may be associated with loci in space. Reference to a previously mentioned noun is performed by pointing again to its specific locus (Klima & Bellugi, 1988).

... universal,

Index finger pointing has been claimed to be a universal ability shared by all human beings (e.g. Povinelli & Davis, 1994). In some cultures, index finger pointing can be replaced with lip pointing (in the Barai in Papua New Guinea, see Wilkins, 2003). But pointing gestures, be they conveyed with the index-finger, the hand, or even the lip or the chin, remain a key *universal* communication tool in humans.

¹ The expression "Pointing is special" refers to the debate on the specialised nature of speech. This debate originates from Chomsky's (1957, 1966) position concerning the special nature of the human language faculty. This position has been taken over by researchers from the Haskins Laboratories, and in particular by Alvin Liberman, who wrote "On Finding that Speech Is Special" in 1982. During the summer school, the "speech-is-special" stance was the object of many questions and debates, driven by the Haskins Lab. alumnus Rudolph Sock.

... has a long historical past,

Although its scientific study is relatively recent, index-finger pointing did not emerge in the XXth century. Traces of it can be found in early European art. The most famous paintings of points are those of Leonardo da Vinci, dating back from the late XVth century (such as the *Virgin of the Rocks*, 1483-1486, with the angel Uriel pointing towards baby John-the-Baptist; or *St John the Baptist*, 1513-1516, with his right hand pointing up toward heaven). But even earlier points can be found, like in the *Bayeux Tapestry*, which was embroidered towards the end of the XIth century, and which tells the 1066 Norman invasion of England (e.g. William of Normandy, on his ducal throne, pointing at Harold making an oath; or men pointing at the ominous Halley's comet). Furthermore, pointing was not only used in European culture. Sculptures of pointing buddhas can be found in XVIIIth century Burmese art, for instance.

... performed by evolved animals,

It has long been claimed that pointing is a human-specific ability and that great apes, for instance, do not point (e.g. Povinelli *et al.*, 2003). Works by Leavens, Hopkins and colleagues show, however, that chimpanzees in captivity can in fact point at unreachable food (Leavens *et al.*, 2005). They sometimes use 'canonical' index finger pointing, but most often, they point with all fingers extended. And, like human infants, chimpanzees point spontaneously, without explicit training.

Therefore, although pointing does not seem to be specific to human primates – since apes, and other animals (think of the vocal reference calls of the vervet monkeys), may well be capable of it – it seems to be a *sophisticated social* behaviour, used by animals capable of some degree of imitation and inter-individual communication (Pollick & de Waal, 2007). Some researchers have even suggested that the index-finger pointing ability might be related to the thumb-index finger opposition and the pincer grip capacity (Butterworth, 2003).

... multimodal,

Pointing involves multiple modalities: it can be manual or digital (with the index finger), labial, facial (chin), ocular, and vocal.

As described above, pointing is most often, at least in Western cultures, digital, i.e. it is conveyed with an extended index finger. It is also very often manual, i.e. the whole extended hand is used to point. But when the hand is used, the handshape may vary (open flat hand held palm up, or palm facing laterally, thumb: see Kendon, 1996). Pointing can also be performed with protruded lips. Labial pointing has been observed in several communities in all inhabited continents (the Kuna indians in Panama, the Arrernte in Australia, the Awtuw and Barai in Papua New Guinea, the Ewe in Ghana, the Navajo in North America; cf. Wilkins, 2003). In fact, as Kendon (1996) notes, many different body parts can be used to perform pointing (head, lip, chin, elbow, foot, arm and hand).

Ocular pointing or deictic gaze is also frequent. It is the ability to (alternately) look at the 'object' then at the viewer's eyes. It is an invitation for the viewer to look at the object which becomes the shared focus of attention. A wonderful example of ocular pointing is the famous

painting by Georges de la Tour, “Le Tricheur à l’as de carreau” (The cheat with the ace of diamonds, 1635). According to our interpretation of this painting, the woman standing next to the cheater gazes at him, thereby designating him for the woman sitting at her left. The sitting lady understands the ocular point, as she responds with a digital point.

Vocal pointing is what is sometimes called deixis or focus. In French, as in many languages, vocal deixis can be conveyed by syntactic extraction, using a deictic presentation form such as in the example below:

C’est Madeleine qui m’amena (It’s Madeleine **who** brought me along.)

It can also be conveyed by prosodic focus, i.e. by using a specific intonational contour on the pointed item, such as in the example below:

MADELEINE_F m’amena. (MADELEINE_F brought me along.)

The effect of this intonational contour, which will be described in more detail in section 2.1., is to highlight the pointed item (‘Madeleine’), the rest of the utterance bearing a flat, post-focal contour.

... *involved in stages of language development*

Ocular pointing (or deictic gazes, at 6 - 9 months) and, later, index finger-pointing (deictic gestures, at 9 - 11 months) have been shown to be two key stages in infant cognitive development that are correlated with stages in oral speech development.

Pointing with the eye is first observed at 8 to 9 months, mutual attention takes place between the adult and the baby. When the adult looks at something, the baby follows his/her gaze and looks in the same direction. At this stage, babies therefore have the ability to look where someone else is looking. Conversely, the baby seems to invite the adult to look at an object, by alternately looking at the object then at the adult’s eyes. Quite early in development, babies can therefore use gaze to manipulate the attention of their carer.

Then, at 9 to 11 months, when infants start to be able to understand a few words, they produce pointing gestures, most often *index-finger pointing*. The emergence of pointing is a good predictor of first word-onset, and gesture production is related to gains in language development between 9 and 13 months (Bates *et al.*, 1979; Butcher & Goldin-Meadow, 2000; Caselli, 1990). As explained in Butterworth (2003), pointing not only serves to single out the object but also to build a connection between the object and the speech sound. Finger pointing therefore seems clearly associated with lexicon construction.

Then at 16 to 20 months, during the transition from the one-word stage to the two-word stage, *combinations of words and deictic gestures* can be observed (such as a pointing gesture to a location and pronouncing ‘dog’, to indicate that a dog is there). Furthermore, the number of gestures and gesture-word-combinations produced at 16 months are predictive of total vocal production at 20 months (Morford & Goldin-Meadow, 1992; Capirci *et al.*, 1996; Goldin-Meadow & Butcher, 2003; Volterra *et al.*, 2005). Finger pointing therefore clearly seems associated with morphosyntax emergence.

As for *vocal pointing*, only a few studies have examined the development of *prosodic focus* and *syntactic extraction* in children.

Concerning the *acoustic realisation of prosody*, pre-school children have been shown to master contrastive focus in English, in absence of any formal teaching (Hornby & Hass, 1970). Little is known about the development of prosodic focus in French, but the ability to realise adult-like pitch movements emerges quite early (Konopczinsky, 1986). Also, it has been shown that adult-like syllable lengthening is realised by young children (from 2 years of age) on accented vowels (Konopczinsky, 1986).

As concerns *articulatory production of prosodic pointing*, what is known is that lexical stress in English is differently realised in young children than in adults (Goffman & Malin, 1999; Connaghan *et al.*, 2001). Children aged 3 to 6 years tend to produce larger movements, with reduced velocities, and increased durations, characteristics of slowly and carefully articulated speech. In a recent video study on French, Ménard *et al.* (2006) showed that French-speaking children (aged 4 and 8) do not differentiate articulatory and acoustic patterns across focused and unfocused syllables as much as adult speakers do. To summarise the findings, although the production of prosodic focus, or vocal pointing, seems to be mastered quite early (as early as 3 years of age) in the acoustic domain, and without formal teaching, its articulatory correlates seem to be acquired much later. But what seems lacking in children is not the capacity to hyper-articulate focused phrases, but to hypo-articulate surrounding phrases. The delay in articulatory performance could thus be related to general articulatory proficiency rather than to specific linguistic mastery. Since manual pointing emerges spontaneously before the end of the first year, it could be that acoustic correlates of prosodic focus are in fact mastered much earlier than 3 years of age.

The development of syntax in young children has been extensively studied by Tomasello and colleagues, among others (see also Brown, 1971; Sheldon, 1974; Tavakolian, 1981; MacWhinney & Pléh, 1988; McKee *et al.*, 1998; Kidd & Bavin, 2000). In a study of the speech of English-speaking children between 1;9 and 5;2 years of age, Diessel & Tomasello (2000) showed that the earliest and most frequent relative clauses that children learn occur in *presentational constructions* that are propositionally simple. They express a single proposition in two finite clauses, such as: "Here's a tiger that's gonna scare him" or "That's the sugar that goes in there". According to the authors, one of the factors that might explain why relative clauses emerge in these presentational constructions is the prefabricated character of the main clause. Interestingly, the main clause contains a deictic pronoun (i.e. this, that, here, there, it). Diessel & Tomasello argued that "*the early use of relative constructions involves a very simple procedure by which the child combines a prefabricated (main) clause (i.e. a clause of the type That's X, There's X, It's X) with a second component [...].*" (p.144). They observed presentational constructions in children aged 2 and younger (1;11). They quoted a study by Jisa and Kern (1998) on French children (aged 5;0-5;11) who also reported extensive use of presentational constructions in young children.

To sum up, pointing is involved at several stages of human communication development. It seems to be one of the first communicative tools used by babies. It is a key part of the shared

attention mechanism in child-adult interaction. It seems to emerge spontaneously, in stages, and in association with oral productions. The crucial role of index-finger pointing in language development and the involvement of multiple forms of pointing at different stages of human communication development, first with the eyes, then the finger, then intonation and finally syntax, suggest that all pointing modalities may share a common cerebral network. The aim of this chapter is to better ground linguistic pointing in somatosensory and cerebral domains and to show that it shares features with other pointing modalities.

First, we will discuss the results of several studies in the literature related to manual and ocular pointing that can provide clues on the cerebral correlates of these two pointing modalities and can shed light on the potential general pointing network.

Then we will focus on oral linguistic pointing. Among the two forms of linguistic pointing, prosodic focus will be described in particular, since it bears interesting physiological characteristics. Its acoustic and articulatory correlates will be given in detail. It will be shown that prosodic focus is signalled using very specific acoustic and articulatory features and that speakers use reliable strategies to organise phonation and articulation together in order to convey prosodic pointing. On the perception side, results from auditory-visual experiments will be presented that indicate that the phonatory and articulatory patterns are not just productive habits but that they are in fact well recovered by listeners and viewers. It will be suggested that integrated representations (acoustic, articulatory, proprioceptive) may be needed in order to produce prosodic pointing adequately and that these multisensory representations may be formed via the activation of associative cerebral areas. It will be hypothesised that these areas should also be involved during the perception of prosodic focus. Two preliminary fMRI studies of the production and perception of prosodic focus will be presented. Finally, we will present the first results of an fMRI study on the different pointing modalities that suggest that manual, ocular and prosodic pointing may well be grounded in a same cerebral network.

The organisation of the chapter is as follows. In section 1.2., a review of the literature provides elements for a description of the cerebral networks of manual and ocular pointings. These networks are the grounding red thread for the study of oral pointing presented in section 2. A summary of several works carried out in our laboratories on the production and perception of prosodic pointing, both in the acoustic and articulatory (or visible) domains, is provided in sections 2.1. and 2.2. This summary is then used to make hypothesis on the potential cerebral correlates of prosodic pointing in both production and perception. Results of a previous work on the cerebral correlates of the production of prosodic and syntactic pointing are provided in section 2.3.1. A work-in-progress on the cerebral correlates of the auditory-visual perception of prosodic focus is presented in section 2.3.2. Finally, preliminary results of an fMRI study of pointing in different modalities (manual, ocular, prosodic, syntactic) are presented in section 3. The implications of these findings are discussed in the conclusion.

1.2. MANUAL AND OCULAR POINTING

As explained above, the crucial role of pointing in communication, be it with the eyes, the finger, intonation or syntax, suggests that all pointing modalities may be grounded in a common cerebral network. The first question we want to address is: what are the cerebral correlates of manual and ocular pointing? Several observations provide preliminary answers to this question. The first set of observations deals with the role of the posterior parietal regions of both hemispheres in manual pointing tasks. It has been suggested that the spatial representations formed in the posterior parietal and premotor frontal regions could provide “*perceptual-premotor interfaces for the organization of movements (e.g. pointing, locomotion) directed towards targets in personal and extrapersonal space*” (Vallar, 1997, p.1401). Patients with left unilateral neglect have been shown to present deficits in pointing tasks (e.g. Edwards & Humphreys, 1999) while PET studies on normal subjects show activation within the left and/or right inferior parietal lobule (IPL) during pointing tasks (e.g. Lacquaniti *et al.*, 1997; Kertzman *et al.*, 1997). The role of the right and left posterior/inferior parietal regions in pointing tasks may further be related to data on brain-damaged deaf signers. Bellugi and colleagues presented a study of deaf signers of American Sign Language (ASL), two of which presented lateralized parietal lesions, one in the right hemisphere and the other in the left (Bellugi *et al.*, 1989). Space in ASL is handled in two ways. The first is topographic: in the description of the layout of objects in space, spatial relations among signs reproduce the actual spatial relations among the objects. The second is deictic: space is used for referential indexing (as explained in the introduction). The right-lesioned signer had difficulty in the use of space for topography: room description was distorted spatially, with left side of signing space neglected. In the use of space for syntax, however, the entire signing space was covered and consistent reference to spatial loci was preserved. By contrast, the left-lesioned signer produced room descriptions without spatial distortions but made errors in the deictic use of space.

More recently a study on *finger pointing and looking* suggests that pointing with the finger and pointing with the eye seem to recruit a common left lateralized network including the frontal eye field and the posterior parietal cortex (Astafiev *et al.*, 2003). In this study, the manual pointing task included a preparation phase and an execution phase. During the preparation phase, the subjects had to prepare to point at a cued location with the right index finger. During the execution phase, the subjects were asked to point towards the target as soon as the target flashed. The saccade task also included preparation and execution phases. Since the instruction was to (prepare to) look at a cued location, it can thus, according to us, be assimilated to ocular pointing. *In saccade preparation*, a transient sensory response to the cue was observed in the bilateral occipital cortex, a sustained response was observed in the bilateral frontal cortex, at the junction of precentral and superior frontal sulci, i.e. in the human frontal eye field (FEF) and a sustained response was observed along the horizontal segment of the intraparietal sulcus (IPS). *In pointing preparation*, bilateral FEF and IPS activations were observed, just as in saccade preparation. These results therefore suggest that the posterior parietal cortex and the frontal cortex contain regions that code preparatory signals for pointing independently of the effector used (eye, index-finger). They are consistent with previous fMRI experiments that reported common activity in the IPS and FEF for visually guided saccades

and pointing movements (Connolly *et al.*, 2000; Simon *et al.*, 2002). Additional left hemisphere activation was observed for the manual pointing task, in the angular gyrus, the supramarginal gyrus, the superior parietal lobule, the dorsal precentral gyrus, and the superior temporal sulcus. The authors checked that the left-lateralization held, even when the left hand was used. Therefore manual pointing-specific responses were observed in the left lateral and medial posterior parietal and frontal cortex. Similar posterior parietal and superior frontal activations were observed for eye as well as arm pointing tasks by Hagler *et al.*, 2007. But no evidence was found for manual-pointing-specific maps.

To summarise, recent neuroimaging studies provide clues to the description of the cerebral networks involved in manual and ocular pointing. They suggest that posterior parietal cortex and frontal cortex are bilaterally activated in ocular and manual pointing tasks. Additional left lateralized activation in the medial posterior parietal cortex could be specific to manual pointing. The question we want to address now is whether a similar network is involved in linguistic pointing. First, we will review several studies carried out in our laboratory that indicate what the speaker's somatosensory representations might be for linguistic pointing as well as to what extent the listener/viewer is able to decode audiovisual signals of vocal pointing. This will enable us to make hypothesis as to which regions the cerebral network for the production and perception of vocal pointing may include.

2. WHAT DO WE KNOW ABOUT VOCAL POINTING?

As mentioned in the introduction, vocal pointing in French (or deixis) can be conveyed by syntactic extraction or prosodic focus (Berthoud, 1990). Prosodic focus, as will be shown in the following, involves specific intonational and articulatory patterns. Syntactic extraction involves the use of a cleft form ("*c'est qui*", "*it's ... who*"). It may or not be accompanied by prosodic focus. It should be mentioned here that there are two types of prosodic focus in French: *contrastive focus* which selects a constituent in the paradigmatic dimension and *intensification focus* which makes a contrast along the syntagmatic axis (Séguinot, 1976). This chapter is only concerned with *contrastive* prosodic focus and the term 'prosodic focus' will be used hereafter to refer to this type of focus.

Compared with a broad focus or non-deictic rendition of an utterance, *prosodic* focus involves precise acoustic and articulatory modifications, whereas *syntactic* extraction may simply use additional words, but no specific somatosensory change. We will first describe what we know about the somatosensory characteristics of *prosodic* focus, namely its acoustic and articulatory correlates. We will also examine how prosodic focus is perceived visually. This will lead to conjectures on the cerebral regions involved in the production and perception of prosodic pointing.

2.1. ACOUSTIC CORRELATES

Acoustic correlates of contrastive prosodic focus have been extensively described (see e.g. Astésano, 2001; Astésano *et al.*, 2004; Dahan & Bernard, 1996; Delais-Roussarie *et al.*, 2002; Di Cristo, 1998; Di Cristo & Jankowski, 1999; Dohen & Løevenbruck, 2004; Jun & Fougeron, 2000; Rossi, 1999; Touati, 1987). Prosodic deixis in French involves an increase in fundamental frequency (F0) on the focused constituent as well as a lengthening of the focal constituent. It

is also associated with an F0 lowering on the pre-focal constituent and a deaccentuation of post-focal constituents. Figure 1 illustrates the effect of prosodic focus on F0. Figure 1a shows the spectrogram and F0 for the neutral rendition of the sentence “[MADELEINE]_{AP} [m’*amena*]_{AP}” (MADELEINE brought me along). Figure 1b shows the same utterance with a focused subject. F0 raising on the first syllable of the Accentual Phrase² ‘Madeleine’ and post-focal deaccentuation are clearly visible. Metrical and rhythmical aspects of prosodic focus have also been studied. Following Di Cristo, Astésano (2001) postulates that the emphatic initial accent in contrastive focus is the ‘hyper’ realisation of the metrical, secondary initial accent in French. In that view, the initial contrastive accent shares common phonetic features with the metrical initial accent: both have similar infra-syllabic lengthening, with significant lengthening of the onset over the rhyme, and similar overall F0 configurations, which clearly distinguish them from final accents. However, these acoustic correlates are twice as large for the emphatic initial accent as for the metrical initial accent, confirming that it is a ‘hyper’ version of the initial accent (Astésano *et al.*, 2007). In addition to the increase in duration of the focused phrase (typically of the focused syllable onset), an anticipatory increase in duration of the pre-focal syllable can sometimes be observed (Dohen & Løevenbruck, 2004).

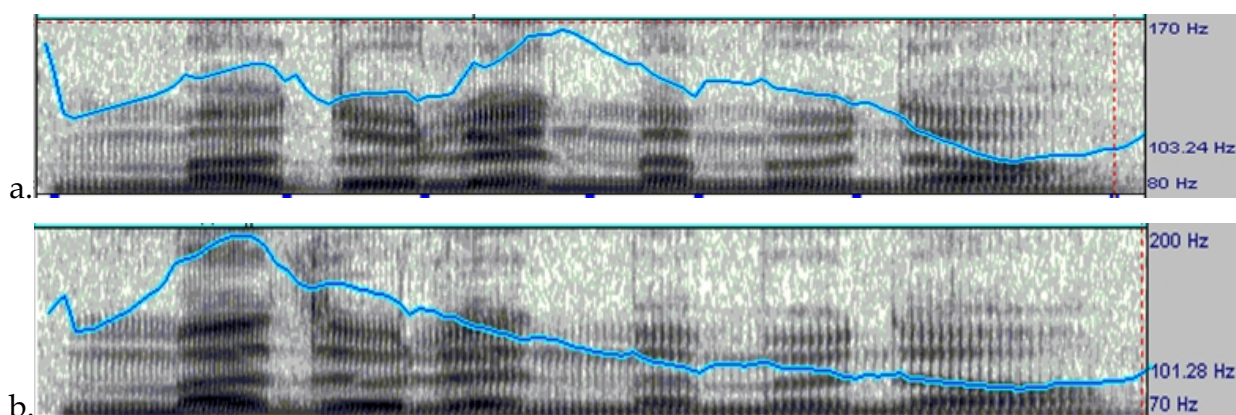


Figure 1: a) Spectrogram with superimposed F0 trace for a neutral rendition of the sentence “Madeleine m’amina”. b) Focused rendition of the same sentence. The increased F0 peak on the first syllable /ma/ is typical of a focused phrase. The post-focal F0 trace falls to reach a flat floor. The second (highest) F0 peak observed in the neutral rendition (typical of the end of an Accentual Phrase) is suppressed.

But there is no decrease in the duration of the post-focal items (no dephrasing). Restructuring of the focused Accentual Phrase is sometimes observed: it can be grouped with other phrases into one Intonational Phrase or divided into several Accentual Phrases.

2.2. ARTICULATORY CORRELATES, INCLUDING VISIBLE CUES

² For the definition of an Accentual Phrase (AP) in French, see Jun & Fougeron (2000). The AP has the default tonal representation /LHi LH*/, with an initial high tone Hi (also described as the peak of the ‘accent secondaire’), a final high tone H*, realised on the phrase-final full syllable (peak of the ‘accent primaire’), and two low L tones realised on the syllable preceding the H-toned syllable.

Summarised results of two research works on French, carried out in our laboratory, are presented here. The first one concerned tongue movements and was carried out using an electromagnetometer. The second set of studies examined lip and other facial movements. It used video and optical recordings. Perceptual tests checked whether the articulatory cues are perceived visually.

2.2.1. TONGUE MOVEMENT

Articulatory movements of the tongue were tracked using an electromagnetometer during the production of prosodic focus in several conditions (Lœvenbruck, 1999). Detailed methods are provided in Appendix A.1. A preliminary F0 analysis showed that, in all focused utterances, the F0 peak was carried by the phrase-initial /la/ syllable. The articulatory characteristics of the phrase-initial /la/ syllable in the neutral conditions were thus compared with those of the focus conditions. Articulatory analyses, shown in Figure 2, consisted in measuring maximal tongue displacement and peak velocity within the syllable.

The displacement data presented in Figure 3a show that, for this speaker, the tongue displacement in the /la/ syllable is larger in focused than in unfocused conditions, when the syllable belongs to a word that is not utterance initial. This means that the tongue movement is larger under focus in these sentence positions. When the word is utterance-initial, the tongue displacement is not larger when the syllable is focused. This is probably due to the fact that word-initial syllables in utterance-initial position are often articulated with more amplitude, even in unfocused contexts (see e.g. Fougeron & Keating, 1997; Tabain, 2003).

The results of the duration data, presented in Figure 3b, show that, for this speaker, the duration of the /la/ syllable is longer in focused than in unfocused conditions, when the syllable is not utterance initial. For utterance initial syllables, there is no increase in duration when the syllable is focused. This is probably due to the fact that word-initial syllables (particularly the onsets) in utterance-initial position are often lengthened, even in unfocused conditions (see Tabain, 2003).

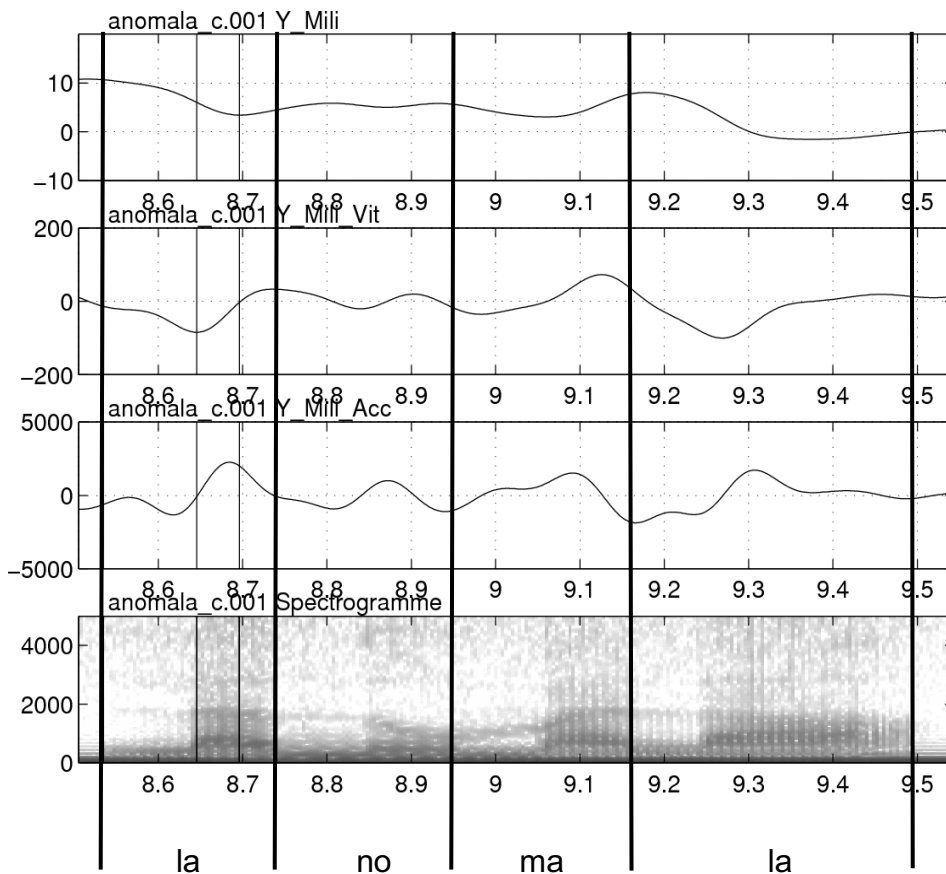


Figure 2: Articulatory labels for the first syllable of [lanomala] in ‘Elle annihilait l’anomala en l’éloignant’. Top: tongue-middle vertical position (in mm); Middle: velocity (mm/s); Bottom: acceleration (mm/s²). The dark vertical lines are acoustic landmarks for syllables [la], [no], [ma] and [la]. The disconnected lines are articulatory labels within the first syllable [la].

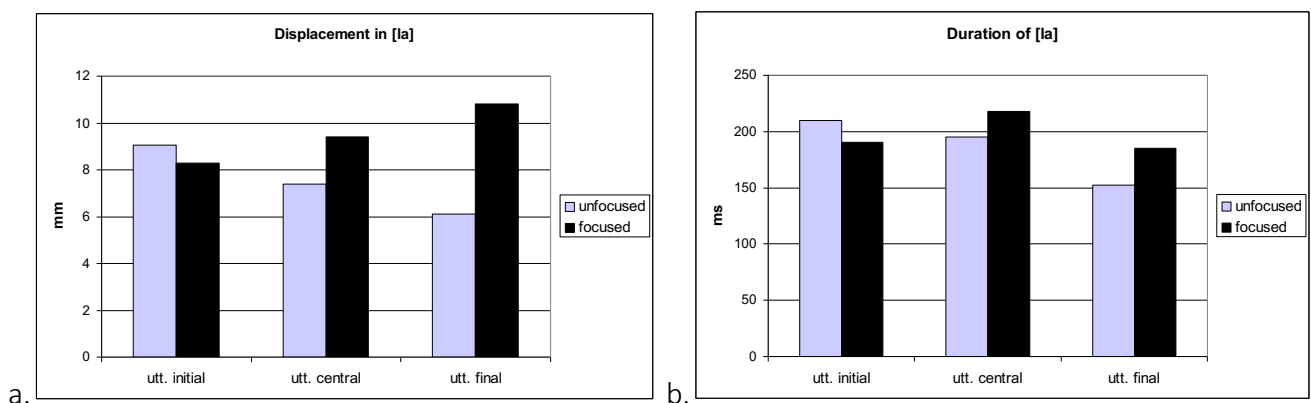


Figure 3: **a)** Tongue vertical displacement (mm) for /la/ in unfocused (light bars) and focused (dark bars) conditions, as a function of the position of the phrase in the utterance. **b)** Duration of /la/ in unfocused and focused conditions.

2.2.2. LIP CUES

In a second set of studies, movements of the lips were tracked during the production of vocal pointing (prosodic contrastive focus) using different measurement techniques.

a. The video study

The results summarised below are detailed in Dohen *et al.* (2004) and Dohen & Løevenbruck (2005). The aim was to examine the possible articulatory visible correlates of contrastive focus in French. Data were collected for two speakers (A & B) for reiterations (A) and real speech (A & B) using a very accurate lip-tracking device (descriptions can be found in Audouy, 2000). It appeared that contrastive focus was characterized by an increase (hyper-articulation) in inter-lip area and in inter-lip area peak velocity for the focal phrase. The amount of hyper-articulation was however highly speaker-dependent. Speaker A hyperarticulated much more than speaker B. Due to corpus constraints, protrusion could be analysed for speaker B only. It was also hyper-articulated and to a greater extent than inter-lip area. In addition, it appeared that speaker B hypo-articulated the post-focal sequence (reduced inter-lip area, inter-lip area velocity and protrusion) while speaker A barely did. Durational measurements were also conducted since duration can be a visual cue as well. These measurements showed that the focused syllables were significantly lengthened for both speakers, the first phoneme of the focused phrase being even more significantly lengthened. For speaker A, it was also observed that the last syllable of a phrase was significantly lengthened and hyper-articulated when the following phrase was focused. This was related to an anticipation strategy. We concluded from these observations that there is a global tendency towards hyperarticulating the focused phrase but that other visible cues are produced and that they appear to be speaker dependent. This is why, in order to be able to identify potential generic strategies, it seemed important to extend this study to a greater number of speakers.

Visual only perception tests were also conducted using the videos of speakers A and B (Dohen & Løevenbruck, 2005). These tests showed that contrastive focus could be perceived through the visual modality alone and that the visual cues used for perception corresponded at least in part to those identified in the production studies. For both speakers, a few stimuli were well perceived even though the visible correlates described above (i.e. focal hyper-articulation and lengthening, with or without post-focal hypo-articulation) were not present. This suggested that other more subtle facial correlates may intervene. Studies on other languages have indeed shown that other facial movements such as eyebrow movements (Krahmer *et al.*, 2002; Grandström & House, 2005) or head movements (Hadar *et al.*, 1983; Cerrato & Skhiri, 2003; Munhall *et al.*, 2004) or both (Graf *et al.*, 2002) could intervene. Cavé *et al.* (1996) also showed that, in French, F0 variations and eyebrow movements could be linked. This is why it also seemed necessary to enlarge the set of facial movements measured by the use of a complementary technique.

b. The 3D motion capture (Optotrak) study

The aim of this second study, detailed in Dohen *et al.* (2006), was to further explore the articulatory visible correlates of contrastive focus in French, by extending the preliminary study in two directions. The first extension consisted in including a greater number of speakers in order to identify possible generic strategies. The preliminary study showed that one speaker's way to single out the focused phrase was to locally increase inter-lip area and its derivative as well as duration. The other speaker's behaviour was spread across the entire utterance. It consisted in creating a contrast within the utterance, by slightly hyper-articulating and lengthening the focused phrase as well as hypo-articulating the post-focal sequence.

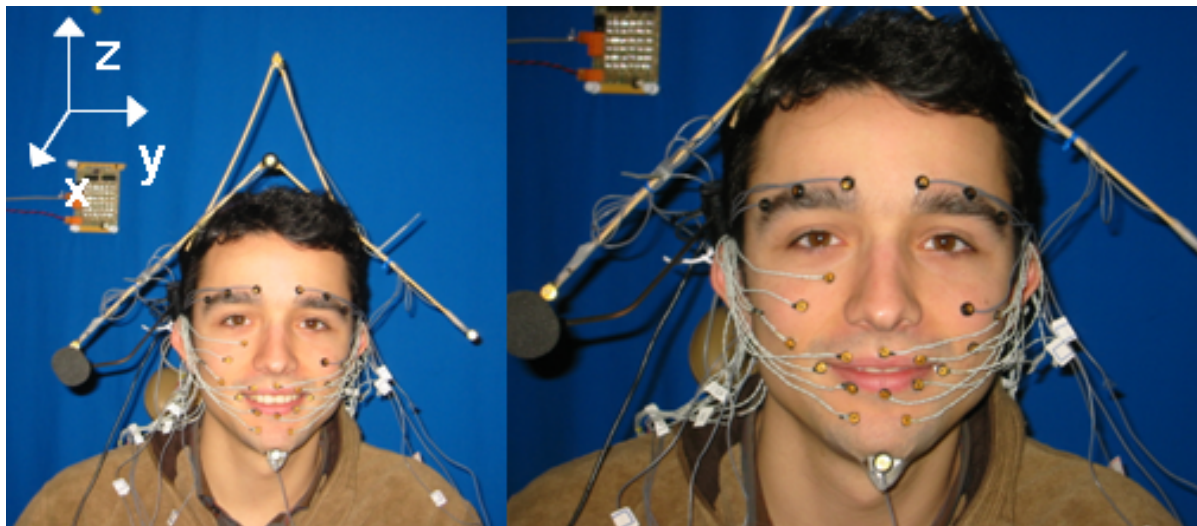


Figure 4: Optotrak measurement device: experimental setup. The 24 IREDs glued to the speaker's face are visible, as well as the 4 additional IREDs attached to a head rig for head motion correction. The x, y, z axes used for the articulatory measurements are shown. See Appendix A.2.

The first aim of this second study was thus to determine whether the two behaviours observed are shared by other speakers or if more variability is at stake. The second addition to the preliminary study consisted in extending the set of facial measurements by the use of a complementary recording technique. Results of the perceptual tests in the preliminary study suggested that other facial correlates than lip parameters may play a role in focus perception. The second study made use of an optical technique that allowed for the monitoring the positions of several markers on the face. Details on the experimental paradigm, the recording procedure and the data processing can be found in Appendix A.2. Five native speakers of French (B, C, D, E and F) were recorded using a 3D optical tracking system: Optotrak (IRED tracking system using markers glued to the speaker's face), which is less accurate on lip contours than the system used in *a.* but provides more facial data. Figure 4 gives an idea of the experimental setup used.

Articulatory and durational analysis

Figure 5 provides a visual overview of the results. Table 1 groups detailed intra- and inter-utterance contrasts (percent changes for each speaker and each parameter).

Speaker B – This speaker significantly lengthened and hyper-articulated (except lip spreading) the focused phrase. He also hypo-articulated the post-focal phrase(s). The

strongest contrasts were measured for lip protrusion and duration and the smallest contrasts for jaw movements.

Speaker C – This speaker significantly lengthened and hyper-articulated the focused phrase. A slight hypo-articulation of the post-focal phrase(s) was also observed but only for jaw movements and lip opening. This speaker anticipated focus on the preceding phrase (slight lengthening and hyper-articulation for jaw movements and lip opening limited to the directly preceding phrase). The strongest contrasts were measured for lip protrusion and duration.

Speaker D – This speaker significantly lengthened and hyper-articulated (except lip spreading) the focused phrase. A post-focal hypo-articulation was measured for lip opening and lip protrusion. This speaker anticipated focus by hyper-articulating (only lip opening) the directly preceding phrase. The strongest contrasts were measured for lip protrusion and the smallest contrasts for lip spreading.

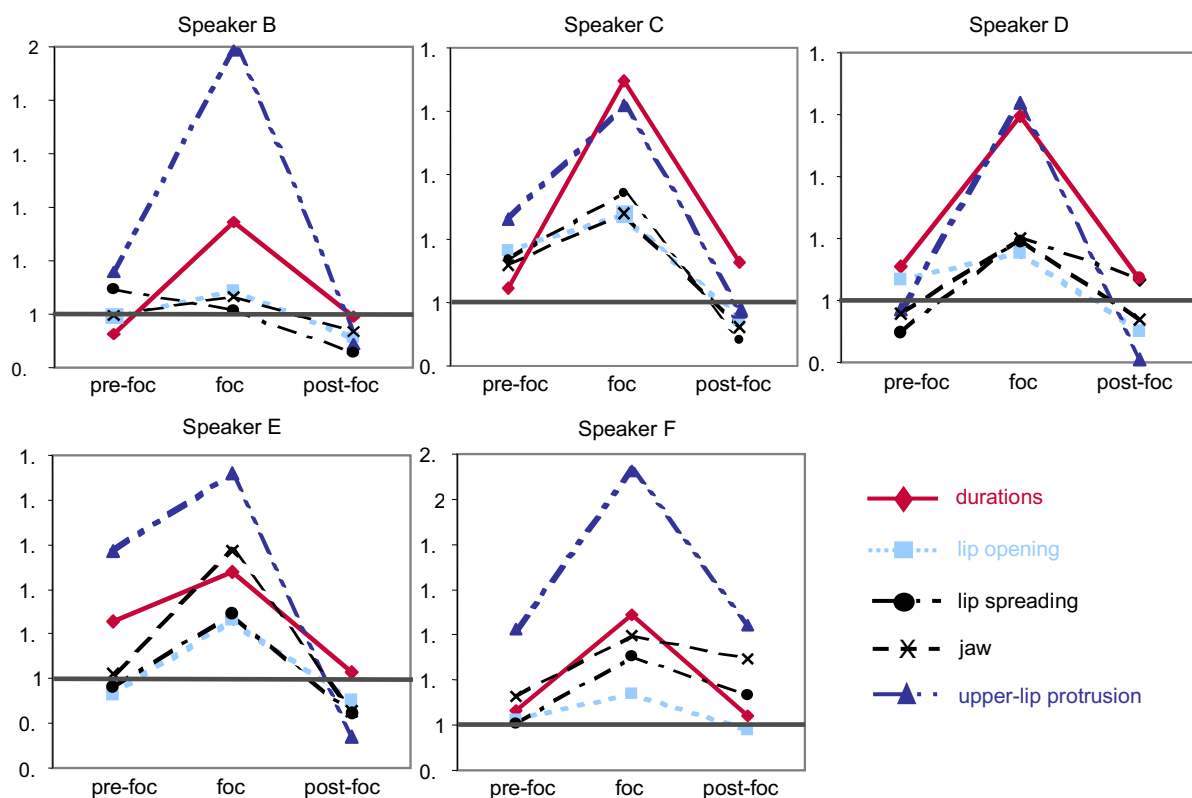


Figure 5: Durational and articulatory measurements for all five speakers: normalised values corresponding to the pre-focal (pre-foc), focal (foc) and post-focal (post-foc) sequences (the dark horizontal lines correspond to the neutral case, i.e. 1).

Speaker E – This speaker significantly lengthened and hyper-articulated the focused phrase. He also hypo-articulated the post-focal phrase(s) and anticipates focus by hyper-articulating (only lip protrusion) the directly preceding phrase. The strongest contrasts were measured for lip protrusion and the smallest contrasts for lip opening and lip spreading.

Speaker F – This speaker significantly lengthened and hyper-articulated the focused phrase. He also anticipated focus by hyper-articulating (only lip protrusion) the directly preceding

phrase. The strongest contrasts were measured for lip protrusion and the smallest for lip opening.

Analysis of the other facial data

Eyebrow movement (raising) – There appeared to be a link between eyebrow raising and the production of prosodic contrastive focus only for three out of five speakers (B, C & E). This eyebrow raising was however not systematic and did not occur whenever focus is produced. Speaker B was the one for which the combined productions (synchronous eyebrow raising and focus production) were the most frequent. The amplitudes of the movements are however very small (largest movement: 2mm). The other speakers either never raised their eyebrows, or did it randomly with no particular link to focus production.

Table 1: Results from the Optotrak study for each speaker (B, C, D, E and F) and each parameter: duration (Dur), vertical jaw movements (Jaw), lip opening (LO), lip spreading (LS) and upper lip protrusion (LP). The table provides the percent changes for the focused phrase compared to the rest of the utterance (intra-utterance contrast) and to the same phrase in the neutral version (inter-utterance contrast). The statistical significance threshold was fixed at $p=0.01$. ‘ns’ corresponds to a non-significant change.

speaker		B	C	D	E	F	
Intra-utterance contrast	Dur	+38.7%	+30.5%	+25.3%	+16.8%	+43.8%	
	Jaw	+9.9%	+14.9%	+12.5%	+31.5%	+18.5%	
	LO	+13.8%	+16.8%	+5.1%	+17.4%	+13.5%	
	LS	ns	+17%	ns	+19.4%	+23.2%	
	LP	+62.8%	+25%	+37.4%	+39.4%	+69.1%	
Inter-utterance contrast	Dur	foc	+20.9%	+34.1%	+29.8%	+23.9%	+49%
		pre-foc	ns	+4.7%	+8.1%	+16.4%	+9%
		post-foc	ns	ns	ns	ns	ns
	Jaw	foc	+6.6%	+14%	+9.9%	+28.5%	+39.5%
		pre-foc	ns	+6%	ns	ns	ns
		post-foc	-6.5%	-8.7%	ns	-6.2%	+23.9%
	LO	foc	+8.7%	+13.9%	+7.6%	+13.3%	+13.6%
		pre-foc	ns	+6.3%	+5.7%	-4.1%	ns
		post-foc	-8.9%	-2.2%	-4.5%	-3.4%	ns
	LS	foc	ns	+17.3%	ns	+14.5%	+30%
		pre-foc	ns	ns	ns	ns	ns
		post-foc	-15.6%	ns	ns	-5.8%	ns
LP	foc	+98.4%	+30.9%	+32%	+46%	+112.4%	

pre-foc	+25.1%	ns	ns	+24.4%	+53.7%
post-foc	-8.1%	ns	-10%	-13%	+41.6%

Head movement – Speaker B was the only one for whom a correlation between head nods and focus production was observed. The correlation was not systematic, however, and the amplitudes and temporal alignment of the movements were highly variable. The other speakers also moved their heads but the movements were not correlated with focus production.

Perception tests

Visual perception tests were carried out on a selection of recordings from four of the speakers. These showed that focus could be correctly detected visually in 66.4% of the cases (chance level 33.3%) and suggested that the articulatory and facial visual cues identified above are well perceived visually.

c. Summary of findings on the articulatory signalling of focus

The production study described above along with that described in the previous section show that there are potential visible articulatory correlates to prosodic contrastive focus in French. One of our main conclusions is the fact that focus affects the whole utterance and not only the specific focused phrase. A number of articulatory gestures are affected by focus. The way and the extent to which these articulatory gestures are affected depend on the speaker. However, two main strategies emerge, an absolute one and a differential one:

Absolute visual signalling strategy: the focused constituent is lengthened and hyper-articulated to a large extent (inter-lip area, protrusion and jaw movements). Previous studies (Dohen & Løevenbruck, 2005) showed that the peak velocities were also increased which signals an increase of the underlying articulatory effort during the gestures (Nelson, 1983). The speakers using this strategy therefore concentrate their efforts on the hyper-articulation of the focused phrase. Some speakers also slightly anticipate focus.

Differential visual signalling strategy: in this case, the focused phrase is also lengthened and hyper-articulated but to a smaller extent. Focus is sometimes anticipated. Additionally, the post-focal sequence is hypo-articulated compared to the neutral case. A visible contrast is thus created within the utterance: the focal hyper-articulation is not very distinct but is reinforced by the post-focal hypo-articulation.

We further observed that the visible articulatory parameter that was the most hyper-articulated under focus was protrusion. This is consistent with the finding that lip protrusion is the most visible lip feature (Benoît *et al.*, 1994). It is a robust feature which is long-anticipated in the articulation of the preceding phonemes (Cathiard, 1994). We also found that there could be a link between prosodic contrastive focus and head (nod) and/or eyebrow (raising) movements. However this link is far from being systematic, particularly for the head movements. There are important inter- and intra-speaker variations in the movement amplitude and in its synchronisation with the acoustic signal.

d. Contribution of articulatory and facial visual cues in global perception

Combined auditory-visual perception studies detailed in Dohen & Lœvenbruck (under revision) analysed the potential contribution of visual information in global auditory-visual perception of prosodic contrastive focus in French. Whispered speech was used to ‘naturally’ degrade part of the acoustic prosodic cues (no fundamental frequency for whispered speech) and thereby the auditory only perception performances. This rendered the measurement of a potential visual advantage possible by avoiding a ceiling effect. The studies showed that the visual information available from articulatory and facial movements was combined to auditory information especially when auditory information was not sufficient to make a straight-forward perceptual judgment (whispered speech). It therefore appears that articulatory and facial cues to prosodic focus are not only produced but that these cues can be used in the perception process.

Altogether, these results suggest that speakers use reliable strategies to organise phonation and articulation adequately in order to convey prosodic pointing. On the perception side, the visual only and auditory-visual perception tests show that these phonatory and articulatory patterns are recovered by listeners/viewers and used for prosodic pointing detection. Our conjecture is that integrated (acoustic, articulatory, proprioceptive) representations may be needed in order to produce prosodic pointing adequately and that these representations may be formed via the activation of associative cerebral areas, such as temporal and/or parietal regions. The fact that manual and ocular pointing both recruit the (associative) posterior parietal cortex adds to the motivation to explore the cerebral correlates of prosodic pointing.

2.3. CEREBRAL CORRELATES OF VOCAL POINTING

2.3.1. PRODUCTION FMRI STUDY

The fMRI production study summarized below is described in more details in Lœvenbruck *et al.* (2005). The aim was to examine the cerebral correlates of vocal pointing in French, conveyed by prosodic focus and syntactic extraction. Sixteen healthy, male, right-handed native speakers of French were examined. The stimuli consisted of visually presented sentences in French. Three isosyllabic sentences were presented, one for each condition:

- baseline condition: “Madeleine m’amena” (Madeleine brought me around).
- prosodic deixis condition: “MADELEINE m’amena” (MADELEINE brought me around), to elicit contrastive focus on the agent.
- syntactic deixis condition: “C’est Mad’leine qui m’am’na” (It’s Mad’leine who brought me ‘round).

The number of syllables in the sentence was maintained equal to 6, using schwa deletion. The methods are described in Appendix A.3.

The results of the fixed effect group and random effects analyses are reported here. Figure 6 represents the functional activations obtained for the main effects with the fixed effect analysis. The pattern of activations common to the two deixis conditions (each compared to the baseline) included Broca’s region (BA 45, 47), the left insula and the premotor cortex (BA 6) bilaterally. Prosodic deixis additionally activated the left anterior cingulate gyrus (BA 24,

32), the left supramarginal gyrus (LSMG, BA 40) and the left postero-superior temporal gyrus (Wernicke's area, BA 22). The (prosodic deixis - syntactic deixis) contrast yielded significant activation in the left posterosuperior temporal gyrus and the LSMG. The results of the random effect analysis, using the same statistical significance threshold ($p < .001$ corrected), for the same contrasts did not provide significant activations. With a less stringent significance threshold however ($p < .05$ non corrected), the contrasts provide a similar pattern of activations as the one obtained with the fixed effect analysis.

This fMRI study shows activation of Broca's region in the deixis conditions compared to the baseline. The Left Inferior Frontal Gyrus (LIFG) was therefore activated during verbal pointing at the agent of the action, through prosody or syntax. This activation is consistent with functional neuroimaging studies on complex syntactic processing (Caplan *et al.*, 2000; Friederici, 2002; Just *et al.*, 1996). These studies have shown the involvement of the LIFG in plausibility judgments about syntactically complex constructions, which require intricate tracking of thematic roles. Our findings are also in line with studies on the observation and mental imagery of action which show LIFG activation during action tracking (Binkofski *et al.*, 2000; Grafton *et al.*, 1996; Iacoboni *et al.*, 1999; Rizzolatti *et al.*, 2007). According to Rizzolatti and colleagues, the role of the LIFG in speech would have evolved from a 'basic mechanism originally not related to communication: the capacity to recognize actions' (Rizzolatti & Arbib, 1998). In addition, Dogil and colleagues' fMRI study on the production of prosodic features at the syllable and phrase levels has also revealed left IFG activation (Mayer *et al.*, 2002; Dogil *et al.*, 2002). Taken together, these observations support the claim that the role of the LIFG is that of an action-structure parser, which, in morphosyntactic encoding and decoding, handles the parsing of the predicate and its arguments, or the attentional monitoring of "who does what to whom".

The left insula was also found activated in both (deixis – baseline) contrasts. The involvement of the left precentral gyrus of the insula in articulatory planning during speech has already been shown (Dronkers, 1996). As described above, prosody has acoustic and articulatory correlates. The production of prosodic focus may require more accurate planning of the movements of the larynx, the tongue and the jaw, which could explain why the prosodic deixis condition yields significant activation of the left insula when compared with the baseline (same words to articulate, but a more stringent prosody). Similarly, the syntactic deixis condition (compared to the baseline) likely requires more accurate articulatory planning, given the larger number of consonant clusters involved (due to schwa deletion).

The activation of the LSMG and of Wernicke's area in the prosodic deixis condition alone suggests that, when deixis is already encoded by syntax, no additional recruitment of the inferior parietal lobule and Wernicke's area is necessary. As mentioned in section 2, several studies suggest that the inferior parietal regions in both hemispheres function as sensory integrators which form representations necessary in the organisation of motor actions, such as linguistic or non-linguistic pointing at targets. The left hemisphere would have a linguistic predominance.

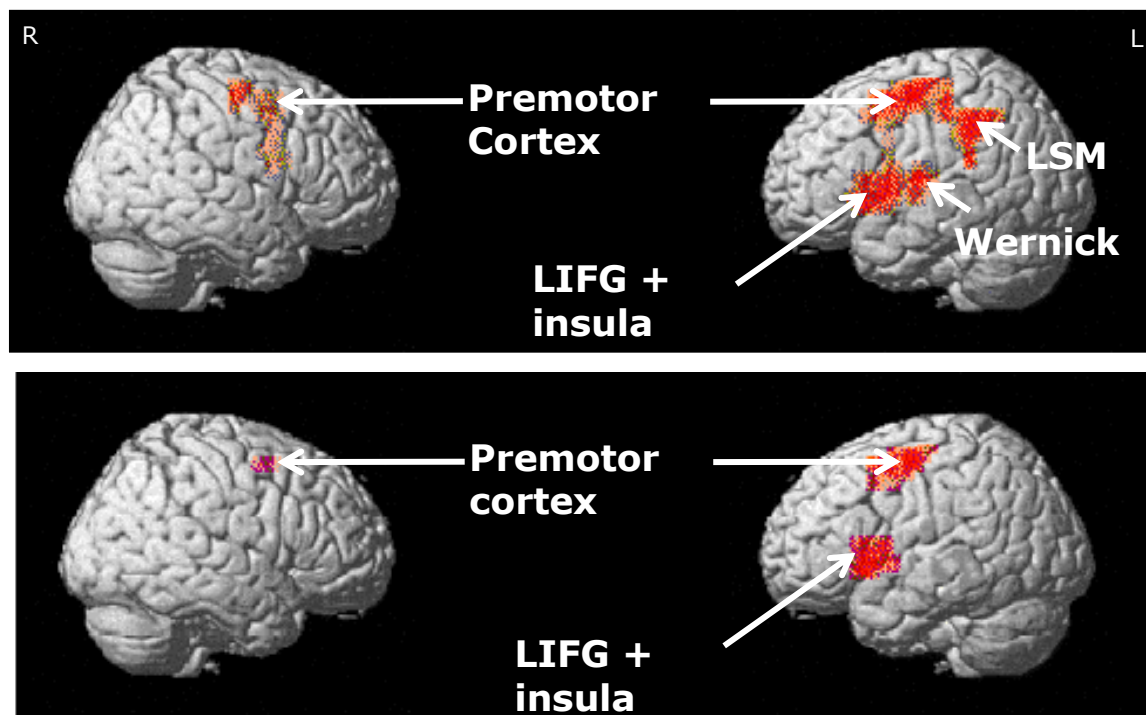


Figure 6: Activations in the prosodic deixis contrast (top) and syntactic deixis contrast (bottom). Both contrasts provide activations of frontal regions. Only the prosodic contrast shows temporal and parietal activations.

In speech, a left temporo-parieto-frontal network might be recruited in the organisation of verbal motor actions from auditory representations. Our results, with the activations of the LSMG, the LIFG and Wernicke's area in prosodic deixis, are in line with this hypothesis. Like visually-guided manual pointing, prosodic pointing may need multisensory representations to be formed via superior temporal and inferior parietal regions to organise articulation and phonation adequately. We therefore suggest that non-grammaticalized linguistic deixis recruits the temporo-parieto-frontal network and that grammaticalized deixis (syntactic deixis) is handled solely by the LIFG.

2.3.2. PERCEPTION FMRI STUDY

The production study described above suggests that associative brain areas are recruited for the production of vocal pointing. As demonstrated in section 2.2., the articulatory and phonatory patterns produced by speakers are also perceived by listeners/viewers. It can therefore be hypothesised that the cerebral network recruited for the production of vocal pointing is at least partly also recruited for its perception. This is what the study presented hereafter aims at testing.

Although, the acoustic and articulatory correlates of prosodic focus have been quite extensively studied (as recalled in section 2.), it remains unclear what neural processes underlie its perception. Meanwhile studies have shown that prosodic processing in general cannot be restricted to the right hemisphere (see Baum & Pell, 1999 for a review). Two studies have analysed the processing of prosodic focus (or closely related prosodic phenomena). The first one (Wildgruber *et al.*, 2004) aimed at contrasting affective vs linguistic prosody. The

linguistic prosodic task was an indirect informational focus detection task (find the most suitable answer to a specific question). For the linguistic prosodic task, the authors found bilateral activations of the primary and secondary auditory cortices, of the anterior insular cortex and of the frontal operculum (BA 6/44/47), right hemisphere activation of the dorso-lateral-frontal regions and left hemisphere activations of the inferior frontal cortex. The second study which examined the processing of prosodic focus (Tong *et al.*, 2005) aimed at differentiating the processing of ‘intonation’ (question/affirmation discrimination) and that of contrastive stress. It additionally compared English and Chinese. For the processing of contrastive stress, the authors put forward bilateral activation of the intra-parietal sulcus (BA 40/7), right hemisphere activation of the medial frontal gyrus (BA 9/46) and left hemisphere activations of the supramarginal gyrus and the posterior medio-temporal gyrus (BA 21/20/37).

Moreover, as we have shown above (in section 2.2.), even though the perception of prosodic focus was often considered as uniquely auditory, it is possible to perceive prosodic focus visually and the visual modality can enhance perception when prosodic auditory cues are degraded (Dohen & Lœvenbruck, under revision). This finding emphasises the necessity to consider the perception of prosodic contrastive focus and speech prosody in general as multimodal. The perception fMRI study presented here aims at analysing the neural processing of prosodic focus from a multimodal point of view.

fMRI recordings were conducted for 12 native speakers of French at the ATR Brain Activity Imaging Center (Japan). Subjects were scanned while they were performing a prosodic focus detection task for three modalities (audio only A, visual only V and audiovisual AV). The stimuli were subject-verb-object (SVO) structured sentences uttered in both normal and whispered speech. In some cases, S was under prosodic contrastive focus. The speaker was a female native speaker of French. After seeing/hearing/seeing and hearing each stimulus, subjects were asked to tell whether they had perceived a correction (i.e. contrastive focus) or not. A detailed description of the stimuli and of the fMRI procedure (paradigm and data collection and analysis) can be found in Appendix A.4.

Behavioral results

Table 2 provides the percentages of correct answers for all conditions. It appears that subjects performed the task correctly: they were able to identify focus cases from non-focus cases (the percentages of correct answers were well above chance, in all conditions).

Table 2: Mean percentages of correct answers and standard deviations (sd) across all subjects for each modality (chance level: 50%).

modality	normal speech		whispered speech	
	% correct	sd	% correct	sd
AV	98.4	2.0	89.6	6.5
A	97.4	3.0	69.9	12.8
V	86.4	7.5	88	8.6

Preliminary fMRI results

The analysis of the fMRI data is still underway and the results presented here are only preliminary. A preliminary analysis was conducted for the focus vs no focus contrast for the auditory and auditory-visual modalities. It appeared that auditory alone (A) detection of prosodic focus involved the right associative auditory cortex and fusiform gyrus (BA 19) as well as the left middle frontal gyrus (BA 6/46) and inferior temporal gyrus (BA 37) and the cerebellum bilaterally. For the auditory-visual modality, we found bilateral activations of the middle and inferior frontal gyri (BA 40), the middle temporal gyrus (BA 21), the inferior parietal lobule (BA 40) and the fusiform gyrus (BA 37) as well as left activation of the supramarginal gyrus (BA 40).

It appears that for all modalities, prosodic focus detection or processing involves bilateral activations of associative brain areas. Auditory perception of prosodic focus (vs no focus) appears to be essentially processed in associative areas right superior temporal gyrus and left inferior temporal gyrus (BA 37). Multimodal (AV) perception of prosodic focus involves bilateral activations of temporal and parietal associative areas as well as inferior and middle frontal regions. This illustrates the underlying necessity of associating various types of information to detect focus (especially auditory and articulatory) and supports the assumption that the articulatory and phonatory patterns produced by the speaker are integrated in perception.

Similar activations of a complex neural network in the processing of focus have been observed in two recent ERP studies (Bornkessel *et al.*, 2003; Magne *et al.*, 2005). The implication of the left parietal lobe in the auditory-visual perception of prosodic pointing is interesting, in the light of our fMRI study of the production of prosodic pointing, and of the studies on manual and ocular pointing.

3. MULTIMODAL POINTING

The aim of this multimodal fMRI study was to investigate the cerebral correlates of pointing in several modalities (Lœvenbruck *et al.*, 2007). To test the hypothesis that multimodal pointing could involve a continuum of cerebral regions, we designed an fMRI paradigm including four conditions: 1) index finger pointing, 2) eye pointing, 3) prosodic focus, 4) syntactic extraction.

Subjects and material

Fifteen healthy right-handed volunteers, aged 18-55 years, native speakers of French were examined. Stimuli consisted of two types (1 and 2) of images consisting of a girl (Lise) and a boy (Jules) next to each other and alternatively located on the right and on the left side of the screen. Type 1 images showed the girl holding a book, while the boy did not; Type 2 images showed the reverse. In the middle of all images, a fixation cross was displayed. A blank screen with a mid-centered black cross preceded each stimulus. The tasks consisted of verbal and non-verbal pointing to a character. Control conditions were included. The same question:

“Est-ce que Lise tient le livre?” (Does Lise hold the book?) was used for all tasks. Subjects were instructed: (a) to confirm the question when type 1 images were presented (control), (b) to point to Jules when type 2 images were presented (contrastive pointing). The tasks were the following: During *prosodic pointing*, subjects uttered «JULES tient le livre» (JULES holds the book), with contrastive focus on “Jules”. During *syntactic pointing*, they uttered «C’est Jules qui tient le livre» (It is Jules who holds the book), with syntactic extraction of “Jules”. During *digital pointing*, they pointed with the right index finger to Jules. During *ocular pointing*, subjects looked in the direction of Jules.

The controls were the following: In prosodic and syntactic controls, subjects neutrally uttered: «Lise tient le livre » (Lise holds the book). In digital control, they performed a downward finger movement. In ocular control, they made a downward eye movement. The detailed methods are presented in Appendix A.5.

Preliminary fMRI results

Patterns of activation were examined for 15 subjects for each of the following contrasts:

- digital pointing tasks vs digital control
- ocular pointing tasks vs ocular control
- prosodic pointing tasks vs prosodic control
- syntactic pointing tasks vs syntactic control

Preliminary results of the random effect group analysis are reported. Figure 7 represents functional activations during each pointing condition vs its control.

Digital pointing (vs its control) yielded bilateral activation of the superior parietal lobule (BA 7), with a left hemisphere dominance. There was bilateral activation of the supramarginal gyrus (BA 40), with a right hemisphere dominance. There was left frontal cortex activation (BA 4, 6). *The ocular condition* activated the left hemisphere predominantly in the post central gyrus (BA 3) and the frontal lobe bilaterally (BA 4, 6). The occipital gyrus (BA 17, 18) was activated bilaterally. *The prosodic condition* recruited the superior parietal lobule (BA 7) bilaterally, with a left dominance. The supramarginal gyrus activation did not reach significance, contrary to our previous finding on covert prosody (see section 2.3.1.). The left postcentral gyrus was activated (BA 2, 3). Perisylvian regions (BA 47, 13, 42) were activated bilaterally. *The syntactic condition* activated the right supramarginal gyrus (BA 40) and did not recruit the left parietal lobule. Right perisylvian regions (BA 44, 47, 21) were activated, as well as the premotor cortex (BA 6) bilaterally.

To summarise, this preliminary analysis revealed that the left superior parietal lobule (BA 7 or its neighbour BA 3) was activated in all three digital, ocular and prosodic pointing but not in syntactic pointing. These results indicate that pointing in different modalities may recruit the left superior parietal lobule, ocular pointing being more anterior than prosodic pointing, itself more anterior than digital pointing.

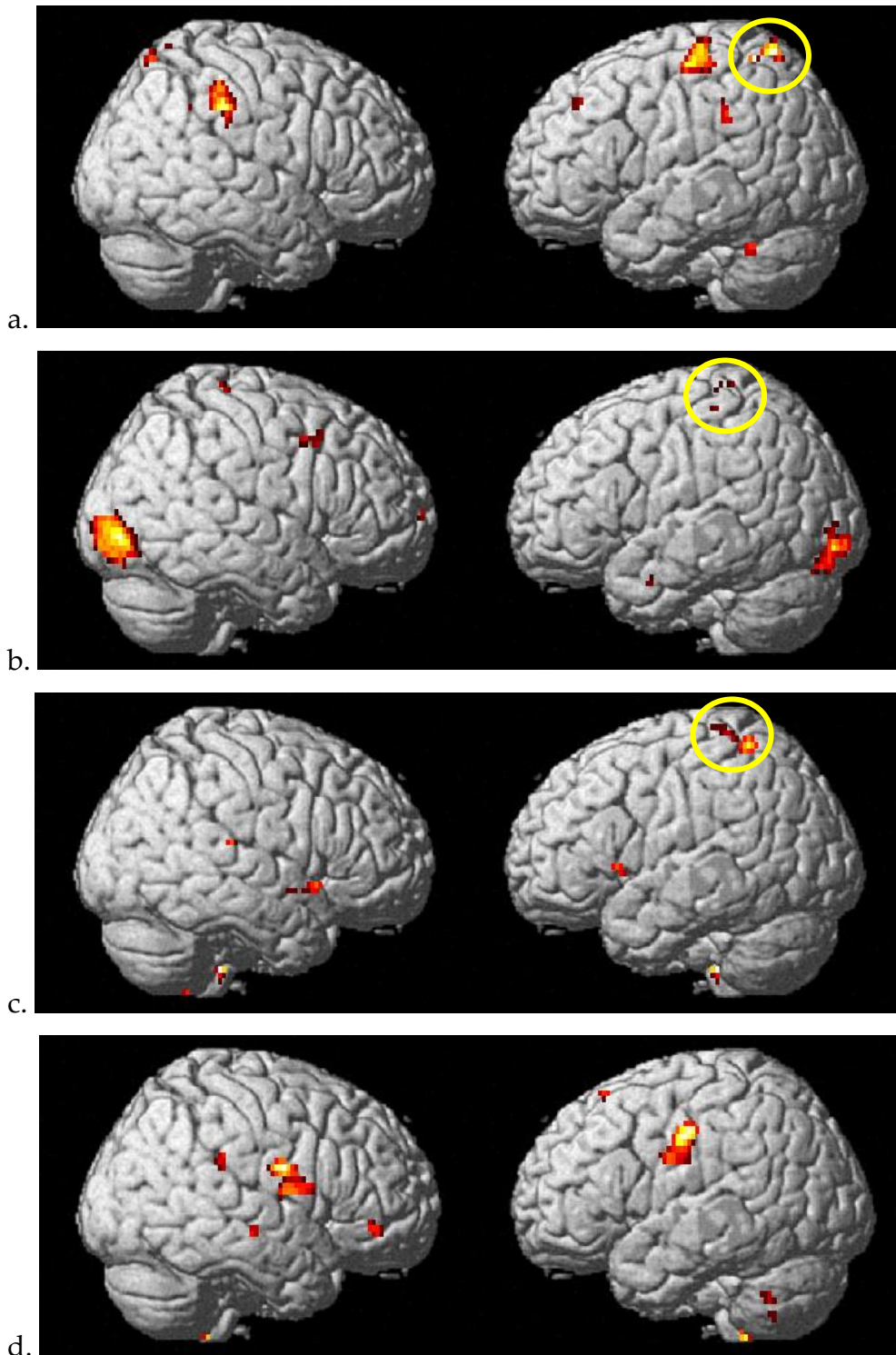


Figure 7: Activations during each pointing condition vs its control, rendered on a sagittal template, in a. digital- b. ocular- c. prosodic- d. syntactic-modes. The left hemisphere is on the right, the right one on the left

These results are in accordance with the results from our behavioural studies which suggest that speakers may use multisensory representations in order to produce prosodic pointing adequately, just like they do to produce a manual or ocular gesture. These representations may be formed via the activation of associative parietal areas. The lack of parietal activation

in syntactic pointing could be due to the absence of on-line multisensory construction in syntactic pointing, compared with prosodic pointing. Syntactic pointing uses a grammaticalized “frozen” construction and may not need integrated representations to be formed. This fMRI study therefore suggests that the left superior parietal lobule could be a potential substrate for pointing with the finger, the eye, and the larynx.

4. CONCLUSION

We have shown that pointing is a critical device to explore the potential common neural networks at play in gesture and language. The role of manual (or digital) pointing in language acquisition strongly suggests indeed that vocal pointing and pointing in other modalities may well be grounded in a common cerebral network.

We have argued from several neuroimaging studies on manual and ocular pointing that these two modalities seem to recruit a network including the left posterior parietal and frontal cortices.

Behavioural studies in our laboratory have shown that prosodic vocal pointing is a sophisticated behaviour that requires accurate laryngeal and articulatory control. Perception studies have revealed that the potential articulatory visible cues to prosodic vocal pointing are in fact visually perceived. In other words, viewers can decipher these sophisticated signals. We have claimed that speakers may need multisensory representations to produce prosodic pointing adequately and that these representations may involve the activation of associative parietal regions, close to the regions involved in manual and ocular pointings.

The results of a first study on covert vocal pointing, including prosodic pointing (i.e. focus) and syntactic pointing (i.e. syntactic extraction) have been presented. It was shown that covert prosodic pointing does recruit left parietal regions, whereas syntactic pointing mainly involves frontal region. The involvement of the left parietal lobe was confirmed by a recent fMRI study of the perception of prosodic pointing. Finally we presented the preliminary results of a new study of multimodal pointing (digital, ocular, overt prosodic and syntactic pointing). They seem to reveal a common pattern of left parietal activation in ocular, digital and prosodic pointing. A grammaticalization process has been suggested to explain the lack of parietal activation in syntactic pointing. Altogether these fMRI studies are in line with our conjecture that linguistic on-line pointing (prosodic focus) is grounded in the same cerebral network as gestural (manual and ocular) pointings. More analyses are underway to consolidate these results. If they come out as robust, then they will contribute to the debate on the gestural origin of language.

ACKNOWLEDGEMENTS

Many researchers were involved in the different works presented here. In alphabetical order, they are C. Abry, M. Baciú, A. Callan, D. E. Callan, F. Carota, M.-A. Cathiard, H. Hill, L. Lamalle, P. Perrier, C. Pichat, C. Savariaux, J.-L. Schwartz, C. Segebarth. We thank them all sincerely.

REFERENCES

- Astafiev S. V., Shulman G. L., Stanley C. M., Snyder A. Z., Van Essen D. C. & 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.
- Astésano C., Magne C., Morel M., Coquillon A., Espesser R., Besson M. & Lacheret A. (2004). Marquage acoustique du focus contrastif non codé syntaxiquement en français. *Proceedings of the XXVIèmes Journées d'Etudes sur la Parole*, 126–129. Fez, Morocco.
- Astésano C., Bard E. G. & Turk A.. (2007). Structural influence on initial accent placement in French. *Language and Speech*, 50 (3), 423-446.
- Audouy, M. (2000). *Traitement d'images vidéo pour la capture des mouvements labiaux*. Final engineering report, Institut National Polytechnique de Grenoble.
- Bates, E., Camaioni, L. & Volterra V. (1975). The acquisition of performatives prior to speech. *Merrill-Palmer Quarterly*. 21: 205-226.
- Bates E., Benigni L., Bretherton I., Camaioni L., Volterra V. (1979). *The emergence of symbols: cognition and communication in infancy*. New York: Academic Press.
- Baum, S. R. & Pell M. D. (1999). The neural bases of prosody: Insights from lesion studies and neuroimaging. *Aphasiology*, 13(8), 581-608.
- Bellugi U., Poizner H. & Klima E. S. (1989). Language, modality and the brain. *Trends in Neurosciences*, 12 (10), 380-388.
- Benoît C., Mohamadi T. & Kandel S. (1994). Effects of Phonetic Context on Audio-Visual Intelligibility of French. *Journal of Speech and Hearing Research*, 37, 1195-1203.
- Berthoud A.-C. (1990). Deixis, thématization et détermination. *La deixis. Colloque en Sorbonne*. M.-A. Morel & L. Danon-Boileau (dir.), PUF, 527-542.
- Binkofski F., Amunts K., Stephan K. M., Posse S., Schormann T., Freund H.-J., Zilles K. & Seitz R. (2000). Broca's region subserves imagery of motion: a combined cytoarchitectonic and fMRI study. *Hum. Brain Mapp.*, 11, 273-285.
- Bornkessel I., Schlesewsky M., & Friederici A. D. (2003). Contextual information modulates initial processes of syntactic integration: The role of inter- versus intrasentential predictions. *Journal of Experimental Psychology: Learning Memory and Cognition*, 29, 269-298.
- Brown H. D. (1971). Children's comprehension of relativized English sentences. *Child Development*, 42, 1923–36.
- Bruner, J. (1975). The ontogenesis of speech acts. *Journal of Child Language*, 2, 1– 19.
- Butcher C. & 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.
- Butterworth G. (2003). Pointing is the royal road to language. *Pointing, where language, culture, and cognition meet*. Kita S. (ed.). Lawrence Erlbaum Associates Publishers, Mahwah, NJ, 9-34.
- Capirci O., Iverson J. M., Pizzuto E., & Volterra V. (1996). Gestures and words during the transition to two-word speech. *Journal of Child language*, 23, 645- 673.

- Caplan D., Alpert N., Waters G. & Olivieri A. (2000). Activation of Broca's area by syntactic processing under conditions of concurrent articulation. *Hum. Brain Mapp.*, 9, 65-71.
- Caselli M.C. (1990). Communicative gestures and first words. In V. Volterra & C.J. Erting (Eds.), *From gesture to language in hearing and deaf children*. New York: Springer-Verlag, pp. 56-67.
- Cathiard M.-A. (1994). *La perception visuelle de l'anticipation des gestes vocaliques : cohérence des événements audibles et visibles dans le flux de la parole*, Unpublished doctoral dissertation, Université Pierre Mendès France, Grenoble.
- Cavé, C.; Guaitella, I.; Bertrand, R.; Santi, S.; Harlay, F.; Espesser, R. (1996). About the Relationship between Eyebrow movements and F0 Variation. *Proceedings of ICSLP 96*, 4, 2175-2179.
- Cerrato, L.; Skhiri, M. (2003). Analysis and measurement of communicative gestures in human dialogues. *Proceedings of AVSP 2003*, 251-256.
- Cheney D. L. & Seyfarth R. M. (1981). Selective forces affecting the predator alarm calls of vervet monkeys. *Behaviour*, 76, 25-61.
- Chomsky N. (1957). *Syntactic Structures*. The Hague: Mouton.
- Chomsky N. (1966). *Cartesian linguistics*. New York: Harper Row.
- Connaghan K. P., Moore C. A., Reilly K. J., Almand K. B., & Steeve R. W. (2001). Acoustic and physiologic correlates of stress production across systems, Poster presented at the *American Speech-Language-Hearing Association*, New Orleans, L. A.
- Connolly J. D., Goodale M. A., Desouza J. F., Menon R. S., Vilis T. (2000). A comparison of frontoparietal fMRI activation during anti-saccades and antipointing. *J Neurophysiol.*, 84, 1645-1655.
- Corballis, M. C. (1991). *The lopsided ape: Evolution of the generative mind*. New York: Oxford University Press; 1991.
- Dahan D. & Bernard J.-M. (1996). Interspeaker Variability in Emphatic Accent Production in French. *Language and Speech*, 39(4), 341-374.
- Delais-Roussarie E., Rialland A., Doetjes J. & Marandin J.-M. (2002). The Prosody of Post Focus Sequences in French. In *Proceedings of Speech Prosody 2002*, Aix-en-Provence, France, 239-242.
- Di Cristo A. (1998). Intonation in French. In D. Hirst; A. Di Cristo (Eds.), *Intonation Systems: a Survey of Twenty Languages*, 195-218. Cambridge University Press.
- Di Cristo A. & Jankowski L. (1999). Prosodic Organisation and Phrasing after Focus in French. *Proceedings of ICPHS 1999*, San Francisco, USA, 1565-1568.
- Diessel H. & Tomasello M. (2000). The development of relative clauses in spontaneous child speech. *Cognitive Linguistics*, 11, 131-151.
- Dogil G., Ackermann H., Grodd W., Haider H., Kamp H., Mayer J., Riecker A., Wildgruber D. (2002). The Speaking Brain. *Journal of Neurolinguistics*, 15 (1), 59-90.
- Dohen M., Løevenbruck H., Cathiard M.-A. & Schwartz J.-L., 2004. Visual perception of contrastive focus in reiterant French speech. *Speech Communication*, 44, 155-172.
- Dohen M. & Løevenbruck H. (2004). Pre-focal Rephrasing, Focal Enhancement and Post-focal Deaccentuation in French. In *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., (2005). Audiovisual Production and Perception of Contrastive Focus in French: a multispeaker study. *Proceedings of Interspeech 2005*, 2413-2416.
- 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.
- Dohen M. & Løevenbruck H. (under revision). Interaction of audition and vision for the perception of prosodic contrastive focus. *Language and Speech*.
- Dronkers N. F. (1996). A new brain region for coordinating speech articulation. *Nature*, 384, 159-161.
- Edwards M. G. & Humphreys G. W. (1999). Pointing and grasping in unilateral visual neglect: effect of on-line visual feedback in grasping. *Neuropsychologia*, 37 (8), 959-973.
- Erickson D. (submitted). Effects of contrastive emphasis on jaw opening.
- Fillmore C. J. (1997). *Lectures on deixis*. Stanford: Center for the Study of Language and Information.
- Fougeron C. & Keating P. A. (1997) Articulatory strengthening at edges of prosodic domains. *Journal of the Acoustic Society of America*, 101, 3728-3740.
- Friederici A. D. (2002). Towards a neural basis of auditory sentence processing. *Trends in Cognitive Sciences*, 6 (2), 78-84.
- Goffman L. & Malin C. (1999). Metrical Effects on Speech Movements in Children and Adults. *Journal of Speech, Language, and Hearing Research*, 42 (4), 1003-1015.
- Goldin-Meadow S. & Butcher C. (2003). Pointing toward two-word speech in young children. In *Pointing: Where language, culture, and cognition meet* S. Kita (Ed.), 85-107. Mahwah, NJ: Earlbaum Associates.
- Grafton S. T., Arbib M. A., Fadiga L. & Rizzolatti G. (1996). Localization of grasp representations in humans by positron emission tomography. 2. Observation compared with imagination. *Exp. Brain Res.*, 112 (1), 103-111.
- Graf, H.P.; Cosatto, E.; Strom, V.; Huang, F.J. (2002). Visual Prosody: Facial Movements Accompanying Speech. *Proceedings of the 5th IEEE International Conference on Automatic Face and Gesture Recognition (FGR'02)*, 381-386.
- Granström, B.; House, D. (2005). Audiovisual representation of prosody in expressive speech communication. *Speech Communication*, 46, 473-484.
- Hadar U., Steiner T. J., Grant E. C., Rose F. C. (1983). Head movement correlates of juncture and stress at sentence level. *Language and Speech*, 26, 117-129.
- Hagler D. J. Jr, Riecke L., Sereno M. I. (2007). Parietal and superior frontal visuospatial maps activated by pointing and saccades. *NeuroImage*, 35 (4), 1562-1577.
- Hewes G. W. (1981). Pointing and language. In T. Myers, J. Laver & J. Anderson (eds.). *The cognitive representation of speech*, 263-269. Amsterdam: North-Holland.
- Hickok G. & Poeppel D. (2001). Towards a functional neuroanatomy of speech perception. *Trends In Cognitive Sciences*, 4 (4), 131-138.
- Higginbotham, D. (1992). Reference and control. In: R.K. Larson, S. Iatridou, U. Lahiri & J. Higginbotham (eds.). *Control and grammar*. Dordrecht: Kluwer, 79-108.

- Hornby P. A. & Hass W. A. (1970). Use of contrastive stress by preschool children. *Journal of Speech and Hearing Research*, 13, 395-399.
- Iacoboni M., Woods R. P., Brass M., Bekkering H., Mazziotta J. C. & Rizzolatti G. (1999). Cortical mechanisms of human imitation. *Science*, 286, 2526-2528.
- Iacoboni M., Koski L. M., Brass M., Bekkering H., Woods R. P., Dubeau M.-C., Mazziotta J. C. & Rizzolatti G. (2001). Reafferent copies of imitated actions in the right superior temporal cortex. *Proc. Natl. Acad. Sci. USA*, 98 (24), 13995-13999.
- Jackendoff R. (2002). *Foundations of Language - Brain, Meaning, Grammar, Evolution*. Oxford University Press.
- Jisa H. & Kern S. (1998). Relative clauses in French children's narrative texts. *Journal of Child Language*, 25, 623-652.
- de Jong K., Beckman M. E. & Fletcher J. (1991). The articulatory kinematics of final lengthening. *J. Acoust. Soc. Am.*, 89, 369-382.
- Jun S.-A. & Fougeron C. (2000). A Phonological Model of French Intonation. In A. Botinis (Ed.), *Intonation: Analysis, modelling and technology*, 209-242. Dordrecht: Kluwer Academic Publishers.
- Just M. A., Carpenter P. A., Keller T. A., Eddy W. F. & Thulborn K. R. (1996). Brain activation modulated by sentence comprehension. *Science*, 274, 114-116.
- Kendon A. (1996). An Agenda for Gesture Studies. *Semiotic Review of Books*, 7(3), 8-12.
- Kendon A. (2004). *Gesture: Visible Action as Utterance*. Cambridge: Cambridge University Press.
- Kertzman C., Schwarz U., Zeffiro T. A. & 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. & 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.
- Kita S. (2003). *Pointing: where language, culture, and cognition meet*. Lawrence Erlbaum Associates Publishers, Mahwah, NJ.
- Klima E. S. & Bellugi U. (1988). *The signs of language*. Harvard University Press.
- Konopscynsky G. (1986). *Du prélangage au langage : acquisition de la structuration prosodique*. PhD Thesis, University of Strasbourg.
- Krahmer, E.; Ruttkay, Z.; Swerts, M.; Wesselink, W. (2002). Pitch, Eyebrows and the Perception of Focus. *Proceedings of Speech Prosody 2002*, 443-446.
- Lacquaniti F., Perani D., Guigon E., Bettinardi V., Carrozzo M., Grassi F., Rossetti Y. & Fazio F. (1997). Visuomotor transformations for reaching to memorized targets: a PET study. *Neuroimage*, 5 (2), 129-146.
- Leavens D. A., Hopkins W. D. & Bard K. A. (2005). Understanding the Point of Chimpanzee Pointing. Epigenesis and Ecological Validity. *Current Directions in Psychological Science*, 14, 185-189.
- Liberman A. M. (1982). On Finding that Speech Is Special. *American Psychologist*, 37, 2, 148-167.
- Lœvenbruck H. (1996). *Pistes pour le contrôle d'un robot parlant capable de réduction vocalique*. Unpublished Doctoral Dissertation in Cognitive Sciences, INPG, Grenoble.

- 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. & 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.
- Lœvenbruck H., Vilain C., Carota F., Baciú M., Abry C., Lamalle L., Pichat C. & Segebarth C. (2007). Cerebral correlates of multimodal pointing: An fmri study of prosodic focus, syntactic extraction, digital- and ocular- pointing. *Proceedings of the XVIth International Congress of Phonetic Sciences*, 6-10 August 2007, Saarbruecken.
- Magne C., Astesano C., Lacheret-Dujour A, Morel M., Alter K., Besson M. (2005). On-Line processing of "pop-out" words in Spoken French dialogues. *Journal of Cognitive Neurosciences*, 17(5), 740-756.
- Macwhinney B., & Ple'h C. (1988). The processing of restrictive relative clauses in Hungarian. *Cognition*, 29, 95–141.
- Mayer J., Wildgruber D., Riecker A., Dogil G., Ackermann H. & Grodd W. (2002). Prosody production and comprehension: converging evidence from fMRI studies. *Proceedings of Speech Prosody 2002*, Aix-en-Provence, France, 11-13 April 2002, 487-490.
- McNeill, D. (1992). *Hand and Mind*. Chicago: University of Chicago Press.
- Mckee C., Mcdaniel D. & Snedeker J. (1998). Relatives children say. *Journal of Psycholinguistic Research*, 27, 573–96.
- Ménard L., Lœvenbruck H. & Savariaux C. (2006). Articulatory and acoustic correlates of contrastive focus in French: a developmental study, in Harrington, J. & Tabain, M. (eds), *Speech Production: Models, Phonetic Processes and Techniques*, Psychology Press : New York, 227-251.
- Morford M. & Goldin-Meadow S. (1992). Comprehension and production of gesture in combination with speech in one-word speakers. *Journal of Child Language*, 19 (3), 559-580.
- Munhall K.G., Jones J. A., Callan D. E., Kuratate T., Vatikiotis-Bateson E. (2004). Visual Prosody and Speech Intelligibility – Head Movement Improves Auditory Speech Perception. *Psychological Science*, 15(2), 133-137.
- Nelson, W.L. (1983). Physical principles for economies of skilled movements. *Biological Cybernetics*, 46(2), 135-147.
- Oldfield R.C. (1971). The assessment and analysis of handedness: the Edinburgh Inventory. *Neuropsychologia*, 9(97), 113.
- Pollick A. S. & de Waal F. B. M. (2007). Ape gestures and language evolution. *P. N. A. S.*, 104, 19,8184-8189.
- Povinelli D. J. & 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.
- Povinelli D.J., Bering J. M., & Giambrone S. (2003). Chimpanzee 'pointing': Another error of the argument by analogy? In *Pointing: Where language culture and cognition meet*. S. Kita (Ed.), 35-68, Lawrence Erlbaum Associates.

- Rizzolatti G., Fogassi L., Gallese V. (1997). Parietal cortex: from sight to action. *Curr. Opin. Neurobiol.*, 7 (4), 562-567.
- Rolfe L. (1996). Theoretical stages in the prehistory of grammar. In A. Lock & C. R. Peters (eds.), *Handbook of human symbolic evolution*, 776-792. Oxford: Clarendon.
- Rossi M. (1999). La focalisation. In *L'intonation, le système du français: description et modélisation*, Chap. II-6, 116-128. Ophrys.
- Séguinot, A. (1976). L'accent d'insistance en français standard. *Studia Phonetica*, 12, 1-58.
- 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.
- Simon O., Mangin J. F., Cohen L., Le Bihan D. & Dehaene S. (2002). Topographical layout of hand, eye, calculation, and language-related areas in the human parietal lobe. *Neuron*, 33, 475-487.
- Tabain M. (2003). Effects of prosodic boundary on /aC/ sequences: Articulatory results. *J. Acoust. Soc. Am.*, 113 (5), 2834-2849.
- Tavakolian S. (1981). The conjoined clause analysis of relative clauses. *Language acquisition and linguistic theory*, S. Tavakolian (ed.), 167-87. Cambridge, MA: MIT Press.
- Tomasello, M., Carpenter, M., & Liszkowski, U. (2007). A new look at infant pointing. *Journal of Child Language*, 34, 1-20.
- Tong Y., Gandour J., Talavage T., Wong D., Dziedzic M., Xu Y., Li X. & Lowe M. (2005). Neural circuitry underlying sentence-level linguistic prosody. *NeuroImage*, 28, 417-428.
- Touati P. (1987). Structures prosodiques du suédois et du français. *Lund Working Papers* 21, Lund University Press.
- de la Tour G. (1635). *Le Tricheur à l'as de carreau* (The cheat with the ace of diamonds), Musée du Louvre, Paris.
- Vallar G., Guariglia C., Nico D. & Pizzamiglio L. (1997). Motor deficits and optokinetic stimulation in patients with left hemineglect. *Neurology*, 49 (5), 1364-1370.
- Volterra V., Caselli M. C., Capirci O. & Pizzuto E. (2005). Gesture and the emergence and development of language. In *Beyond Nature-Nurture. Essays in Honor of Elizabeth Bates*. M. Tomasello & D. Slobin (eds.), Mahwah, NJ: Lawrence Erlbaum Associates. 3-40.
- Wechsler, S. & Wayan, A. (1998). Syntactic ergativity in Balinese: an argument structure based theory. *Natural Language and Linguistic Theory* 16, 387-441.
- Wildgruber D., Hertrich I., Riecker A., Erb M., Anders S., Grodd W. & Ackerman H. (2004). Distinct Frontal Regions Subserve Evaluation of Linguistic and Emotional Aspects of Speech Intonation. *Cerebral Cortex*, 14(12), 1384-1389.
- Wilkins D. (2003). Why pointing with the index finger is not a universal (in sociocultural and semiotic terms). *Pointing: where language, culture, and cognition meet*. Kita S. (ed.). Lawrence Erlbaum Associates Publishers, Mahwah, NJ, 171-216.
- Wurmbrand, S. (2001). *Infinitives: restructuring and clause structure*. Berlin: Mouton de Gruyter.

APPENDICES

A.1. Methods for the EMA Study

Recordings

Simultaneous acoustic and articulatory recordings were collected using EMA (Carstens AG100), for a female native speaker of French. Five pellets were glued midsagittally to the apex, middle and dorsum of the tongue, and to the lower and upper incisors. The EMA data were sampled at 500 Hz, the acoustic data at 16000 Hz. The data were corrected for rotation, low-pass filtered and normalized by the reference pellet to correct for head movements. Velocity and acceleration traces were obtained using a finite difference method.

Corpus

The corpus consisted of several read sentences, containing 4-syllable target words, which constituted an Accentual Phrase. Position of the target word in the sentence was varied (initial, central, final), as in the example below for which the target word is “l’illuminée” (the crank):

‘L’illuminée a allumé néanmoins le monument.’ (Initial).

The crank has nevertheless lighted the monument.

‘Il a humilié l’illuminée en l’éloignant.’ (Central).

He humiliated the crank by moving her away.

‘L’aumonier a néanmoins éloigné l’illuminée.’ (Final).

The chaplain has nevertheless moved the crank away.

The corpus was designed so that analyses were made easier. To facilitate F0 tracking, only sonorants were used, and to obtain clear tongue movements, syllables containing vowels /i/ and /a/, which correspond to large tongue displacements in the horizontal and vertical dimensions, were selected. Several target words were pronounced. Results are presented for the target word: “l’anomala” (/la-no-ma-la/, beetle), which entails clear tongue movements on the initial and final syllables. As described in Jun & Fougeron (2000), in the contrastive focus rendition, initial and final syllables are potential spots for the F0 peak to occur.

The sentences were pronounced under two conditions: a neutral (broad focus) condition and a contrastive focus condition. Three speaking rates were produced: slow, normal, fast. Contrastive emphasis or focus was elicited on the target word as follows. Before each recording, a sentence was played to the subject, where the target word had been replaced by a wrong 4-syllable word. The subjects had to correct the sentence, placing contrastive focus on the target word.

Articulatory data analysis

Among the 8 traces (horizontal and vertical positions of the tongue -apex, -middle and -dorsum and of the lower incisor), the vertical position of the tongue-middle showed the most variation. It was thus chosen as the representative articulator. For each syllable, the tongue-middle vertical position at the time the vowel was fully reached, the peak velocity of the movement from the consonant to the vowel, and the duration of the syllable were measured, using hand-labelled events. The spectrogram of the acoustic signal, as well as velocity and acceleration traces, were used to mark these events. As shown in figure 2 for [la-no-ma-la], the beginning and end of the syllable were marked using the spectrogram (and listening to the

signal) and provided the duration of the syllable. The tongue-middle minimum vertical position for /a/ was measured using zero-crossing of the velocity trace. The peak velocity was measured using zero-crossing of the acceleration trace.

A.2. Methods for the Optotrak Study

For more details the reader can refer to Dohen *et al.* (2006).

Corpus

The corpus used for this study consisted of 13 subject-verb-object (SVO) structured sentences with CV syllables and, whenever possible, sonorants. Below is an example of one of the sentences used:

Lou ramena Manu. Lou gave a lift back to Manu.

Recordings

Four focus conditions were elicited: subject-, verb- and object-focus (narrow focus) and a neutral version (broad focus). In order to trigger contrastive focus, the speakers had to perform a correction task. They heard a prompt in which two speakers were talking and they were then asked to correct a phrase (S, V or O) which had been mispronounced. Two repetitions of each utterance were recorded.

The Optotrak system consists of three infrared (IR) cameras used to track the motion of infrared emitting diodes (IREDs) which in this case were glued to the speaker's face. The 3D coordinates of each IRED were automatically detected. For this experiment, we used two Optotraks in order to compensate for missing data. A total of 24 IREDs were glued to the speakers' faces. An additional 4 IREDs were attached to a head rig and were used to correct for head motion. IRED positions were sampled at 60 Hz and low-pass filtered. The acoustic signals were recorded simultaneously and sampled at 22 kHz.

Acoustic validation

The data recorded were first acoustically validated i.e. it was checked whether focus had actually been produced acoustically. For all the speakers, we checked that the focused utterances displayed a typical focused intonation and that focus was well perceived auditorily (informal auditory perception tests).

Articulatory measurements

In our previous studies (see *a.*), two articulatory features had been analyzed, namely inter-lip area and protrusion. These parameters best represented the high segmental variability of speech and were the most relevant parameters to isolate supra-segmental originating variations. However, it is not possible to compute inter-lip area accurately from Optotrak data. This is why, in this study, we separately analyzed lip opening (difference between the vertical coordinates of the upper and lower middle lip markers) and lip spreading (difference between the horizontal coordinates of the two lip corner markers). Jaw vertical movements were analyzed using the chin marker. Upper lip protrusion was computed as well.

Facial movements

Based on other studies of the facial movements accompanying speech and more specifically prosody, we decided to limit our study to the head and eyebrow movements. Cavé *et al.* (1996) showed that eyebrow movements accompanying prosody were mainly raising movements. Therefore we decided to study the raising of both the left and the right eyebrows (middle eyebrow markers). As for head movements, the three rotations and translations were available. Since Munhall *et al.* (2004) and Graf *et al.* (2002) had found that the main movements related to prosody were nods, we analyzed the rotation of the rigid body around the horizontal axis.

Data shaping

In order to be able to isolate and compare supra-segmental variations for different segmental constituents, we used a normalization technique. After normalization, a value of 1 corresponds to no variation of the considered parameter compared to the neutral version, a value above 1 corresponds to an increase and a value below 1 to a decrease. For both articulatory and facial movement parameters, we analyzed the inter- and intra-utterance contrasts related to focus (inter: comparison of a phrase in its focused and neutral versions; intra: comparison of a focused phrase with the other phrases of the same utterance).

A.3. Methods for the fMRI production Study

Stimuli

Each sentence was presented for 3 seconds at the beginning of the corresponding condition. Then a fixation mark, alternating every 3 seconds between a '+' and a 'x' sign, appeared in the middle of the screen. This alternation aimed at triggering the silent (covert speech) repetition (14 times per condition) of the sentence presented.

Subjects' performance

Before entering into the magnet, the subjects were trained to execute the tasks, first in overt mode, then in covert mode. In addition, pre- and post-scan audio recordings were carried out to estimate the subjects' task performance during the fMRI scan. Subjects were prompted by exactly the same script as during the scans. They produced each of the sentences 4 times. For the post-scan recording, the instruction was to reproduce in overt speech the intonation patterns mentally produced during the scans. Overall, the subjects' performance, as measured by the audio recordings before and after the scans, indicated that their production varied neither in rhythm nor in intonation between recordings.

fMRI paradigm

Three functional scans were performed. A block paradigm was used. A scan comprised 8 epochs (two repetitions of each condition) of 42 seconds each. The order of presentation of the four conditions was varied across scans and across subjects. Functional MR imaging was performed on a 1.5T imager (Philips NT) with echo-planar (EPI) acquisition. Twenty-five adjacent, axial, slices (5mm thickness each) were imaged 10 times during each epoch. The imaging volume was oriented parallel to the bi-commissural plane. An EPI MR pulse sequence was used. The MR parameters were: TR = 3700ms, TE = 45ms, pulse angle = 90°, acquisition

matrix = 64x64, reconstruction matrix = 128x128, field-of-view = 256x256mm². Between the first and second functional scans, a high-resolution 3D anatomical scan was acquired.

fMRI data processing

Data analysis was performed using the SPM-99 software (Wellcome Department of Cognitive Neurology, London). First, motion correction was applied. All images within a functional scan were realigned by means of a rigid body transformation. Then, the anatomical volume was spatially normalized into a reference space using the Montreal Neurological Institute template. The normalization parameters were subsequently applied to the set of functional images. Finally, to conform to the assumption in SPM that the data are normally distributed, the functional images were spatially smoothed.

Contrasts between conditions were determined voxelwise using the General Linear Model. Statistical significance threshold for individual voxels was established at $p = 0.001$. Clusters of activated voxels were then identified, based on the intensity of the individual responses and the spatial extent of the clusters. Finally, a significance threshold of $p = 0.05$ (corrected for multiple comparisons) was applied for identification of the activated clusters.

A.4. Methods for the fMRI perception study

Subjects

Twelve healthy right handed volunteers (Edinburgh Inventory; Oldfield 1971), native speakers of French (4 women and 8 men), with normal or corrected-to-normal vision and normal neurological history were examined. The study was approved by the ATR Human Subject Review and subjects gave informed written consent for experimental procedures.

Stimuli

The stimuli consisted of sentences spoken by a native French speaker (women). 24 were pronounced in normal speech and 6 in whispered speech. The structure of the sentences was: subject (two-syllable first name) - two-syllable action verb - object (two-syllable noun preceded by a one-syllable singular determiner). All syllables were CV syllables (sonorant consonants). An example of one of the sentences used is given below:

Nico mangea le bonbon. *Nico ate the candy.*

The audiovisual stimuli were recorded at ex-ICP (now Speech & Cognition Department – GIPSA-lab) in a sound attenuated room. Two focus conditions were recorded for each sentence: no-focus (neutral) and focus on the subject. A question/answer procedure was used to elicit prosodic focus as naturally as possible.

After the recordings, audio data was normalized regarding intensity. Video clips were re-centred for the head of the speaker to be in the middle of the image. Three types of stimuli were designed: audio only (A), visual only (V) and audiovisual (AV). The AV stimuli (avi) were directly extracted from the recordings. Both audio and video streams were then isolated to design the A (wav) and V (avi) stimuli. The duration of all the stimuli was 3s.

During the functional MRI sessions, whenever no video stimulus was displayed on the screen (A only condition and Null Events), the subjects saw a mid-centred black cross that they were asked to fixate (fixation cross).

Procedures

The subjects were told that they would see and/or hear a sentence extracted from a dialogue of the following type:

A: Nico mangea le bonbon. *Nico ate the candy.*

B: Sarah mangea le bonbon ? *Sarah ate the candy?*

A: NICO mangea le bonbon. *NICO ate the candy.*

Two situations were possible: 1. B did not understand who performed the action (as in the example above) and A corrected him (focus case); 2. B understood well but was unsure and repeated the sentence in a question mode and A simply repeated the correct sentence in a neutral mode (no focus case). The subjects were asked to identify the situation: A corrected B or not. They were therefore indirectly asked to identify focus cases.

The stimuli (video and/or audio) were presented to the subjects inside the MR scanner using Neurobehavioral Systems' Presentation software. Audio was presented via MR-compatible headphones (Hitachi Advanced Systems' ceramic transducer headphones). Video was presented through a projector located outside the MR room and a mirror positioned inside the head coil just above the subject's eyes. The subjects responded via a MR-compatible two-button response box placed in their right hand. They were told beforehand that they had to provide an answer even if they were not sure and that they should respond only once the stimulus was finished (fixation cross back).

Prior to the experiment, subjects were trained with the experimental task. Once inside the MR-scanner, the subjects could train using the response box and audio volume was adjusted.

fMRI paradigm

The fMRI procedure consisted of an event-related pseudo-random design. Four functional scans were acquired: two for normal speech and two for whispered speech. Each functional scan consisted of 12 sentences in six conditions: AV+focus, AV+no focus, A+focus, A+no focus, V+focus and V+no focus (72 stimuli). For normal speech, one functional scan used the first 12 sentences available in all conditions and the second the last 12 sentences. For whispered speech, the two sessions both consisted of the same 6 sentences replicated twice for all conditions. A total of 14 null events (NE, fixation cross) were added to the 72 stimuli to vary intertrial interval times. Trials were presented as events lasting 5.1s: stimulus (3s) + response delay (2.1s). NE also lasted 5.1s. The trial sequences were presented following a pseudo-random order. Total duration of a scan was approximately 7 min (5.1s × 86 trials).

fMRI data collection and processing

Blood Oxygenation Level-Dependent (BOLD) contrast was measured over the whole brain using a Shimadzu-Marconi Magnex Eclipse 1.5 Tesla Power Drive 250 imager at the ATR Brain

Activity Imaging Center. T2*-weighted images were acquired using a gradient echo-planar sequence (repetition time: 2000ms; echo time: 48ms; flip angle: 80°; 5 initial dummy scans). A total of 20 axial sequential slices with a 3×3×6 mm voxel resolution were acquired. They covered the cortex and cerebellum. A high-resolution 3D anatomical MR scan of the volume examined functionally was acquired after the four functional MR scans (191 slice RF-FAST sequence: repetition time: 12ms; echo time: 4.5ms; flip angle: 20°; voxel size: 1×1×1 mm).

Functional data analysis was performed using the SPM2 software (Statistical Parametric Mapping-Wellcome Department of Cognitive Neurology, London, UK) running on a PC under MATLAB (Mathworks, Sherbon, USA). All the functional images were pre-processed as follows. Movement artifact was corrected for by realigning the images. Differences in acquisition time between slices were accounted for and images were spatially normalized to a standard space. Finally, the images were smoothed using a 6×6 ×14 Gaussian kernel.

The haemodynamic response to the onset of each event was modeled using a delayed haemodynamic response function (HRF). Contrasts were determined voxelwise using a General Linear Model. Statistical significance threshold for individual voxels was established at $p < .01$ uncorrected. Clusters of activated voxels were then identified, based on the intensity of the individual responses and the spatial extent of the clusters (10 voxels). A random effect analysis was then conducted for the group of 12 subjects.

A.5. Methods for the multimodal fMRI study

fMRI paradigm

A pseudorandom, event-related fMRI paradigm was employed. Four functional scans were acquired, one for each type of pointing. Each functional scan included a sequence of the following five conditions: Preparation of the control task (Pc), Preparation + Execution of the control task (PEc), Preparation of the pointing task (Pp), Preparation + Execution of the pointing task (PEp) and a null event (NE, black fixation cross). The five conditions were alternated between scans and between subjects. 24 repetitions of each condition were presented. Trials were presented as events lasting 4.5 s. The duration of the PEp and PEc conditions was of 0.5 s for the initial fixation cross + 2 s for the preparation phase + 2 s for the execution phase. The Pp and Pc conditions consisted of 0.5 s for the fixation cross + 4 s for the preparation phase. NE condition also lasted 4.5 s. The trial sequences were presented in a pseudo-random optimized order. Total duration of a scan was approximately 9 min (4.5 s × 5 conditions × 24 trials).

Apparatus

Three types of behavioural responses were recorded: vocal production, eye movement and index finger movement. Verbal responses were recorded using an fMRI-compatible microphone. To minimize the amount of noise recorded, the microphone was positioned out of the scanner, at one extremity of a wave guide consisting of a soft plastic tube. The other extremity of the tube was connected with a mask placed over the subjects' mouth. This apparatus reinforced the signal-to-noise ratio. Eye position was monitored using an ASL 504 eye-tracker (Applied Science Laboratories, Bedford, MA) coupled with the scanner. Right

index finger responses were recorded using a digital camera placed out of the fMRI room, behind the window.

fMRI data acquisition and processing

A whole-body 3 Tesla MRI imager (Bruker) with gradient echo acquisition was used to measure blood oxygenation level-dependent contrast over the whole brain (repetition time: 2.5s; echo time: 30 ms; field of view: 216 x 216 mm, acquisition matrix: 72x72; reconstruction matrix: 128 x 128; 7 dummy scans). Forty-one 3.5 mm axial interleaved slices were imaged parallel to the bi-commissural plane, encompassing the whole brain and the cerebellum. A high-resolution 3D anatomical scan was obtained. Anatomical images were acquired using a sagittal MPRAGE sequence (inversion time: 900ms, volume: 176 x 224 x 256 mm; resolution: 1.375 x 1.750 x 1.33 mm; acquisition matrix: 128 x 128 x 192; reconstruction matrix: 256 x 128 x 128). A B0 fieldmap was acquired twice. Functional data analysis was performed using SPM2 software (Wellcome Department of Cognitive Neurology, London). Functional data were realigned to correct for head motion using a rigid body transformation. A spatial normalisation was applied. The functional images were spatially smoothed.

To examine cerebral activation during the crucial part of the trials (after the preparation phase), the haemodynamic response to the onset of each event was modelled with a delayed haemodynamic response function, empirically shifted to onset 3.5 s later. Contrasts between conditions were determined voxelwise using the General Linear Model. To perform a random effect analysis, the contrast images (pointing vs control) were calculated for each subject individually and were then entered into a one-sample t test with a significance threshold of $p < .005$ uncorrected.

Behavioural results

Monitoring the subjects' utterances included word accuracy and correct prosodic contour productions. Word recognition was possible in the audio signals from all subjects but 2. F0 examination was possible on 9 out of 16 subjects. In the other cases, technical problems with the microphone or the mask drastically degraded the signal. Word accuracy and F0 patterns were correct in all available prosodic and syntactic conditions (and their controls). Eye-tracking data were analyzed with the eye-tracker software. Horizontal positions of the eyes were checked using a Matlab script. Subjects behaved according to the instructions. An overview of the video data on finger movements shows that all the subjects performed adequately.

Preliminary version produced by the authors.

In Lœvenbruck H., Dohen M. & Vilain C. (2009). In Fuchs S., Lœvenbruck H. Pape D. & Perrier P. (eds.), *Some Aspects of Speech and the Brain*, Peter Lang, 211-258, ISBN 978-3-631-57630-4. - **hal-00360758**

Running title: EMG lip activity during AVHs

An EMG study of the lip muscles during covert auditory verbal hallucinations in schizophrenia

Lucile Rapin^{1,2}, Marion Dohen², Mircea Polosan³, Pascal Perrier², Hélène Løevenbruck²

¹Département Education et Pédagogie, Université du Québec à Montréal, Québec

²GIPSA-lab, Speech and Cognition Department, UMR 5216, CNRS, Université de Grenoble, France

³Pôle de Psychiatrie et de Neurologie, CHU de Grenoble, France

Corresponding author:

Hélène Løevenbruck

GIPSA-lab, Département Parole et Cognition

and

Laboratoire de Psychologie et NeuroCognition

CNRS UMR 5105 Université de Grenoble-Alpes

Bâtiment Sciences de l'Homme et Mathématiques BP47

38040 Grenoble Cedex 9 France

Helene.Loevenbruck@upmf.-grenoble.fr

Tel: +33 4 76 82 56 75

Fax: +33 4 76 82 78 11

Word counts:

Abstract: 198

Abstract

Purpose: Auditory verbal hallucinations (AVHs) are speech perceptions in the absence of a external stimulation. An influential theoretical account of AVHs in schizophrenia claims that a deficit in inner speech monitoring would cause the verbal thoughts of the patient to be perceived as external voices. The account is based on a predictive control model, in which verbal self-monitoring is implemented. The aim of this study was to examine lip muscle activity during AVHs in schizophrenia patients, in order to check whether inner speech occurred.

Methods: Lip muscle activity was recorded during covert AVHs (without articulation) and rest. Surface electromyography (EMG) was used on eleven schizophrenia patients.

Results: Our results show an increase in EMG activity in the orbicularis oris inferior muscle, during covert AVHs relative to rest. This increase is not due to general muscular tension since there was no increase of muscular activity in the forearm muscle.

Conclusion: This evidence that AVHs might be self-generated inner speech is discussed in the framework of a predictive control model. Further work is needed to better describe how the inner speech monitoring dysfunction occurs and how inner speech is controlled and monitored. This will help better understanding how AVHs occur.

Keywords

Auditory verbal hallucinations, inner speech, EMG, orbicularis oris, self-monitoring deficit, internal models, predictive control

INTRODUCTION

Auditory verbal hallucinations (AVHs) are a common feature of the schizophrenia psychosis, affecting 50% to 80% of the patients (Andreasen, & Flaum, 1991; Nayani and David, 1996). Hallucinations are defined as “a sensory experience which occurs in the absence of corresponding external stimulation of the relevant sensory organ, has a sufficient sense of reality to resemble a veridical perception over which the subject does not feel s/he has direct and voluntary control, and which occurs in the awake state” (David, 2004, p. 110). AVHs or “hearing voices” can thus be considered as speech perception in the absence of a relevant external acoustic input. AVHs are often distressing and engender suffering, functional disability as well as social marginalization (Franck, 2006). They are a complex and non-unitary phenomenon that can be explained by different underlying mechanisms (Larøi & Woodward, 2007).

Many theoretical models have been proposed to explain AVHs in schizophrenia (see e.g. David, 2004, for a review). A widely cited and influential model formulates AVHs as a dysfunction of the monitoring of inner speech (Feinberg, 1978; Frith, 1992). Inner or covert speech can be defined as internal non-vocalized, non-articulated speech, produced in one’s mind (MacKay, 1992). It has been shown to activate a similar network of cerebral regions as overt speech (McGuire, Silbersweig, Murray, David, Frackowiak, & Frith, 1996; Shergill, Bullmore, Brammer, Williams, Murray, & McGuire, 2001). The model claims that, due to a failure of the self-monitoring mechanism, the inner speech of the patient is not identified as self-generated and is experienced as coming from an external source, i.e., as external voices. In other words, the model posits that the mechanism for distinguishing self-generated from externally-generated percepts is deficient.

Specifically, it has been suggested that inner speech may share some properties with motor actions (Feinberg, 1978; Frith, 1992; Jones & Fernyhough, 2007). Although the hypothesis that “inner speech is a kind of action” (Gallagher, 2004, p. 11; Jones & Fernyhough, 2007, p. 396) is still a matter of debates (see e.g. Vosgerau & Newen, 2007) it has led to interesting theoretical modelling, via the motor control framework (Frith, 1996) and typically via the ‘predictive model’ (also called ‘comparator model’). In this framework, it is suggested that the brain uses forward modelling to predict the consequences of action (Wolpert, Ghahramani, & Jordan 1995; Wolpert, 1997). When motor commands are sent to the motor system to achieve a particular intended end-state, an efference copy is issued in parallel. This efference copy is used to calculate a prediction of the sensory state of the motor plan.

As illustrated in Figure 1, the predictive model includes two internal models (Kawato, Furukawa & Suzuki, 1987; Wolpert, 1997; Wolpert, *et al.*, 1995): (a) an *inverse model* which computes motor commands based on the intended action and (b) a *forward model* which predicts the sensory consequences from a copy of the motor commands. The inverse model transforms the intention of action into motor commands, which are sent to the motor system and generate movement. A copy of the motor commands is simultaneously sent to the forward model, which generates a prediction of the sensorimotor consequences of the motor commands.

PLACE FIGURE 1 AROUND HERE

The predictive model has many advantages. First, with a predictive system, errors can be detected before sensory feedback reaches the central nervous system. The propagation of the sensory feedback to the central nervous system is not instantaneous, due to axon transmission times and synaptic delays. The predictive system can overcome this problem. The forward model predicts the consequences of implementing motor commands in terms of kinematics and sensations. If we can predict where and when our tongue is going to be after the motor command has been issued, then we can check whether the correct command has been selected before we actually make the movement. This internal feedback is available much more rapidly than the actual feedback signals resulting from the movement. Motor commands can therefore be adjusted online to reach the intended goal (integrating the error in the comparison between predicted state and intended state, see Figure1), without having to rely on delayed sensory feedback, which makes for smooth motor actions (Kelso, 1982; Miall & Wolpert, 1996).

Secondly, the predictive system provides information about the source of sensations, or agency. It is suggested that if the actual sensory feedback matches the predicted state then self-authorship is experienced. Frith (1996) claims that a defective predictive system could explain why a self-initiated action may be experienced as

controlled by someone else in the ‘delusion of control’ symptom: if the predicted sensory feedback (corollary discharge) does not match the actual sensory feedback then some external influence must have been at work.

A further advantage of the predictive model is that when actual and predicted states match, the sensory effect of the motor act is modulated, attenuating it perceptually compared with identical stimulation that is externally produced (Blakemore, Frith, & Wolpert, 1999; Blakemore, 2003; Frith, 2002). This suggests that a modulation/attenuation of sensory cortex activity occurs when an action is self-initiated. In other words, if we can predict the sensations we are going to feel then they are not needed and they can be attenuated.

Forward and inverse models can be improved or adapted on the basis of the errors resulting from the comparison between the actual sensory experience and the predicted and intended states respectively (see Figure 1). Although the use of internal models of sensory expectations in motor control is a fruitful hypothesis in motor control research and specifically in speech monitoring (Guenther, Ghosh & Tourville, 2006; Houde & Nagarajan, 2011; see also Postma, 2000) some researchers have provided data that question the fact that internal models can be adapted or generalised. Tremblay, Houle & Ostry (2008) for instance, show no generalisation of adaptation to altered jaw dynamics during word production, contrary to the predictions of control based on internal model improvement or generalisation.

Predictive control modelling of movement (including speech) has been suggested to also apply to inner speech or conscious thought (Feinberg, 1978, Frith, 1996). Implication of a corollary discharge mechanism in inner speech control is supported by several neuroimaging studies demonstrating a frontal modulation of the temporal regions during inner speech production (Ford & Mathalon, 2004; Paus, Perry, Zatorre, Worsley & Evans, 1996 (for silent speech, without vocalization but with articulation); Shergill *et al.*, 2002). In addition, the involvement of a motor control system during inner speech production is also supported by empirical research suggesting that motor commands are emitted. Subtle muscle activity has been detected in the speech musculature using electromyography (EMG) during verbal mental imagery, silent reading and silent recitation (Jacobson, 1931; Livesay, Liebke, Samaras, & Stanley, 1996; McGuigan & Dollins, 1989). This suggests that motor commands are generated during inner speech production in healthy participants.

AVHs in schizophrenia can be explained in the framework of predictive control of inner speech, as proposed by Seal, Aleman & McGuire (2004) as well as Jones & Fernyhough (2007) (see Feinberg, 1978 for an earlier similar account). These authors suggest that, during AVH, because of a distortion or absence of a predicted state, the actual sensory consequences of the inner speech are not attenuated. Either because of attributional biases (Seal *et al.*, 2004), or because the emotion of self-authorship is not felt (Jones & Fernyhough, 2007), inner speech is then experienced as other-generated. Therefore, a dysfunction in the predictive model would result in the experience of an AVH.

Neuroimaging studies of schizophrenia patients have provided evidence for the suggestion that the corollary discharge from the frontal speech areas fails to inform the temporal lobes that the auditory input is self-generated. Disrupted fronto-temporal pattern of cerebral activation has been shown in schizophrenia patients relative to healthy participants (Ford & Mathalon, 2005; Lawrie *et al.*, 2002; McGuire *et al.*, 1995; Rapin *et al.*, 2012).

Within this framework, i.e., if AVHs are the results of inner speech not being identified as self-produced due to a predictive control deficiency, then AVHs should be associated with inner speech production. Therefore, the presence of speech-related motor commands during AVHs would confirm that AVHs are self-generated inner speech production. If they exist, these motor commands could result in very small and visually undetectable activity of the orofacial speech muscles. Several studies have tried to examine orofacial muscular activity during AVHs.

Gould (1948, 1949) measured the myoelectric activity of the vocal musculature (lower lip and chin) during a resting period in patients with and without AVHs as well as in healthy participants. He concluded that AVHs are associated with activity of the speech musculature. Nevertheless, he did not control for the presence vs. absence of AVHs and he also observed high amplitude muscular activity in 28% of the control participants, suggesting that the recorded activity might in fact have corresponded to a generalized muscular tension and not to AVHs. In one of the patients, subvocal speech (slight whisper) was also present during AVHs, and the content of the audio

recording corresponded to the patient's report of the voices' content. Therefore, Gould's findings are in line with the account of AVH as self-generated soft (or quiet) speech, but not exactly inner speech. The activity measured by Gould could in fact correspond to a subtype of AVHs, produced with subvocalization. AVHs accompanied by movements of the lips and tongue have been described as a specific form of AVHs (Séglas, 1892) but many hallucinating patients do not subvocalize or articulate. Therefore Gould's finding could be specific to a particular type of hallucinations. Few studies tried to replicate Gould's observations. McGuigan (1966) studied one patient who signalled his AVHs by pressing a button at the start of a voice until the end of its occurrence. McGuigan observed a significant increase in the chin muscles activity two seconds prior to an AVH report but not six seconds before. Crucially, the increase in the chin muscle activity was accompanied by a faint whisper. Inouye and Shimizu (1970) also reported increased invasive EMG activity from the speech musculature in nine hallucinating patients in 47.6% of their hallucinations. Similar to McGuigan's findings, they observed a muscular activity increase prior to the start of an AVH. Nonetheless, the authors did not specify whether AVHs were accompanied by subvocalization. In the same theoretical framework, Bick and Kinsbourne (1987) found that patients report less AVHs when they are asked to perform a motor task involving the mouth (vs. control motor tasks).

In parallel with these few findings in which AVHs seemed to be associated with EMG activity, contrary results were also found. Junginger and Rauscher (1987) recorded vocal and non-vocal invasive EMG activity in 19 hallucinating and 22 non-hallucinating psychiatric patients. The authors failed to find neither a relationship between increases in vocal EMG and reports of hallucinations nor an acoustic subvocal activity. Green and Kinsbourne (1990) did not find EMG activity during AVHs either. These results suggest that when AVHs are not subvocalized they would not systematically involve speech muscle activity. This suggestion is supported by a review by Stephane and colleagues (2001).

In sum, some studies have suggested the presence of orofacial muscular activity during AVHs. The findings were not systematically replicated, however, making it impossible to conclude on a speech-related muscular activity during AVHs. These studies had important methodological limitations (limited number of participants, lack of clear correlation between EMG measurements and AVH reports, imprecision in the time-windows taken into account for analysis). Moreover, most of the studies concluding positively on a muscular activity associated with AVHs either also clearly showed the existence of a subvocal activity or did not specify whether the muscular activity was accompanied by subvocalization. This suggests that these positive results could be valid only for a specific type of hallucinations. Therefore it seems crucial to reproduce muscular recordings in schizophrenia patients suffering from non-subvocalized AVHs in order to check that the inner-speech-monitoring-dysfunction account is suited for all types of AVHs.

The question whether "inner" AVHs result in myoelectrical activity still holds, as the EMG studies of AVH provide conflicting results. The aim of this study was thus to examine lip muscle activity during the occurrence of AVHs in schizophrenia patients, without co-occurrent subvocalization, using a replicable protocol. If the inner-speech-monitoring-dysfunction account is correct, then we predicted that lip muscle activity should be higher during AVHs than during a resting state without AVHs.

Because of the exploratory nature of this study, no average measurements were known beforehand, making it impossible to predict magnitude effects and to design for a power analysis. An emphasis was put on designing an efficient and suitable protocol for the patients. We were interested in collecting data from patients to obtain objective measurements on speech muscular activation during AVHs as well as to derive an order of magnitude usable in further studies.

METHODS

Participants

14 schizophrenia patients (mean age=37.8, standard deviation (SD) =12.1) participated in this study. Three participants were discarded because the symptom of AVH was not confirmed. The group of 11 patients analyzed (mean age=37.17, SD=12.5) included five women. Eight patients were right-handed, two were left-handed and one was ambidextrous (Oldfield, 1971). All patients were native speakers of French. Under the DSM-IV-TR criteria (APA, 2000), 10 patients were diagnosed paranoid schizophrenic and one undifferentiated schizophrenic. All patients were receiving antipsychotic medication and were experiencing frequent hallucinations at the time of

the study (as assessed by an AVH questionnaire). Patients presenting language and/or auditory deficits were excluded from the study. Patients were recruited at the *Centre Hospitalier Universitaire* of Grenoble (Grenoble university hospital) and gave written consent to participate in the study, which was approved by the university and hospital ethics committee of Clinical Research (CPP-09 CHUG-17).

Preliminary Study

To measure possible muscular activity associated with the beginning of an AVH, it was necessary to know when the phenomenon happened by asking the patient to signal its occurrence. Obviously, the patient cannot report her/his AVHs orally since this would result in an artefactual activation of the speech musculature. A number of previous studies (McGuigan, 1966; Inouye & Shimizu, 1970) used a button pressing method to signal AVHs. The patients pressed a button once they heard a voice and kept it pressed until the voice ended. This method was used in our experiment. In order to take into account the possible delay between hearing an AVH and the report of this perceptual experience, it was necessary to estimate the button pressing reaction time to hearing speech. In order to do so, six schizophrenia patients (mean age=36, SD=13.9) from our sample additionally underwent a behavioral experiment examining this reaction time. A speech sequence consisting of a combination of six isolated words and six full sentences, separated by silences of random durations, was played through loudspeakers. Patients were instructed to press the button when they heard an utterance and to keep it pressed until the end of the utterance. Results showed a reaction time to hearing speech of 501 ms (SD=385ms). This finding was taken into account in the data processing to label the beginning of a hallucinatory period (see section on data processing).

Setup for recordings

Data were recorded with a Biopac MP150 surface EMG system, using the Acqknowledge software (www.cerom.fr). Bipolar surface electromyography (sEMG) recordings were obtained from two orofacial speech muscles: orbicularis oris superior (OOS) and orbicularis oris inferior (OOI). The orbicularis muscles were chosen because they showed the most activation in the EMG studies presented in the introduction. Figure 2 shows the electrodes placed on the two lip muscles. A muscle in the forearm (brachioradialis) served as a control to ensure that the measured orofacial muscular activations were associated with speech and not with a generalized muscle contraction. The forehead muscles were not used because of their role in emotion-related face motions, which are common in patients experiencing AVHs.

The experiment was video-monitored using a Canon video camera to track visible facial movements. A microphone was placed 15cm away from the patient's lips to record the patient's speech during the overt reading conditions and to monitor any faint speech production during AVHs. sEMG and audio data were all recorded through the Biopac acquisition system and were thus synchronized. The recorded signals were: the muscular activities (sEMG electrodes), the audio signal, the button pressing signal and the e-prime stimuli presentation signal when used (see hereafter). These signals were recorded at the following sample frequencies: muscles: 3125Hz; audio: 25000Hz; button: 3125Hz; e-prime: 3125Hz. The experiment took place at the Grenoble university hospital.

PLACE FIGURE 2 AROUND HERE

Procedure

A training session was first executed and then two conditions (control and test) were run.

In the *training* condition, patients were asked to read aloud isolated words and sentences at a comfortable rate. This condition lasted approximately five minutes and was used to familiarize the patient with the setup, and to check the electrodes were well positioned on the orbicularis oris muscles.

In the *resting condition*, patients were asked to remain silent and not to move for 10 seconds. This condition was used to obtain a baseline measurement of muscle activity.

In the *hallucinatory condition*, patients were asked to remain silent, not to move, and not to prevent voices from arising. They were instructed to signal when they heard a voice by pressing a button^{bb} until the end of the utterance. They were told not to verbally interact with the voice, in order to limit the occurrence of subvocal speech associated with their own response to the voice. This condition lasted approximately 15 minutes, depending on the patient's responsiveness and degree of anxiety.

Our expectation was to find higher sEMG activations in the hallucinating state than in the resting state (in which quasi-constant baseline activation was expected).

The patients underwent a pre-experiment questionnaire on the frequency, latency, nature and content of their everyday AVHs and a post-experiment questionnaire inspired from the Hoffman AVH rating scale (Hoffman *et al.*, 2003), on AVHs experienced during the hallucinatory condition.

The experiment thus unfolded as follows: pre-experiment questionnaire, silent condition, training condition, silent condition, hallucinatory condition, silent condition, post-experiment questionnaire.

Data Processing

With regards to the hallucinatory periods, audio, video and EMG signals were used to eliminate any AVHs with subvocalizations. More specifically, AVHs in which sound or visible muscular activity was detected (in the audio, video or EMG signals) were excluded from the analysis. If either the audio recording showed any lip or mouth noise, or if the video signal showed any lip or mouth movement, or if the EMG signal contained a visually obvious burst of activity corresponding to lip contraction, that part of the signal was excluded. This represented less than 5% of the data. Therefore, the AVHs data analyzed consisted of non-subvocalized AVHs occurrences only.

Data were exported from the Acqknowledge software to a Matlab format (www.mathworks.fr. 9.0). EMG data were pre-processed as follows. A 50Hz frequency comb filter was applied to eliminate power noise. A [10Hz-300Hz] bandpass filter was then applied in order to focus on the orofacial sEMG frequency band (DeLuca, 2002). The filtered EMG signal was then centered by subtracting its mean.

The analysis of the muscle activity was performed on specific time windows focused on the tasks. In the hallucination condition, a window starting 500ms before each button press was chosen to take the reaction time into account (see above). We had originally planned to set the end of the analysis window at the time of button release, *i.e.* at the end of the hallucination as signaled by the patient. However after examining several patients, we were not sure of the ability of the patient to rapidly signal the end of a hallucination. If, say, the hallucination began with several sentences, then the patient would wait after the last hallucinatory sentence, to make sure another one was not coming. This would make the hallucinatory period longer than it really was and possibly lead to the presence of potential interferences such as swallowing (visible muscle activity) or self-intended verbal thinking (non-visible muscle activity). Moreover, for some patients, it appeared that AVHs were so frequent that they maintained the button pressed at all times even if there were some short breaks between AVHs. We therefore restricted the analysis to the beginning of the AVHs to make sure the analysis window only contained hallucinatory events. A time window of [-500, 1500] ms around the button press thus corresponded to an interval where an AVH had most certainly occurred. Some patients nevertheless experienced AVHs lasting less than 2,000 ms and when this was the case, the actual time of button release was used for the end of the time window. This analysis strategy leads to parts of the AVH data not being used since the long AVHs were not fully taken into account (only their beginnings).

For the baseline, it was finally impossible to use the signals from the resting condition since, upon asking the patients whether they had experienced AVHs during these periods, most of them answered positively. In order to have baseline signals without AVHs, the rest signals were manually detected using parts of the signals recorded during the hallucinatory condition in which there was no button pressing *i.e.* no reported AVHs. These signals will be referred to as "Rest condition", even though in most cases they were not recorded during the rest period. These Rest signals were chunked into two-second periods to match the AVH time window length. We took the

^{bb} In a first version, patients were asked to press a button as soon as they heard a voice and to press it again at the end of the hallucinatory fragment (sentence, word or discourse).

same number of Rest and AVH trials. The Rest trials were picked in a time window close to an AVH in order to maintain physiological state homogeneity.

Once each trial was isolated for each condition and each participant, the maximum of the absolute value of each pre-processed sEMG signal (one per muscle) was computed in the corresponding time window using a 200ms sliding window. Maximum activation was chosen because any lip muscle activity observable in the AVHs should be of extremely low amplitude (since no movement of the lips were involved), and because lip muscle activity is not constant, even in overt speech. Therefore the potential bursts of activation could be masked by noise, if an average activation pattern were calculated over each period. The maximum of the absolute value extracted was then transformed with a natural logarithm. The level of the EMG signal measured can be considered to result not only from muscular activity but also from skin conductivity and other factors that differ for each participant. If these factors (supposedly constant for each participant) are considered as multiplicative, the logarithmic transformation makes it possible to overcome them in the statistical analysis by making the model additive and therefore insensitive to this constant. It also served to smooth the edges of the distribution and reduced the influence of potential outliers on the statistical results.

A repeated-measures analysis of variance (ANOVA) was then applied to the mean activation measures over all patients with *Condition* (two levels: AVH, baseline) and *Muscle* (three levels: OOS, OOI, BR) as within-subjects factors. The analysis was carried out with SPSS 18.0 software (<http://www.spss.com/fr/>).

RESULTS

AVH reports

All participants experienced covert AVHs during the experiment. Overall, the number of “voices” heard was one for just over a third of the patients and more than five for just under a third of the patients. The voice was heard clearly equivalent to an external voice by 27% of the patients, and was reported as producing an intense sense of reality to them. The voices uttered short sentences in more than half of the cases. The content of some was reported as follows: “*Arrête, arrête maintenant*” (‘Stop, stop it now’); “*Tu es stupide*” (‘you are stupid’); “*Aide-moi*” (‘help me’); “*Arrête maintenant, ça va agir sur ton cerveau*” (‘stop it now, this is going to influence your brain’); “*accident de voiture*” (‘car accident’); “*âme, âme, pars au large*” (‘soul, soul, go offshore’). The voice’s influence on the patient was moderate, probably due to the experimental setup the patients were put in. They reported they were able to abstract themselves from the voice’s content.

EMG results

First, to check that the electrodes were well positioned and that only the electrodes on the lips were associated with speech, we compared EMG signals measured during the training condition with those measured during the experimental conditions. As expected, we found that lip muscle EMG activations were significantly higher in the (overt speech) training condition than in the other two conditions. In addition, we found that arm muscle activation was not higher in the training condition than in the other two conditions. After this preliminary assessment, the training condition data were not included in the analysis.

Table 1 provides the mean values and standard deviations of the peak activation for each muscle in each of the conditions of interest (AVH and Rest) for each participant. Note that the muscle activations in the Rest condition were not null, due in part to physiological noise (Tassinary, Cacioppo & Vanman, 2000). Figure 3 provides the mean of the peak activation in the AVH and Rest conditions over all patients for the orbicularis oris superior, the orbicularis oris inferior, and the brachioradialis. It can be seen that orbicularis oris inferior activation is stronger in the AVH condition than in the Rest condition, whereas the brachioradialis activation remains constant across conditions.

PLACE TABLE 1 AROUND HERE

The three by two repeated-measures ANOVA revealed a significant effect of *Condition* ($F(1,10)=11.54, p<.007$) as well as a significant interaction between *Condition* and *Muscle* factors ($F(2,20)=4.24, p<.029$). The *Muscle* effect was non-significant ($F(2,20)=3.11, p=.067$). In average, across all patients, the activation values were higher for the three muscles in the AVH condition than in the Rest condition (see table 1). However, the statistical

interaction between *Condition* and *Muscle* suggests that this difference is not significant for all muscles. Univariate Student t-tests were therefore performed to examine the effect of condition on each muscle separately.

Orbicularis Oris Superior

The Student t-test showed no significant difference of activation between AVH and Rest for the OOS ($t(10)=-1.65$, $p=.13$). In other words, the activation values of the OOS were not statistically significantly higher in the AVH condition than in the rest condition (figure 3, table 1).

Orbicularis Oris Inferior

The Student t-test showed significant difference of activation between AVH and Rest for the OOI ($t(10)=-3.48$, $p=.006$). This suggests an OOI activation increase in the AVH condition compared to the Rest condition (figure 3, table 1).

Brachioradialis

The Student t-test revealed no effect of condition ($t(10)=-1.54$; $p=.155$), as shown in figure 3 and table 1.

PLACE FIGURE 3 AROUND HERE

DISCUSSION

This study examined the potential activation of the speech musculature during non-subvocalized AVHs in 11 schizophrenia patients using sEMG under two conditions: during auditory verbal hallucination (AVH) and during rest. Two lip muscles involved in speech production were examined —orbicularis oris superior (OOS) and inferior (OOI)— as well as one control muscle in the forearm (BR). A condition effect was revealed suggesting that, for the three muscles, the AVH condition resulted in higher muscular average activation values than the rest condition. However the analysis of the interaction between the muscles and the condition revealed that the activation increase from rest to AVH is only significant for one speech muscle, namely the OOI. Thus, AVH is associated with a lower lip muscle activity that is significantly higher than at rest. This suggests that AVHs are associated with speech specific labial motor commands. These labial motor commands could be associated with self-generated inner speech. The fact that the activity of the brachioradialis did not significantly vary across conditions confirmed that the increase in muscular activity measured in the OOI speech muscle is actually related to speech and not to general muscular tension. The difference between the two orbicularis oris activation patterns can be explained by muscle anatomy: the sEMG electrode positioned on the OOI site could in fact capture activation of muscular fibers from other muscles involved in speech, entangled with those of the OOI, such as the depressor angular muscle (Blair & Smith, 1986) or the mentalis (Nazari, Perrier, Chabanas, & Payan, 2011). This would increase the global level of activation in the lower lip, attributed to the OOI, as compared to the OOS during speech. Thus, by its proximity with other speech muscles, the OOI site better reflects muscular activity during speech production. This finding is consistent with the literature describing OOI activity (rather than other orofacial muscle) during speech and inner speech tasks (Livesay *et al.*, 1996).

The high inter-subject variability observed in this study can be explained by the difficulty in measuring precise muscular activity using surface electrodes due to intrinsic morphological differences. Skin thickness, subcutaneous fat and facial hairs were among the non-controllable factors that influenced conductivity. Moreover, electrode placement, ambient temperature and humidity were also factors producing variability. Another factor which could potentially increase inter-subject variability is the diversity of the participants. The patients' age ranged from 19 to 61 years and the illness duration ranged from 3 to 26 years. All patients received antipsychotic medication but the treatment varied between participants. Early-onset schizophrenia is associated with a significant clinical severity and poor response to antipsychotic treatment. Age may have an impact on the patient phenotype and may affect clinical response and susceptibility to side effects (Kumra *et al.*, 2008). Moreover, the antipsychotic/anti-hallucinatory effect, based on the antidopaminergic effect of the medication, varies depending on the affinity for dopaminergic D2 receptor and the dose of antipsychotics (Kapur & Remington, 2001). The post-experiment questionnaire did reveal some differences between participants in the properties of their hallucinations. Over a third of the patients only heard one voice during the hallucinatory

periods, but some heard several voices, more than 5 voices for under a third of the patients. About a third of the participants judged the voices they heard as very clear and real. And for more than a half of the participants, the AVH consisted of very short sentences. These differences could be related to differences in age, illness duration or medication. No correlation could be found, however, between the degree of EMG activity in each patient and any of the hallucinatory properties reported by each patient in the post-experiment questionnaire, nor any of the medication. Future works should address the issue of participant diversity. Nevertheless, the statistical significance of the results obtained in this study, despite the participant diversity, is evidence for their robustness.

Our findings suggest that the hallucinatory crisis is associated with a surface electromyographic activity increase in the OOI muscle compared to baseline (i.e. rest with no AVHs). This indicates that there would be subtle speech-related muscular activity during covert AVHs providing evidence that covert AVHs could involve self-generated inner speech. This result is of importance as it provides support to the inner-speech-monitoring-dysfunction account of AVHs (Frith, 1992; Jones & Fernyhough, 2007; Seal, Aleman & McGuire, 2004). In the framework of a motor control model based on predictive control, the lower lip activity observed during AVHs could well reflect the emission of motor commands associated with the generation of inner speech. As hypothesized by the inner-speech-monitoring-dysfunction account of AVHs, in schizophrenia patients, during inner speech, motor commands would correctly be sent to the speech muscles, but the verbal self-monitoring mechanism, or the efference system, would dysfunction, resulting in inner speech being misattributed to an external agent. Our results show that AVHs are indeed associated with increased speech-related muscular activation and support the assumption that a deficit in the monitoring of inner speech is at the origin of AVHs.

As noted by Vanderwolf (1998), the neurophysiological source of this low amplitude EMG activity (during imagined action) has not been elucidated. It has been suggested that part of it arises from intrafusal muscle fibers (Malmo, 1975). Jeannerod & Decety (1995) have also suggested a mechanism by which residual EMG may be present during action imagery. According to them, during mental simulation of an action, “it is likely that the excitatory motor output generated for executing the action is counterbalanced by another parallel inhibitory output. The competition between two opposite outputs would account for the partial block of the motoneurons, as shown by residual EMG recordings and increased reflex excitability” (p. 728). In the same vein, Rossini *et al.* (1999) showed that mental simulation of finger movement increases excitability of the spinal motoneuronal pools governing muscles involved in finger movement. Furthermore, a recent examination of hand-related motoneuron activity has shown that a mental arithmetic task affects the rate and variability of the tonic discharge of motor units (Bensoussan *et al.*, 2012). The increases observed in hand EMG activity were consistent with the modulation of the motor unit discharge rate induced by mental arithmetics. This finding suggests that mental arithmetic may influence the state of the motor system, including in its most peripheral spinal component, i.e. the motoneuron. More research is necessary to better understand how (and at which neurophysiological level, cerebral, spinal, peripheral) the motor commands associated with inner speech are inhibited but may still result in residual EMG activity.

To summarise, our findings suggest that EMG activity is present in speech muscle during inner speech and are coherent with the view that “inner speech is a kind of action”. Motor commands issued during inner speech, could modulate motoneuron activity related to the lips and could result in the observed subtle increase of EMG activity.

Although the inner-speech-monitoring-dysfunction account of AVHs has been extensively described and cited (Frith, 1992; Jones & Fernyhough, 2007; Seal *et al.*, 2004), several crucial parts of it remain unexplained by its proponents. As mentioned in the introduction, the account was based on a speech motor control model involving predictive control applied to inner speech. Specifically, the account relies on the deficient comparison between the corollary discharge (or predicted sensory feedback) and the actual sensory-motor feedback leading to a lack of attenuation of the sensory consequences of self-produced actions. This would make self-produced actions, therefore, indistinguishable from externally generated sensations.

The application of the predictive control account to inner speech is problematic, however. As questioned by Frith (2012) himself, it is not clear what the actual sensory consequences of inner speech are. Inner speech generates neither kinematics nor auditory sensations, since it is internally generated. So the comparison between the

predicted state and the actual state is irrelevant in the case of inner speech. We suggest two alternative interpretations.

First, our understanding of inner speech is that motor commands are generated, that are sent to appropriate speech muscles. Inhibitory signals may be sent in parallel to prevent the intensity of the motor commands from reaching a sufficient threshold for movements of the speech organs to occur. But even though the speech organs may not move during inner speech, the presence of motor commands in the speech muscles could slightly increase muscle tension and could correspond to detectable proprioceptive feedback. This suggests that in the instance of inner speech production the actual sensory feedback would in fact be a proprioceptive sensory feedback. This residual proprioceptive feedback could then be the sensory feedback that lacks a corresponding prediction during AVH and that becomes reinterpreted as an external voice.

The second possible interpretation is that during inner speech, motor commands may be sent to the motor system, but they would be inhibited and hence irrelevant as no auditory output would be produced. A copy of these motor commands would also be sent to the forward model, the output of that being the auditory voice we hear in our head (inner speech has sensory qualities and can be associated with rich suprasegmental acoustic representations, cf. Yao *et al.*, 2012). We suggest that the relevant comparison for agency monitoring during inner speech is not that between predicted and actual sensory states but instead, the one between the intended (the input to the inverse model) and the predicted (the output of the forward model) states (the top comparator on Figure 1). Then the AVH symptoms could be explained as follows: if a deficit arises somewhere in the efference copy system, there is no match between predicted and intended states and the inner voice (predicted output) could feel alien.

The presence of EMG activity during AVH has a further possible interpretation, in agreement with mirror system or interaction theories of speech production/perception (Fadiga *et al.*, 2002; Schwartz *et al.*, 2010) or with the Motor Theory of speech perception (Liberman & Mattingly, 1985). The motor activity observed during inner speech could simply be an epiphenomenon of a sensory (auditory) processing of the inner voice. Any motor cortex activity or EMG activity recorded during AVH could in fact reflect the motor system resonance with the auditory images evoked by the AVH. The measurement used in our experiment cannot distinguish between this interpretation and our interpretation that AVH is a kind of action. We plan further cerebral imaging experiments to derive fine temporal analyses of cortical activity during inner speech production, which will allow us to examine whether motor cortex activity arises before sensory cortex activity (i.e. inner speech is a kind of action with sensory consequences) or the reverse (i.e. inner speech is a sensory experience with motor resonance).

Another part of the model that deserves further description is how “other” voices can be heard. Schizophrenia patients often report that they can precisely identify the voice they hear as being clearly that of someone they know and as addressing them in the second-person (Hoffman, Fernandez, Pittman, & Hampson, 2011). Similarly, inner speech can involve interior dialogues with oneself and other people, their imagined voices being clearly distinguishable from that of oneself. This suggests that inner speech can take several forms, own-inner-speech and imagined-other-inner-speech. Several studies have examined the behavioural and neural correlates of imagining speech in the voice of another person. Johnson *et al.* (1988), for instance, had participants listen to words spoken by a given speaker and imagine words in their own voice or in the speaker’s voice. Participants were then asked to then to discriminate words they had heard from words they had imagined. Their accuracy in discriminating between perceived and produced items was higher when had imagined the words with their own voice than with the speaker’s voice. This is evidence that participants can successfully produce inner speech in the voice of someone else. Mc Guire *et al.* (1996) examined the neural correlates of inner speech (inner generation of short sentences) and of auditory verbal imagery (imagining sentences being spoken to them in another person's voice) using positron emission tomography. Own-inner-speech and imagined-other-inner-speech were both associated with increased activity in the left inferior frontal gyrus. In addition, imagined-other-inner-speech was associated with increase in the left premotor cortex, the supplementary motor area and the left temporal cortex, suggesting more involvement of speech perception processes, when mentally using another person’s voice. Shergill *et al.* (2001) used fMRI to examine own-inner-speech and imagined-other-inner-speech. Imagined-other-inner-speech was associated with activation in the areas engaged during own-inner-speech task (namely left inferior frontal/insula region, the left temporo-parietal cortex, right cerebellum and the

supplementary motor area) plus the left precentral and superior temporal gyri, and the right homologues of all these areas. This result was interpreted as a greater engagement of verbal self-monitoring during imagined-other-inner-speech.

Further studies should investigate whether different internal models are used for different voices. If different internal models are used for different voices then inducing an adjustment of one's internal model (using somatosensory feedback alteration, for instance) should not modify the other-voices models and thus should not modify (inner) production with someone else's voice. Future studies should also examine whether speech muscles are significantly activated as compared to resting when participants are asked to imagine sentences uttered by someone else.

These hypotheses remain highly speculative and demonstrate the need of further work to examine exactly what precise part of the speech motor control system is impaired; to understand what exactly are the sensations that lead to voice hearing during inner speech; to understand how own-inner-speech and imagined-other-inner-speech are controlled and monitored. Furthermore, factors such as attention, intentionality, externalization bias, and metacognitive beliefs are of importance in the genesis of AVHs and can assist in understanding why all inner speech production is not mistaken as coming from an external agent (Aleman & Larøi, 2008; Larøi & Woodward, 2007). They should be incorporated in any thorough account of AVHs.

This preliminary study has some limitations. First, the patients were not recorded during an inner speech condition. This condition could have been compared to the AVH condition and their potential similarity could have provided additional support to the hypothesis of inner speech involvement in AVH. A pilot study showed that it was difficult, sometimes impossible, to have the patients whisper. Some patients found the whispering task very difficult. Explaining the inner speech task could then prove even more difficult. In addition, in inner speech, there is no behavioral way of controlling the patient's production. Another limitation concerns the resting state, which was not optimally controlled. In fact, it is quite possible that intrusive inner speech (flowing unbidden thoughts) was also occurring during the rest periods and could thus result in lip muscle activity. Some authors have proposed relaxation techniques to minimise myoelectric signals during rest periods (Jacobson, 1931; Vanderwolf, 1998). This strategy could be used in further experiments.

Note that an alternative interpretation of our results could be that the collected muscular activities during the hallucinatory period correspond to the patients' replies to the voices and not to the voices themselves, which would also result in lip muscle activity. It has been shown that some patients suffering from AVHs develop a dialogue with their voices in order to tell them to stop for example (Nayani & David, 1996). In our study, the instructions were however to let the voices come and not to reply to them. All patients understood these instructions, which were moreover repeated throughout the condition.

In summary, we collected lip muscle activity during hallucinatory periods and rest in 11 schizophrenia patients. A significant increase in the activity of the lower lip muscle was observed during AVHs controlled for subvocalization compared to rest, suggesting that speech motor commands are emitted during AVHs, even without co-occurrent subvocalization. This result supports the hypothesis according to which AVHs correspond to self-generated inner speech, which would be mistakenly attributed to an external agent. Furthermore, it underlines the motor nature of AVHs, which can be analyzed in the framework of a speech motor control model including predictive control and is in agreement with theories accounting for AVHs as a misattribution of inner speech. Yet it questions the inner-speech-monitoring-dysfunction account in several of its crucial assumptions, which further studies should explore.

ACKNOWLEDGMENTS

This project was partially funded by a grant from the Région Rhône-Alpes Cluster 11 "Handicap Vieillesse Neurosciences" to H. Løevenbruck and M. Polosan. We are very thankful to Lionel Granjon (Gipsa-lab) for his assistance in data preprocessing and analysis. We thank our colleagues from Gipsa-lab, Coriandre Vilain, Alain Arnal and Christophe Savariaux for technical help. We also thank Stéphane Rousset (LPNC, Grenoble) and Jean-Luc Schwartz (Gipsa-lab) for helpful advice.

REFERENCES

- Aleman, A., & Larøi, F. (2008). *Hallucinations: the science of idiosyncratic perception*. Washington, DC : American Psychiatric Association.
- American Psychiatric Association (2000). *Diagnostic and statistical manual of mental disorders: DSM-IV-TR* (4th ed.). Washington, DC: American Psychiatric Association.
- Andreasen, N.C. & Flaum, M. (1991). Schizophrenia: the characteristic symptoms. *Schizophrenia Bulletin*, 17(1), 27-49.
- Bensoussan, L.; Duclos, Y. & Rossi-Durand, C. (2012). Modulation of human motoneuron activity by a mental arithmetic task. *Human Movement Science*.
- Bick, P.A., & Kinsbourne, M. (1987). Auditory hallucinations and subvocal speech in schizophrenic patients. *American Journal of Psychiatry*, 144 (2), 222-225.
- Blakemore, S-J. (2003). Deluding the motor system. *Conscious Cognition*, 12, 647-655.
- Blakemore, S.-J., Frith, C. D., & Wolpert, D. W. (1999). Spatiotemporal prediction modulates the perception of self-produced stimuli. *Journal of Cognitive Neuroscience*, 11, 551–559.
- Blair, C. & Smith, A. (1986). EMG recording in human lip muscles: can single muscles be isolated? *Journal of Speech and Hearing Research*, 29, 256-266.
- David, A.S. (2004). The cognitive neuropsychiatry of auditory verbal hallucinations: an overview. *Cognitive Neuropsychiatry*, 9(1-2), 107-123.
- DeLuca, C.J. (2002). Surface electromyography: Detection and recording. <http://www.delsys.com/library/papers/SEMGintro.pdf>
- Fadiga, L.; Craighero, L.; Buccino, G. & Rizzolatti, G. (2002). Speech listening specifically modulates the excitability of tongue muscles: a TMS study. *European Journal of Neuroscience*, 15, 399-402.
- Feinberg, I. (1978). Efference copy and corollary discharge: implications for thinking and its disorders. *Schizophrenia Bulletin*, 4, 636-640.
- Ford, J.M., & Mathalon, D.H. (2004). Electrophysiological evidence of corollary discharge dysfunction in schizophrenia during talking and thinking. *J. Psychiatr Res.*, 38(1), 37-46.
- Ford, J.M., & Mathalon, D.H. (2005). Corollary discharge dysfunction in schizophrenia : can it explain auditory hallucinations? *International Journal of Psychophysiology*, 58, 179-189.
- Franck, N. (2006). *La schizophrénie: La reconnaître et la soigner*. *Schizophrenia: Detection and care*]. Paris: Odile Jacob Publishing.
- Frith, C.D. (1992). *The cognitive neuropsychology of schizophrenia*. London: Lawrence Erlbaum, associates.
- Frith, C.D. (1996). The role of the prefrontal cortex in self-consciousness: the case of auditory hallucinations. *Philos. Trans. R. Soc. Lond. B Biol. Sci*, 351 (1346), 1505–1512.
- Frith, C.D. (2002). Attention to action and awareness of other minds. *Conscious Cognition*, 11, 481-487.
- Frith, C. (2012). Explaining delusions of control: The comparator model 20 years on *Consciousness and cognition*, 21, 52-54.
- Gallagher, S. (2004). Neurocognitive models of schizophrenia: a neurophenomenological critique. *Psychopathology*, 37, 8-19.
- Gould, L.N. (1948). Verbal hallucinations and activity of vocal musculature: an electromyographic study. *American Journal of Psychiatry*, 105, 367-372.
- Gould, L.N. (1949). Auditory hallucinations in subvocal speech: objective study in a case of schizophrenia. *Journal of Nervous Mental Disorders*, 109, 418-427.
- Green, M.F., & Kinsbourne, M. (1990). Subvocal activity and auditory hallucinations: clues for behavioral treatments? *Schizophrenia Bulletin*, 16(4), 617-625.
- Guenther, F.; Ghosh, S. & Tourville, J. (2006). Neural modeling and imaging of the cortical interactions underlying syllable production. *Brain and language*, 96, 280-301.
- Hoffman, R.E., Fernandez, T., Pittman, B., & Hampson, M. (2011). Elevated functional connectivity along corticostriatal loop and the mechanism of auditory/verbal hallucinations in patients with schizophrenia. *Biological Psychiatry*, 69, 407-414.
- Hoffman, R.E., Hawkins, K.A., Gueorguieva, R., Boutros, N.N., Rachid, F., Carroll, K., *et al.* (2003). Transcranial magnetic stimulation of left temporoparietal cortex and medication-resistant auditory hallucinations. *Archives of General Psychiatry*, 60, 49-56.

- Houde, J. & Nagarajan, S. (2011). Speech production as state feedback control. *Frontiers in Human Neuroscience*, 5 :82.
- Inouye, T., & Shimizu, A. (1970). The electromyographic study of verbal hallucinations. *Journal of Nervous Mental Disorders*, 151, 415-422.
- Jacobson, E. (1931). Electrical measurements of neuromuscular states during mental activities. VII. Imagination, recollection, and abstract thinking involving the speech musculature. *American Journal of Physiology*, 97, 200-209.
- Jeannerod, M. & Decety, J. (1995). Mental motor imagery: a window into the representational stages of action. *Current Opinion in Neurobiology*, 5, 727-732.
- Johnson, M.; Foley, M. & Leach, K. (1988). The consequences for memory of imagining in another person's voice. *Memory & Cognition*, 16, 337-342.
- Jones, S.R., & Fernyhough, C. (2007). Thought as action: inner speech, self-monitoring, and auditory verbal hallucinations. *Conscious Cognition*, 16(2), 391-399.
- Junginger, J., & Rauscher, F.P. (1987). Vocal activity in verbal hallucinations. *Journal of Psychiatry Research*, 21(2), 101-109.
- Kapur, S. & Remington, G. (2001). Dopamine D2 Receptors and Their Role in Atypical Antipsychotic Action: Still Necessary and May Even Be Sufficient. *Biol Psychiatry*, 50, 873–883.
- Kawato, M., Furukawa, K., & Suzuki, R. (1987). A hierarchical neural network model for the control and learning of voluntary movements. *Biological Cybernetics*, 56, 1-17.
- Kelso, J. A. S. (1982). Concepts and issues in human motor behavior: Coming to grips with the jargon. *Human motor behavior: An introduction*, Kelso J. A. S. (ed.), Lawrence Erlbaum Associates Hillsdale, NJ, 21-62.
- Kumra, S., Oberstar, J.V., Sikich, L., Findling, R.L., McClellan, J.M., Vinogradov, S. & Schulz, S.C. (2008). Efficacy and tolerability of second-generation antipsychotics in children and adolescents with schizophrenia. *Schizophr Bull.*, 34, 60–71.
- Larøi, F., & Woodward, T.S. (2007). Hallucinations from a cognitive perspective. *Harvard Reviews of Psychiatry*, 15(3), 109-117.
- Lawrie, S.M., Buechel, C., Whalley, H.C., Frith, C.D., Friston, K.J., & Johnstone, E.C. (2002). Reduced frontotemporal functional connectivity in schizophrenia associated with auditory hallucinations. *Biological Psychiatry*, 51, 1008-1011.
- Lieberman, A. & Mattingly, I. (1985). The motor theory of speech perception revised. *Cognition*, 21, 1-36.
- Livesay, J., Liebke, A., Samaras, M., & Stanley, A. (1996). Covert speech behavior during a silent language recitation task. *Perception and Motor Skills*, 83 (3 pt 2), 1355-1362.
- MacKay, D.G. (1992). Constraints of theories of inner speech. In: Reisberg D, editor. *Auditory imagery*. New Jersey: Lawrence Erlbaum Associates; 121-142.
- Malmö, R. (1975). *On emotions, needs, and our archaic brain*. New-York : Holt, Rinehart & Winston.
- McGuigan, F.J. (1966). Covert oral behavior and auditory hallucinations. *Psychophysiology*, 3(1), 73-80.
- McGuigan, F.J., & Dollins, A.B. (1989). Patterns of covert speech behavior and phonetic coding. *Pavlov Journal of Biological Sciences*, 24(1), 19-26.
- McGuire, P.K., Silbersweig, D.A., Wright, I., Murray, R.M., David, A.S., Frackowiak, R.S.J., *et al.* (1995). Abnormal monitoring of inner speech: a physiological basis for auditory hallucinations. *Lancet*, 346, 596-600.
- McGuire, P.; Silbersweig, D.; Murray, R.; David, A.; Frackowiak, R. & Frith, C. (1996). Functional anatomy of inner speech and auditory verbal imagery. *Psychological medicine*, 26, 29-38.
- Miall, R.C., & Wolpert, D.M. (1996). Forward models for physiological motor control. *Neural Networks*, 9, 1265–1279.
- Nayani, T.H., & David, A.S. (1996). The auditory hallucination : a phenomenological survey. *Psychological Medicine*, 26, 177-189.
- Nazari, M. A., Perrier, P., Chabanas, M. & Payan, Y. (2011). Shaping by stiffening: a modeling study for lips . *Motor control*, 15, 1, 141-168.
- Oldfield, R.C. (1971). The assessment and analysis of handedness : the Edinburgh inventory. *Neuropsychologia*, 9, 97-113.

- Paus, T., Perry, D.W., Zatorre, R.J., Worsley, K.J., & Evans, A.C. (1996). Modulation of cerebral blood flow in the human auditory cortex during speech: role of motor-to-sensory discharges. *Eur. J. Neurosci.*, 8(11), 2236-2246.
- Postma, A. (2000). Detection of errors during speech production: A review of speech monitoring models. *Cognition*, 77, 97-132.
- Rapin, L.A., Dohen, M., Løevenbruck, H., Whitman, J.C., Metzack, P.D. & Woodward, T.S. (2012). Hyperintensity of functional networks involving voice-selective cortical regions during silent thought in schizophrenia, *Psychiatry Research: Neuroimaging*, doi:10.1016/j.psychresns.2011.12.014.
- Rossini, P.; Rossi, S.; Pasqualetti, P. & Tecchio, F. (1999). Corticospinal excitability modulation to hand muscles during movement imagery. *Cerebral Cortex*, 9, 161-167.
- Schwartz, J.; Basirat, A.; Ménard, L. & Sato, M. (2012). The Perception-for-Action-Control Theory (PACT): A perceptuo-motor theory of speech perception. *Journal of Neurolinguistics*, 25, 336-354.
- Seal, M.; Aleman, A. & McGuire, P. (2004). Compelling imagery, unanticipated speech and deceptive memory: Neurocognitive models of auditory verbal hallucinations in schizophrenia. *Cognitive Neuropsychiatry*, Psychology Press, 9, 43-72.
- Ségla J. (1892). *Des troubles du langage chez les aliénés*. [Language disorders in mentally disturbed patients]. Paris: J Rueff et Cie.
- Shergill, S.S., Brammer, M.J., Fukuda, R., Bullmore, E., Amaro, E., Murray, R.M., *et al.* (2002). Modulation of activity in temporal cortex during generation of inner speech. *Hum Brain Mapp.*, 16, 219-227.
- Shergill, S.S., Bullmore, E.T., Brammer, M.J., Williams, S.C., Murray, R.M., & McGuire, P.K. (2001). A functional study of auditory verbal imagery. *Psychol Med.*, 31(2), 241-253.
- Stephane, M., Barton, S., & Boutros, N.N. (2001). Auditory verbal hallucinations and dysfunction of the neural substrates of speech. *Schizophrenia Research*, 50, 61-78.
- Tassinari, L. G., Cacioppo, J. T., & Vanman, E. J. (2007). The skeletomotor system: Surface electromyography. In J. T.Cacioppo, L. G. Tassinari, & G. G. Berntson (Eds.), *Handbook of psychophysiology* (3rd ed., pp. 267–297). Cambridge, England: Cambridge University Press.
- Tremblay, S., Houle, G., & Ostry, D.J. (2008). Specificity of speech motor learning. *J Neurosci*, 28, 2426–2434.
- Vanderwolf, C.H. (1998). Brain, behavior, and mind: what do we know and what can we know? *Neuroscience and Biobehavioral Review*, 22, 125-142.
- Vosgerau, G. & Newen, A. (2007). Thoughts, motor actions, and the self. *Mind & language*, Wiley Online Library, 22, 22- 43.
- Wolpert, D.M. (1997). Computational approaches to motor control. *Trends in Cognitive Sciences*, 1(6), 209-216.
- Wolpert, D.M., Ghahramani, Z., & Jordan, M.I. (1995). An internal model for sensorimotor integration. *Science*, 269, 1880-1882.
- Yao, B., Belin, P., & Scheepers, C. (2012). Brain ‘talks over’ boring quotes: Top down activation of voice selective areas while listening to monotonous direct speech quotations *NeuroImage*. 60 (3), 1832-1842.

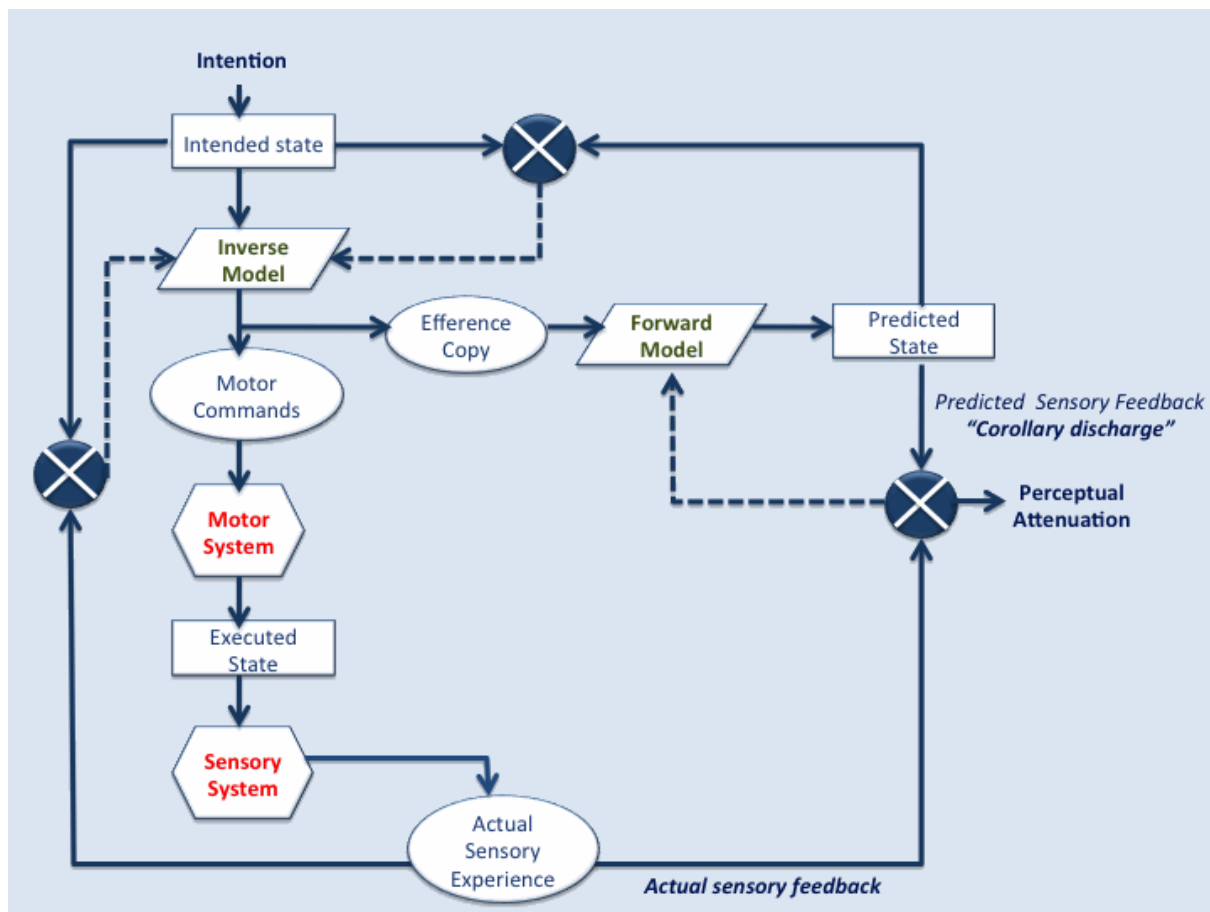


Figure 1: The Predictive Model: a forward model predicts the sensory consequences of action (estimated sensory feedback). This feedback can be used to cancel out self-generated sensory events, thus distinguishing them from sensory events with an external cause. Adapted from “Explaining the Symptoms of Schizophrenia: Abnormalities in the Awareness of Action,” by C. Frith, S. Blakemore, and D. Wolpert, 2000, *Brain Research Reviews*, 31, p. 359. Copyright 2000 by Elsevier. Adapted with permission.

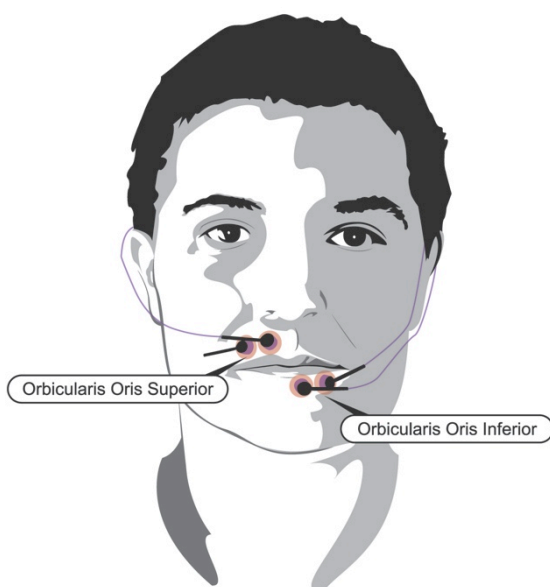


Figure 2: Surface electromyography electrodes placed on the lip muscles examined.

		P1	P2	P3	P4	P5	P6	P7	P10	P11	P13	P14	Mean Value	Standard Deviation
Orbicularis Oris Superior	AVH	2.36	4.10	5.38	3.45	3.36	2.57	3.60	4.39	3.35	2.75	1.74	3.37	1.020
	Rest	2.44	3.95	5.13	3.48	3.46	2.54	2.66	3.95	3.34	2.49	1.88	3.21	0.926
Orbicularis Oris Inferior	AVH	2.60	4.74	4.83	4.47	3.26	2.66	3.03	3.80	3.63	4.11	1.71	3.53	0.985
	Rest	2.27	3.87	3.73	4.06	3.33	2.51	2.10	3.62	3.68	3.17	1.55	3.08	0.839
Brachio-radialis	AVH	3.62	2.81	3.67	2.87	2.77	2.70	2.32	2.51	2.61	2.69	2.46	2.82	0.439
	Rest	3.66	3.01	3.09	2.60	2.69	2.60	2.14	2.37	2.49	2.99	2.15	2.71	0.452

Table 1: Mean values and standard deviations of peak muscular activation for each muscle and in each condition, for each patient.

Note. Patients P8, P9, and P12 were excluded because the symptom of AVH was not confirmed. AVH = auditory verbal hallucination.

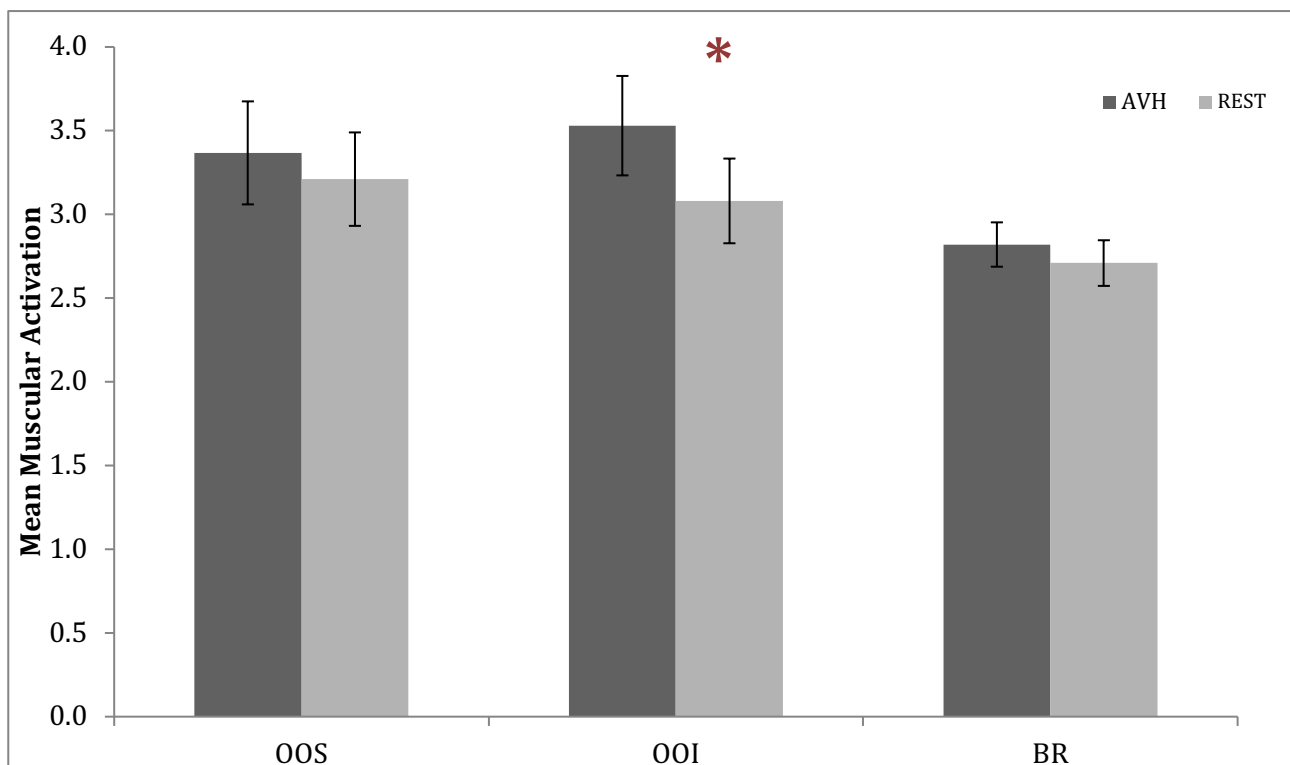


Figure 3: Means of the maxima of activation of the orbicularis oris superior (OOS), the orbicularis oris inferior (OOI) and the brachioradialis (BR) in the AVH condition (dark grey) and in the Rest condition (light grey). The error bars represent one standard error.

Preliminary version produced by the authors.

In Rapin L., Dohen M., Polosan M., Perrier P., Lœvenbruck H. (2013). *Journal of Speech, Language, and Hearing Research*, 56, S1882–S1893. [PMID : 24687444](#) - [hal-00926630](#)

Un cas d'incarnation linguistique : implication du système moteur orofacial dans les ruminations mentales

Hélène Lœvenbruck¹, Ladislav Nalborczyk^{1,2}, Marcela Perrone-Bertolotti¹, Céline Baeyens³, Romain Grandchamp¹

¹LPNC, CNRS UMR 5105, Université Grenoble Alpes, Grenoble, France

²Department of Experimental Clinical and Health Psychology, Université de Ghent, Belgique

³Université Grenoble Alpes, LIP, EA 41-45, Grenoble, France

Résumé

Deux conceptions de la nature de la parole intérieure s'opposent sur le plan théorique. Dans la lignée de la conception incarnée du langage, l'hypothèse de la *Simulation Motrice* conçoit la parole intérieure comme un acte moteur imaginé, simulé. Dans la conception mentaliste du langage, les tenants de l'hypothèse de l'*Abstraction* soutiennent que la parole intérieure est abstraite, appauvrie sur le plan articulatoire, voire amodale, et ne nécessite pas l'implication du système moteur. Prenant pour cas d'étude les ruminations mentales verbales, une forme de parole intérieure négative, répétitive, excessive et par conséquent abondante, nous proposons deux tests expérimentaux de ces hypothèses opposées. Le premier test examine si les ruminations mentales provoquées par un paradigme classique d'induction sont associées ou non à une activation motrice labiale, en mesurant l'activité électromyographique des lèvres. Le deuxième test évalue la contraposée de l'hypothèse de la Simulation Motrice, c'est-à-dire examine l'effet de la réduction de l'activité motrice orofaciale sur les ruminations, grâce à un protocole de relaxations ciblées. Nos résultats préliminaires semblent dans l'ensemble en faveur d'une nature incarnée et motrice des ruminations mentales. Le premier test montre en effet l'implication de la musculature labiale durant les ruminations. Le deuxième test suggère que le type de rumination induit par notre paradigme peut être réduit en relaxant spécifiquement les muscles de la parole, tels que les muscles des lèvres. Des études complémentaires sont envisagées pour confirmer ces résultats préliminaires, qui semblent soutenir l'hypothèse de la *Simulation Motrice*, selon laquelle la parole intérieure correspond à de la parole à voix haute simulée articulatoirement et inhibée avant l'exécution motrice.

Mots-clés

Parole intérieure - ruminations mentales - Simulation Motrice – Abstraction – EMG faciale - Relaxation

INCARNATION LINGUISTIQUE

En linguistique, l'approche incarnée trouve son origine chez Merleau-Ponty qui dès 1945 proposait que la parole soit une action, un geste et que le « geste linguistique comme tous les autres, dessine lui-même son sens » (Merleau-Ponty, 1945, p.235). Plus précisément, selon Merleau-Ponty, « La prédominance des voyelles dans une langue, des consonnes dans une autre, les systèmes de construction et de syntaxe ne représenteraient pas autant de conventions arbitraires pour exprimer la même pensée, mais plusieurs manières pour le corps humain de célébrer le monde et finalement de le vivre » (p.236). Le corps ferait donc partie intrinsèque de la parole et de la communication avec autrui. Le langage, dans sa fonction sociale de communication avec autrui, serait ainsi incarné, selon les termes actuels des sciences cognitives (e.g. Wilson, 2002). Le corps jouerait également un rôle crucial dans la fonction réflexive du langage. Selon Merleau-Ponty, la parole permet de construire, d'élaborer la pensée. « La pensée (...) n'existe pas hors du monde et hors des mots », elle coïncide avec l'acte d'expression, avec la gesticulation verbale, même intérieure. La parole (intérieure ou non) est donc conçue par Merleau-Ponty comme un geste, un processus sensori-moteur qui constitue et qui actualise le sens.

Dans la perspective théorique de Merleau-Ponty, l'approche incarnée du langage décrit le langage comme un acte moteur corporel en interaction avec l'environnement (Clark, 2006 ; Johnson & Lakoff, 2002 ; Wilson, 2002). Cette approche se place en opposition avec la conception mentaliste, selon laquelle le langage est l'expression

de représentations mentales universelles et correspond à des symboles abstraits, manipulés par des règles déconnectées du corps (cf. Chomsky, 1965). Dans une conception incarnée du langage, sémantique et pragmatique ont leur racine dans un processus sensorimoteur, une expression corporelle en interaction située avec autrui et le monde (tel que l'envisagent par exemple Bottineau, 2014 ou Perrin, dans ce volume). La parole est donc une gesticulation, ritualisée selon des normes propres à chaque culture et à chaque langue, et peu à peu intégrées par chaque individu. La question que nous posons dans cet article est la suivante : ces gestes culturels, constitutifs de la communication avec autrui et avec soi-même, sont-ils aussi en jeu, sous forme simulée, lors de la parole intérieure ? La parole intérieure peut-elle elle-même être conçue comme un acte moteur, imaginé, et peut-on observer dans certains cas des traces physiologiques de cette simulation motrice ?

LES RUMINATIONS MENTALES COMME UN CAS DE PAROLE INTERIEURE PERTINENT

Les ruminations mentales sont des pensées répétitives qui tournent autour d'un thème commun et qui se répètent en l'absence d'exigences environnementales immédiates requérant ces pensées (Martin & Tesser, 1996). Les ruminations sont fréquentes chez tous les individus. Cependant elles peuvent être excessives et devenir pathologiques (Davis & Nolen-Hoeksema, 2000). Dans ce cas, les mêmes pensées, en général négatives, sont alors ressassées sans que les personnes ne puissent s'en empêcher, comme l'illustre si puissamment Fiodor Dostoïevski dans *Crime et Châtiment* (1866) : « *Il y avait une idée étrange qui lui poussait, à coups de bec, à l'intérieur du crâne, comme un poussin qui voudrait naître, une idée qui l'occupait beaucoup, vraiment beaucoup* ». Ruminer de façon excessive entraîne une modification dans le fonctionnement cognitif impactant les performances des individus dans différents domaines. Par exemple, Davis et Nolen-Hoeksema (2000) ont montré que les ruminations excessives mènent à de plus faibles capacités de résolution de problèmes, un biais de rappel vers des souvenirs négatifs, un pessimisme excessif et une flexibilité cognitive appauvrie, avec une tendance à commettre des erreurs de type persévération dans des tests neuropsychologiques. Les ruminations, via leurs conséquences psychologiques, constituent un facteur de risque de dépression et d'autres troubles psychiatriques, tels que la phobie sociale, le trouble obsessionnel compulsif, le trouble de stress post-traumatique, les troubles des conduites alimentaires, les troubles de l'image du corps, les troubles liés à l'alcool (Davis & Nolen-Hoeksema, 2000 ; Baeyens, Kornacka, & Douilliez, 2012).

Si les ruminations prennent parfois la forme d'images visuelles, ou de sensations, elles sont le plus souvent de nature verbale (Watkins *et al.*, 2005 ; Rood *et al.*, 2012). Les ruminations mentales de type verbal sont donc une forme de parole intérieure négative, répétitive et excessive. Leur qualité abondante et excessive et leur fréquence dans la population générale en font par conséquent un cas de parole intérieure spontanée intéressant, pouvant se soumettre à l'étude objective.

NATURE ABSTRAITE OU INCARNEE DE LA PAROLE INTERIEURE ?

La parole intérieure concerne les neurosciences et la psychologie cognitives car elle joue un rôle central dans la mémoire de travail, la mémoire autobiographique, la planification, la résolution de problème, la lecture, l'écriture, la pensée, la conscience de soi, l'auto-régulation et le développement du langage chez l'enfant (pour une revue détaillée, voir Perrone-Bertolotti, Rapin, Lachaux, Baciù, & Lœvenbruck, 2014). Toutefois, les premiers à s'y être intéressés sont les écrivains (en particulier F. Dostoïevski, E. Dujardin, J. Joyce, V. Woolf, S. Beckett), les philosophes (Platon, Aristote, T. Ribot, V. Egger), les psychanalystes (notamment J. Lacan) et les chercheurs en sciences humaines, avec comme principal instrument d'exploration l'introspection et la réflexion. L'introspection a ainsi fourni des descriptions détaillées de la forme du langage intérieur. On en trouve des illustrations dans de nombreux romans, notamment ceux inscrits dans le courant du « monologue intérieur » initié par Edouard Dujardin (cf. Smadja, à paraître). La forme linguistique du langage intérieur, telle que la rapportent philosophes et écrivains, semble se distinguer de celle de la parole à haute voix. Par exemple, Victor Egger (1881) note que la parole intérieure, étant dirigée vers nous-même et ne devant être comprise que de nous-même, peut être déformée, modifiée, abrégée et est donc différente de la parole à voix haute. Vygotski (1934/1985) partage cette conception et affirme que la forme linguistique du langage intérieur le rend difficile à comprendre : « *Le discours intérieur est abrégé, décousu, fragmenté, incompréhensible pour un auditeur extérieur.* ». Il ajoute que des mots importants ou des affixes grammaticaux peuvent être omis. Selon d'autres chercheurs en sciences humaines —

psychologues, psycholinguistes ou linguistes — la forme phonologique des mots produits en parole intérieure peut être incomplète (Dell & Repka, 1992 ; Sokolov, 1972). Dans cette lignée, Bergounioux (2001, p.120), précise que « *l'endophasie ne semble différer de la parole explicite ni par sa grammaire, ni par son lexique, à la réserve d'un emploi généralisé de l'asyndète et de l'anaphore, et d'une surreprésentation de la prédicativité* ». La parole intérieure est même considérée par certains chercheurs comme une abstraction (Oppenheim & Dell, 2008). Bergounioux (2001) précise que, d'un point de vue phénoménologique, l'endophasie, correspond à la parole sans signal (« *Endophasia, phenomenologically speaking, is speech without a signal* »). Cette caractérisation abstraite de la parole intérieure peut être considérée comme s'inscrivant dans la lignée de la conception chomskyenne du langage, selon laquelle des représentations profondes, abstraites, universelles, sont sous-jacentes aux formes de surface exprimées oralement (Chomsky 1957, 1965 ; cf. aussi Pinker, 1989). Elle est également cohérente avec l'hypothèse de l'existence d'un « mentalais » universel, supposé sous-tendre les processus mentaux et les pensées complexes (Fodor, 1975).

En opposition à cette conception « abstraite » ou mentaliste de la parole intérieure, d'autres chercheurs, dans une démarche que l'on peut rapprocher de celle du programme « *embodied cognition* » ou « cognition incarnée », ont cherché à recueillir des signaux concrets émanant de la parole intérieure. Une des premières formulations de la parole intérieure comme un processus impliquant des actions motrices remonte aux travaux de Watson au début du XX^{ème} siècle. Watson décrivait la parole intérieure (qu'il désignait par le terme de « langage implicite ») comme une forme affaiblie de parole à voix haute, impliquant des actions ténues, infimes, « abrégées, court-circuitées et économisées » (Watson, 1919, p. 327). Ainsi, Watson considérait que la parole intérieure impliquait des mouvements articulatoires infimes. Cette hypothèse extrême, qui inclut la présence de mouvements articulatoires pendant la pensée verbale, a cependant été réfutée. Smith *et al.* (1947) ont administré du curare à un participant pour induire une paralysie orofaciale temporaire. Alors que le participant était devenu complètement incapable d'articuler, il était resté capable de comprendre les questions qu'on lui posait et de répondre par des contractions du sourcil, ou bien de se remémorer les questions après la séance. Cette expérience montre donc que la pensée verbale et la parole intérieure peuvent avoir lieu sans articulation et que l'hypothèse extrême de Watson n'est pas tenable. Selon une conception plus nuancée, la parole intérieure ne doit pas être considérée comme de la parole à voix haute affaiblie, mais comme une simulation mentale de la production de parole à voix haute. Selon cette conception physicaliste ou incarnée, la production de parole intérieure est considérée comme similaire à la production de parole à voix haute, le processus d'exécution motrice étant bloqué, l'articulation inhibée et la parole restant silencieuse (Grèzes & Decety, 2001 ; Postma & Noordanus, 1996). Ainsi un continuum existerait entre la parole à voix haute et la parole intérieure, en lien avec le continuum décrit par Decety et Jeannerod (1996) pour les actions imaginées et effectives. Cette hypothèse dite de la « Simulation Motrice » a conduit certains auteurs à considérer que la parole intérieure, par essence, partage certaines des caractéristiques des actions motrices de parole (Feinberg, 1978 ; Jones & Fernyhough, 2007). Elle est confortée par un certain nombre d'observations empiriques.

En premier lieu, la parole à voix haute et la parole intérieure ont des corrélats acoustiques subjectifs similaires : la qualité sonore, le rythme, le tempo et la hauteur (grave ou aigu) du son sont rapportés comme similaires en parole intérieure et en parole à voix haute (Mackay, 1992). On peut s'en convaincre aisément en lisant mentalement les chiffres d'un numéro de téléphone. On entend alors intérieurement les pauses, les regroupements par nombre, les montées et descentes mélodiques typiques de la parole à voix haute. Ceci suggère que la parole intérieure génère, tout comme la parole à voix haute, des sensations rythmiques et prosodiques.

En second lieu, la parole à voix haute et la parole intérieure sont associées à des corrélats physiologiques comparables : des mesures objectives de taux de respiration, de débit d'élocution et d'activité musculaire suggèrent en effet que la parole intérieure est planifiée physiquement, de façon semblable à la parole à voix haute. En ce qui concerne le taux de respiration, Conrad et Schönle (1979) ont examiné la respiration durant le repos, la parole intérieure (sans articulation ni vocalisation), la parole subvocale (avec des mouvements articulatoires infimes mais sans production sonore) et la parole à voix haute. Ils ont montré que le cycle respiratoire varie selon un continuum, du repos jusqu'à la parole à voix haute, en passant par la parole intérieure puis la parole subvocale. Cette transition continue se manifeste par une augmentation progressive de la durée

de la phase expiratoire. Alors que pour le repos, le cycle respiratoire est symétrique, avec des phases d'inspiration et d'expiration de durées similaires, il est fortement asymétrique pour la parole à voix haute, avec une phase d'inspiration courte et une phase d'expiration longue permettant l'émission de parole. Conrad et Schönle ont montré que la parole intérieure est elle aussi caractérisée par une durée de la phase expiratoire plus longue que celle de la phase d'inspiration. Ils en concluent que cette modification du cycle respiratoire entre le repos et la parole intérieure suggère que les processus moteurs de la parole interviendraient lors de la production intérieure. En ce qui concerne le débit d'élocution, Netsell *et al.* (2010), par exemple, ont examiné la production de phrases en parole à voix haute et en parole intérieure chez des locuteurs typiques. Les participants marquaient le début et la fin de leur énonciation en appuyant sur un bouton. Netsell *et al.* ont ainsi montré que dans ce type de tâche, les débits d'élocution en voix haute ou intérieure ne sont pas différents. Enfin, en ce qui concerne l'activité musculaire orofaciale, il a été suggéré dès la fin du XIX^{ème} siècle que la parole intérieure pourrait être associée à une contraction infime des muscles des lèvres. On peut faire soi-même l'expérience de cette activité labiale en effectuant l'exercice proposé par Stricker (1885, chapitre II) : *Ouvrez vos lèvres pour qu'elles prennent la forme requise pour prononcer le son « o » et essayez d'imaginer que vous êtes en train de prononcer le son « m »*. Dans la posture labiale ouverte et arrondie requise pour prononcer « o », le fait de prononcer mentalement le son « m » fait imperceptiblement resserrer les lèvres et l'on sent une légère contraction. Ainsi, selon Stricker (p. 21) : « I. A la représentation de chaque son oral se rattache inséparablement un sentiment (plus ou moins distinct) dans les organes articulatoires. II. Ces sentiments ont leur siège dans les muscles. III. Ces sentiments sont semblables à ceux qui président à la prononciation réelle des sons articulés. » Ces « sentiments musculaires » qui interviendraient lors de la « pensée des sons articulés » (la parole intérieure), ont été mis en évidence par quelques rares enregistrements électromyographiques lors de tâches contrôlées impliquant la parole intérieure. Les signaux électromyographiques (EMG) sont des potentiels électriques associés aux contractions musculaires. Ils prennent leur origine dans le cortex moteur qui génère des potentiels d'action qui transitent le long des cellules nerveuses et sont transmis aux fibres musculaires, résultant en une contraction du muscle. Lors de la parole à voix haute, on peut enregistrer une activité EMG dans les muscles orofaciaux qui contrôlent les mouvements articulatoires, notamment dans les lèvres. Certains chercheurs sont parvenus à mettre en évidence une activité EMG orofaciale dans des situations mettant en jeu la parole intérieure. Une activité EMG a ainsi été détectée dans les muscles liés à la production de parole lors de tâches d'imagerie verbale ou de répétition silencieuse (Jacobson, 1931 ; Sokolov *et al.*, 1972). McGuigan et Dollins (1989) ont recueilli des données EMG qui montrent que les muscles des lèvres sont activés lorsque les participants lisent silencieusement la lettre « P » (qui, à voix haute, recrute les lèvres), mais pas la lettre « T » (qui, à voix haute, recrute la langue). De façon réciproque, les muscles de la langue sont seulement activés pour la lettre « T ». Livesay *et al.* (1996) ont obtenu une activité EMG dans les lèvres lors d'une tâche de récitation mentale, mais pas lors d'une tâche de visualisation (non verbalisée). Les résultats d'une étude sur les rêves suggèrent que la parole intérieure qui survient dans les rêves est accompagnée d'une activité EMG dans les muscles de la parole (Shimizu & Inoue, 1986). Il a également été montré qu'une activité EMG dans le muscle orbiculaire de la bouche (*orbicularis oris*) inférieur peut être détectée lors des hallucinations auditives verbales chez des patients souffrant de schizophrénie (Rapin, Dohen, Polosan, Perrier, & Lœvenbruck, 2013).

En dernier lieu, la parole à voix haute et la parole intérieure présentent également des corrélats cérébraux comparables. De façon générale, la parole intérieure et la parole à voix haute activent toutes deux les aires essentielles du langage, dans l'hémisphère gauche, *i.e.* les aires de Broca et de Wernicke ainsi que le lobule pariétal inférieur (Baciu *et al.*, 1999 ; Basho *et al.*, 2007 ; Bookheimer *et al.*, 1995 ; Huang *et al.*, 2002 ; McGuire *et al.*, 1996 ; Palmer *et al.*, 2001 ; Partovi *et al.*, 2012 ; Shergill *et al.*, 2001). Des différences existent cependant. Un certain nombre d'études IRMf obtient plus d'activation du cortex moteur primaire pour la parole à voix haute que la parole intérieure (e.g. Palmer, 2001), ce qui conforte l'hypothèse de la *Simulation Motrice*, selon laquelle la parole intérieure est considérée comme de la parole à voix haute avec un blocage de l'exécution motrice. De plus, des activations plus importantes pour la parole à voix haute ont également été observées dans l'insula et dans les régions sensorielles, auditives notamment (Bookheimer *et al.*, 1995 ; Huang *et al.*, 2002 ; Palmer *et al.*, 2001 ; Shuster & Lemieux, 2005 ; Owen *et al.*, 2004). La plus forte activation de l'insula peut être reliée au rôle de cette région dans la planification articulatoire (Dronkers, 1996), cruciale en voix haute, et probablement moins nécessaire en parole intérieure. L'activation plus importante des aires auditives reflète le fait que la production

de parole à voix haute est associée à la perception de notre propre parole. En effet, on s'entend parler à voix haute et notre système auditif est plus stimulé que dans la parole intérieure (au cours de laquelle on a aussi des sensations auditives, mais l'absence de son rend inutile le traitement complet du signal sonore depuis l'oreille jusqu'au cortex auditif). De façon réciproque, certaines études IRMf ont montré plus d'activation dans la parole intérieure que dans la parole à voix haute dans un certain nombre de régions cérébrales, notamment le gyrus frontal supérieur rostral droit, le gyrus temporal moyen gauche, le gyrus cingulaire droit, les lobules pariétaux inférieurs et supérieurs droits (Basho *et al.*, 2007 ; Owen *et al.*, 2004 ; Shuster & Lemieux, 2005). Ces observations suggèrent que la parole intérieure ne peut être simplement considérée comme la parole à voix haute à laquelle on aurait soustrait l'exécution motrice. Certains résultats peuvent s'expliquer par une meilleure qualité du signal dans les conditions de parole intérieure, due à l'absence de mouvements de la tête, l'articulation à voix haute entraînant de faibles mouvements irrépressibles. De façon plus intéressante, les activations dans les régions listées ci-dessus peuvent aussi être expliquées par des processus d'inhibition de la réponse à voix haute (lorsqu'on doit produire un mot mais ne pas le dire à voix haute).

En résumé, l'ensemble de ces observations suggère que la parole intérieure et la parole à voix haute partagent des corrélats subjectifs, physiologiques et neuraux et corrobore l'hypothèse selon laquelle la parole intérieure serait une simulation motrice de la parole à voix haute.

L'hypothèse de la *Simulation Motrice* a cependant été remise en question par certains résultats expérimentaux issus de la psycholinguistique. L'analyse des erreurs de production de parole peut renseigner sur la façon dont nous planifions et coordonnons notre articulation. Le fait que des erreurs surviennent également en parole intérieure nous indique que des processus de planification similaires sont en jeu. Oppenheim et Dell (2010) ont ainsi comparé les types d'erreur observés en parole intérieure et en parole à voix haute lors de la répétition de virelangues. Leur analyse des erreurs commises se fonde sur le modèle de production de la parole de Dell (1986). Dans ce modèle, lors de l'énonciation d'un mot, la première phase consiste en la récupération du concept, au niveau sémantique (par exemple « un petit objet rond avec lequel on joue au tennis »). Dans la deuxième étape, au niveau lexical, on accède au mot (dans notre exemple : « balle »). Puis, au niveau phonologique, les phonèmes sont spécifiés, *i.e.* les sons qui composent ce mot (/b/, /a/ et /l/). Enfin dans une dernière étape, au niveau articulatoire, sont choisis les traits ou les gestes articulatoires à produire pour cet enchaînement de sons (geste labial voisé et occlusif pour /b/, geste d'abaissement de la mandibule et de la langue pour /a/, geste de contact lingual alvéolaire avec écoulement d'air latéral pour /l/). Selon Dell *et al.*, la parole intérieure peut être vue comme de la parole à voix haute tronquée, n'impliquant que la première partie des étapes impliquées dans la parole à voix haute. Leur argument se fonde sur les biais observés lors des erreurs de production dans les deux types de parole. Les psycholinguistes ont en effet montré que les erreurs de production à voix haute comportent un biais lexical (*e.g.* Baars *et al.*, 1975), les erreurs correspondant plus souvent à des substitutions vers des mots de notre lexique (ainsi « balle » pourra être substitué par « dalle », « salle » ou « malle » plus que par « nalle » qui n'est pas un mot). Ce biais est expliqué par Dell et ses collègues par des activations réciproques ascendantes (*bottom-up feedbacks*) entre les niveaux phonologique et lexical. Ainsi, lors de la répétition de virelangues (tels que « balle datte balle datte »), lorsqu'au niveau phonologique un phonème est activé trop précocement, le signal de *feedback* remontant vers le niveau lexical peut activer le mot du lexique qui contient ce phonème (« dalle »). Ce biais lexical a également été observé en parole intérieure dans plusieurs études (notamment Oppenheim & Dell, 2008 et Corley *et al.*, 2011). Il paraît donc vraisemblable que la parole intérieure requiert les étapes lexicales et phonologiques, et pas uniquement l'étape conceptuelle.

La question qui reste ouverte est celle de savoir si la parole intérieure recrute également la dernière étape, qui inclut des spécifications articulatoires. Or les erreurs en production de parole à voix haute présentent, en plus du biais lexical, un biais de similarité phonémique (ou biais articulatoire), c'est-à-dire une tendance à échanger des phonèmes présentant des traits articulatoires communs : on observera plus fréquemment des erreurs de « balle » vers « malle » que de « balle » vers « salle », les deux consonnes initiales de la première paire partageant les traits de labialité et de voisement alors que les deux consonnes de la deuxième paire sont très différentes articulatoirement. Oppenheim et Dell (2008) n'ayant pas observé de biais articulatoire en parole intérieure, ils en ont conclu que la parole intérieure n'est pas spécifiée au niveau articulatoire et qu'elle reste de nature abstraite. Toutefois, dans des expériences similaires, Corley *et al.* (2011) obtiennent bien ce biais, même en

parole intérieure, et suggèrent que la parole intérieure requiert bien des spécifications articulatoires, en accord avec l'hypothèse de la *Simulation Motrice*. Néanmoins, d'autres observations peuvent encore remettre en cause cette hypothèse. Les expériences de Netsell *et al.* (2010) sur le débit d'élocution, citées plus haut, ont également été conduites chez des personnes qui bégaièrent. Leurs résultats montrent que les personnes qui bégaièrent ont un débit d'élocution significativement plus lent en parole à voix haute qu'en voix intérieure. De plus, parmi les sept personnes bègues enregistrées, six ont rapporté que leur parole intérieure est complètement fluide. Ceci peut être interprété en faveur de l'hypothèse de l'*Abstraction*, selon laquelle la parole intérieure ne nécessite pas une spécification intégrale du niveau articulatoire, ce qui expliquerait pourquoi elle n'est pas perturbée chez les personnes qui bégaièrent. De plus, l'hypothèse de la *Simulation Motrice* implique que la parole intérieure partagerait un grand nombre de processus avec la parole à voix haute, sans aller jusqu'au processus d'exécution motrice, et que par conséquent la parole à voix haute engloberait la parole intérieure. Certaines études de patients cérébro-lésés semblent corroborer l'hypothèse que parole à voix haute et parole intérieure partagent certains processus. En effet, un grand nombre d'étude de patients souffrant d'aphasie de Broca (région impliquée, entre autres, dans la production de parole) a montré que le déficit de parole à voix haute s'accompagne d'une perturbation de la parole intérieure (Levine, Calvanio & Popovics, 1982) ou que la parole intérieure reste intacte (Baddeley & Wilson, 1985 ; Vallar & Cappa, 1987), ce qui signifie bien que seules les étapes ultimes de production articulatoire sont endommagées. Toutefois, certaines études semblent contradictoires et ont rapporté des cas de patients présentant des difficultés à effectuer des jugements de rime en parole intérieure sans déficit dans les mêmes tâches en parole à voix haute. Ces cas ont été interprétés comme le fait que la parole intérieure ne peut être décrite comme de la parole à voix haute sans la composante motrice (Geva *et al.*, 2011 ; Martin *et al.*, 2014). Toutefois, nous proposons une explication alternative. Les tâches de jugement de rimes s'appuient sur une image auditive des mots à évaluer et impliquent le cortex auditif (cf. par exemple Paulesu, Frith & Frackowiak, 1993). La parole à voix haute fournit une trace acoustique qui stimule le cortex auditif et permet la comparaison sonore des mots pour juger s'ils riment. Dans le cas de la parole intérieure, cette trace acoustique n'est que simulée mentalement, avec une saillance moindre et une forme auditive plus évanescence que lors de la parole à voix haute, comme le suggère la plus faible activité du cortex auditif lors de la production de parole intérieure comparée à la parole à voix haute (Perrone-Bertolotti *et al.*, 2014). Ainsi, le déficit observé ne serait pas la preuve d'une dissociation entre les processus de parole intérieure et de parole à voix haute mais indiquerait simplement une plus faible saillance auditive des sensations acoustiques évoquées lors de la parole intérieure.

Ainsi, pour résumer, la question reste ouverte : la parole intérieure semble bien correspondre à de la parole à voix haute tronquée ou inhibée mais le niveau où a lieu la troncation fait encore l'objet de débats. Il reste à déterminer si le niveau articulatoire, et donc le système moteur, est systématiquement impliqué lors de la production de parole intérieure. Deux conceptions de la parole intérieure s'opposent. Selon l'hypothèse de la *Simulation Motrice*, la parole intérieure est incarnée. C'est une action simulée, elle implique le système moteur de la parole. Les mouvements articulatoires sont spécifiés mais inhibés. Selon l'hypothèse de l'*Abstraction*, la parole intérieure est de nature abstraite, intégrant éventuellement des représentations lexicales et phonologiques, mais ne requérant pas le niveau articulatoire, *i.e.* ne recrutant pas le système moteur.

DEUX PROPOSITIONS DE TEST EXPERIMENTAL DE L'HYPOTHESE DE LA SIMULATION MOTRICE

Afin d'éclairer le débat entre les conceptions incarnée et abstraite de la parole intérieure, nous présentons ici brièvement l'étude empirique détaillée dans Nalborczyk *et al.* (2017), dont le but était de tester l'hypothèse de l'implication du système moteur orofacial dans un cas de parole intérieure fréquent et récurrent : la rumination mentale.

Une prédiction de l'hypothèse de la *Simulation Motrice* est que les ruminations, en tant que cas particulier de la parole intérieure, requièrent l'activation du système moteur orofacial. Les ruminations devraient donc être accompagnées d'une activité dans les muscles faciaux liés à la production de parole, tels que les muscles labiaux. Elles devraient également être accompagnées d'activité dans les muscles liés aux émotions négatives, tels que les muscles du front (Jäncke *et al.*, 1996). Elles ne devraient pas cependant impliquer les muscles non-faciaux, tels que les muscles du bras. De plus, si les ruminations requièrent le système moteur, alors, par contraposée,

bloquer le système moteur devrait interférer avec la production de rumination. Une prédiction de l'hypothèse de la *Simulation Motrice* serait donc que la relaxation des muscles orofaciaux, réduisant l'implication du système moteur, devrait faire diminuer les ruminations mentales de façon plus importante que la relaxation des muscles du bras ou que l'écoute d'une histoire distrayante.

L'hypothèse de l'*Abstraction*, quant à elle, prédit que les ruminations devraient être associées à une activité dans les muscles faciaux liés aux émotions, tels que les muscles du front, mais ne devraient pas requérir une activité des muscles liées à la parole, tels que les muscles labiaux, ni des muscles non faciaux, tels que les muscles du bras. De plus, la relaxation centrée sur les muscles orofaciaux ne devraient pas avoir un effet d'interférence supérieur à celle centrée sur les muscles du bras.

Pour tester ces prédictions, notre étude comportait deux parties. La première partie visait à examiner si les ruminations mentales induites sont associées ou non à une activation motrice labiale. Les activités EMG de différents muscles ont été enregistrées : les muscles *orbicularis oris* supérieur et inférieur (muscles labiaux), le muscle *frontalis* (muscle du front) et le muscle *flexor carpi radialis* (muscle de l'avant-bras). Ainsi si l'hypothèse de la *Simulation Motrice* est correcte, les ruminations devraient être associées à une activité musculaire des lèvres et du front plus importante que l'état de repos. Si l'hypothèse de l'*Abstraction* est correcte, alors seule l'activité du muscle du front devrait augmenter lors des ruminations.

La deuxième partie visait à examiner l'effet de la réduction de l'activité motrice orofaciale par relaxation sur les ruminations. Trois types de relaxation ont été évalués : la relaxation orofaciale, centrée sur les muscles du visage et en particulier des lèvres et du front, la relaxation brachiale, centrée sur les muscles des bras et la relaxation par l'écoute d'une histoire distrayante. Ainsi, si l'hypothèse de la *Simulation Motrice* est correcte, relaxer les muscles impliqués dans la production de parole devrait diminuer l'intensité des ruminations de façon plus importante que la relaxation brachiale ou l'écoute d'une histoire. Si l'hypothèse de l'*Abstraction* est correcte, alors les deux relaxations orofaciale et brachiale devraient détendre les participants et diminuer les ruminations. L'écoute d'une histoire devrait également contribuer à diminuer les ruminations, mais de façon moindre.

PREMIER TEST :

MESURE DE L'ACTIVITE OROFACIALE PENDANT LES RUMINATIONS MENTALES

Le premier test visait à examiner l'activité des muscles orofaciaux pendant les ruminations mentales. Pour ce faire, nous avons provoqué des ruminations chez des participants sains, en utilisant une technique classique d'induction. Il a été choisi de ne sélectionner que des femmes, des études antérieures ayant montré que celles-ci sont plus susceptibles de ruminer que les hommes (Nolen-Hoeksema, 1987, 1991). Soixante-douze étudiantes ont été sélectionnées au sein de l'Université Grenoble Alpes. Elles étaient de langue maternelle française, d'âge moyen 20,58 ans (écart-type : 4,99) et ont rapporté ne pas présenter de troubles neurologique, neuromusculaire, cardiaque ou psychiatrique. L'absence de symptômes dépressifs a été vérifiée en utilisant la version française de l'échelle de dépression *Center for Epidemiologic Studies Depression scale* (CES-D ; Fuhrer & Rouillon, 1989) qui permet d'évaluer le niveau de symptômes dépressifs dans la population générale. Le but de l'expérience n'était pas connu des participantes. Elles étaient recrutées sous le prétexte de contribuer au développement d'un nouveau test de quotient intellectuel (QI) plus sensible aux traits de personnalité et à l'humeur. Toutes les participantes ont donné par écrit leur consentement pour participer à l'étude, cette dernière ayant été validée au préalable par le comité éthique local (CERNI, N° 2015-03-03-61).

Pour mesurer l'activité des muscles orofaciaux, nous avons utilisé des capteurs EMG de surface bipolaires, placés sur le visage et le bras des participantes. Ces capteurs étaient positionnés au-dessus de la lèvre supérieure pour recueillir l'activité du muscle *orbicularis oris* supérieur (noté *OOS*), en-dessous de la lèvre inférieure pour l'orbiculaire inférieur (*OOI*), sur le front pour le muscle *frontalis* (*Front*) et à l'intérieur de l'avant-bras non-dominant pour le *flexor carpi radialis* (*Bras*). Ce dernier capteur servait de contrôle pour s'assurer que les signaux recueillis à la surface des muscles étaient spécifiquement associés à une activité de parole ou un activité faciale et non pas à une tension générale de tout le corps.

Avant la procédure d'induction de rumination, des mesures de ligne de base ont été effectuées. Pour s'assurer que toutes les participantes étaient dans un état émotionnel similaire, une séance de relaxation leur a été proposée avant de commencer. Cette séance n'était centrée sur aucun muscle en particulier et visait simplement à détendre les participantes. Le but était simplement de minimiser l'influence de l'état thymique initial, intrinsèque à chaque participante et ainsi, de minimiser la variabilité inter-individuelle. Après cette séance, il était demandé aux participantes de ne pas bouger ni parler pendant une minute, le temps d'effectuer des mesures EMG au repos. Il leur a été ensuite demandé de décrire leur état initial de rumination en évaluant sur une échelle de 0 à 100 la pertinence de la phrase : « En ce moment je ressasse des choses négatives ».

Ensuite, les participantes étaient invitées à répondre au test de QI. Le test était d'une durée d'environ 15 minutes et comportait des questions de logique, de mathématiques et de raisonnement spatial. La plupart des questions était difficile voire impossible à résoudre dans le temps imparti (30 secondes). Cette tâche est inspirée des tâches de *forced failure* (e.g., LeMoult & Joorman, 2014) utilisant la mise en échec dans le but de mettre les participants dans une humeur légèrement négative, propice à l'induction de ruminations. Les participantes étaient ensuite invitées à réfléchir aux causes et conséquences de leurs sentiments, pendant cinq minutes. Cette méthode est inspirée du paradigme classique d'induction de ruminations proposé par Nolen-Hoeksema et Morrow (1993). Pendant ce laps de temps, de nouvelles mesures EMG étaient réalisées, suivies d'une réponse à l'échelle de rumination décrite ci-dessus.

Les signaux EMG enregistrés chez les 72 participantes ont été filtrés et les artefacts de mouvement éventuels ont été supprimés, en suivant une procédure de pré-traitement décrite dans Rapin *et al.* (2013). Pour chaque signal EMG ainsi pré-traité, nous avons calculé par fenêtres consécutives de 200ms la moyenne sur chaque fenêtre et retenu la valeur moyenne de ces moyennes sur nos sessions d'intérêt. Cela nous a permis d'obtenir un score pour chaque muscle (OOS, OOI, Front, Bras) pour chaque session (Repos, Post-Induction) et pour chaque participante. Après suppression des données bruitées ou présentant des artefacts, il restait des données fiables chez 71 participantes.

L'analyse de l'évolution du score sur l'échelle de rumination montre que la procédure d'induction était satisfaisante puisque le score après induction a augmenté de façon importante et significative par rapport au repos : les participantes ont ainsi rapporté plus de ruminations après l'induction qu'avant.

La comparaison des scores EMG avant et après induction de ruminations montre une augmentation de l'activité des deux muscles faciaux liés à la parole (OOI, OOS) ainsi que du muscle du front, lié aux émotions négatives. L'activité EMG du bras quant à elle est restée constante et n'a pas augmenté après induction.

Ces résultats sont globalement en accord avec l'hypothèse de la *Simulation Motrice*. En effet, tout d'abord, le fait que l'activité du bras soit restée stable après l'induction de rumination suggère que l'augmentation d'activité observée dans les muscles faciaux n'est pas simplement due à une augmentation générale de la tension du corps qui aurait été induite par un état mental négatif. Ensuite, l'augmentation de l'activité EMG labiale, sans augmentation de l'activité EMG du bras, mais bien associée à une augmentation du score subjectif de rumination suggère la présence d'une activité labiale spécifique lors des ruminations mentales. Les muscles des lèvres étant impliqués dans la production de parole à voix haute, cette augmentation de l'activité labiale lors de ruminations suggère que le système moteur de la parole est recruté dans la phase ruminative. Des études antérieures ont montré que le muscle du front est activé lors d'émotions négatives (Jäncke *et al.*, 1996). L'augmentation de l'activité du front enregistrée pendant la phase de rumination s'explique par la valence négative des ruminations induites par notre procédure. Ainsi, ces résultats EMG faciaux semblent bien aller dans le sens de l'hypothèse selon laquelle les ruminations sont un cas de parole intérieure incluant toutes les étapes de la planification de la parole, jusqu'au niveau articulatoire, puisque les muscles labiaux sont eux-mêmes recrutés.

SECOND TEST :

EFFET DE LA RELAXATION OROFACIALE SUR LES RUMINATIONS MENTALES

Le second test visait à examiner l'effet de la réduction de l'activité motrice orofaciale sur les ruminations. Trois types de relaxation ont été évalués : la relaxation orofaciale, centrée sur les muscles du visage et en particulier

des lèvres et du front, la relaxation brachiale, centrée sur les muscles des bras et la relaxation par l'écoute d'une histoire distrayante.

Les mêmes participantes que celle du test précédent ont été observées. Après la procédure d'induction de rumination décrite ci-dessus, les participantes étaient réparties aléatoirement dans 3 groupes expérimentaux. Ces trois groupes différaient par le type de relaxation proposé. La séance pouvait ainsi consister en une séance de relaxation spécifique des muscles orofaciaux (Groupe « Relax Orofaciale »), en une relaxation des muscles du bras (Groupe « Relax Bras »), ou en une condition de distraction dans laquelle un conte d'une durée sensiblement égale était diffusé (Groupe « Conte »). Les 3 séances avaient été enregistrées au préalable par une sophrologue et étaient diffusées par haut-parleur. De nouvelles mesures étaient ensuite réalisées (EMG et échelle de rumination) afin d'évaluer et comparer les effets des différents types de relaxation.

Les données EMG de 71 participantes ont été pré-traitées et analysées de la même façon que pour le premier test, décrit ci-dessus.

L'analyse de l'évolution du score de rumination montre que les trois séances de relaxation sont bénéfiques puisque le score après relaxation diminue de façon importante et significative par rapport à l'état précédent (post-induction) : les participantes ont ainsi rapporté moins de ruminations après la relaxation qu'avant. De façon intéressante, la diminution la plus importante est observée dans le groupe de participantes ayant bénéficié d'une relaxation orofaciale, i.e. centrée sur la détente des muscles du visage.

La comparaison des scores EMG avant et après relaxation montre des résultats différents selon les muscles. Ainsi pour les trois muscles faciaux, on observe de façon générale une diminution de l'activité EMG après relaxation orofaciale, alors que l'activité du bras reste stable. Pour le muscle OOI, l'écoute d'un conte semble également diminuer l'activité EMG. Pour le muscle du front, la diminution de l'activité EMG est la plus importante dans les deux conditions de relaxation orofaciale et brachiale de façon semblable, et cette diminution est moindre pour l'écoute d'un conte. Le fait que l'activité du bras reste constante quelle que soit la condition de relaxation s'explique en rappelant que cette activité après induction (avant relaxation) était la même que celle au repos (cf. les résultats du premier test, présentés ci-dessus). L'activité du bras est donc restée stable tout au long de cette expérience en deux parties et ne semble pas jouer de rôle dans la production de rumination mentale.

La première interprétation de ces résultats est que la production verbale générée par l'induction de rumination a été plus atténuée chez le groupe de participantes ayant ensuite bénéficié de la relaxation orofaciale que chez les autres groupes, puisque les scores de rumination ont le plus diminué chez ce premier groupe. La diminution plus importante de l'activité des muscles labiaux OOS et OOI dans ce groupe reflèterait cette réduction de production verbale et serait donc en accord avec l'hypothèse de la *Simulation Motrice*. Le fait que l'activité du front diminue de façon similaire dans les deux types de relaxation orofaciale et brachiale suggère que la détente musculaire quelle qu'elle soit (centrée sur le visage ou non) permettrait de réduire les affects négatifs et pourrait ainsi diminuer la contraction musculaire du front.

Une deuxième interprétation est que la diminution plus importante d'activité labiale après la relaxation orofaciale s'explique de façon triviale par le fait qu'une relaxation centrée sur le visage devrait être plus efficace pour réduire la contraction musculaire labiale qu'une relaxation non-orofaciale. Ainsi, les effets différentiels des deux types de relaxation orofaciale et brachiale sur les muscles des lèvres ne seraient pas dus à une plus forte diminution des ruminations verbales dans le cas de la relaxation orofaciale, mais seraient simplement dus à une différence de cible anatomique de chacun des types de relaxation. Cependant, le fait que cette diminution plus importante de l'activité musculaire labiale lors de la relaxation orofaciale soit accompagnée d'une diminution de la production de rumination mentale suggère que la réduction d'activité labiale est bien liée à la réduction des ruminations. De plus, si l'anatomie expliquait seule ces résultats, la diminution comparable de l'activité du muscle du front dans les deux types de relaxation orofaciale et brachiale serait problématique. En effet, si la réduction de l'activité musculaire était simplement liée à la relaxation des muscles faciaux, alors la diminution de l'activité du muscle du front devrait elle aussi être plus importante dans la condition de relaxation orofaciale que dans les autres conditions de relaxation, ce qui n'est pas le cas.

Nous suggérons donc que la première interprétation est la plus plausible. Ces résultats semblent donc s'interpréter comme une diminution des activités des muscles labiaux OOS et OOI induite par la relaxation orofaciale et associée à une diminution des ruminations mentales. Le fait que la simple écoute d'un conte ait contribué à diminuer les activités des muscles OOI et Front (de façon moindre que la relaxation orofaciale cependant) mérite une explication. L'écoute d'un conte pourrait en effet avoir réduit les ruminations en déplaçant le focus attentionnel des participantes : alors que notre procédure d'induction de rumination les incitait à porter leur attention sur leurs problèmes, l'écoute d'un conte distrayant permet de centrer l'attention sur l'histoire et de sortir du processus ruminatif. Ceci suggère que des séances de relaxation corporelle pourraient contribuer à réduire les ruminations en recentrant l'attention sur le corps plutôt que sur les problèmes sources des ruminations.

DISCUSSION GENERALE

Le but de l'étude rapportée brièvement ici était d'examiner si les ruminations requièrent une simulation motrice, impliquant les muscles de la parole et sont ainsi une forme incarnée de parole intérieure ou si elles sont de nature abstraite, impliquant des représentations linguistiques appauvries sur le plan articulatoire. Nous avons enregistré des données EMG labiales, faciales et brachiales pour évaluer les corrélats articulatoires potentiels des ruminations. Nos résultats semblent dans l'ensemble en faveur d'une nature incarnée et motrice des ruminations mentales.

Dans le premier test, notre procédure d'induction de rumination a en effet augmenté les scores de rumination, ce qui confirme que les participantes étaient alors dans un état ruminatif négatif. Cet état était accompagné d'une augmentation de l'activité des deux muscles impliqués dans la parole (OOS et OOI) sans modification de l'activité du muscle du bras. Ceci semble bien indiquer que la rumination est associée à une activité spécifique des muscles articulatoires de la parole. L'augmentation concurrente de l'activité du muscle du front peut s'expliquer par une augmentation des émotions négatives, induite par notre procédure d'induction à valence négative. Les résultats du premier test semblent donc bien montrer une implication des muscles de la parole lors des ruminations mentales et semblent corroborer l'hypothèse de la *Simulation Motrice*, selon laquelle la parole intérieure est intégralement spécifiée, depuis le niveau conceptuel jusqu'au niveau articulatoire.

Dans le second test, les séances de relaxation musculaire suggestive ont conduit à une diminution de l'activité des muscles de parole, la diminution étant plus forte après la relaxation orofaciale qu'après la relaxation brachiale. Pour le muscle du front, l'effet a été le même pour les deux types de relaxation musculaire (orofaciale et brachiale). Cette plus forte diminution de l'activité des muscles de la parole lors de la relaxation orofaciale a été associée à une plus forte diminution des scores de rumination auto-rapportée. Ces observations suggèrent qu'une réduction de l'activité des muscles de la parole pourrait entraver la simulation articulatoire et ainsi limiter la production de parole intérieure, entraînant une réduction des ruminations. Cette interprétation est cohérente avec l'hypothèse de la *Simulation Motrice*.

Les scores de rumination ont aussi diminué après la relaxation brachiale et l'écoute d'un conte, bien que de façon moins importante qu'après la relaxation orofaciale. Ceci suggère qu'une relaxation généralisée ou une distraction peuvent également diminuer les ruminations négatives.

En résumé, nos deux tests semblent bien corroborer l'hypothèse de la *Simulation Motrice*, selon laquelle l'activité des muscles de la parole est inhérente à la production de parole intérieure. Le premier test a montré l'implication de la musculature labiale durant les ruminations négatives. Le deuxième test suggère que le type de rumination induit par notre paradigme requiert l'activité des muscles de la parole et qu'il peut donc être réduit en bloquant ou en relaxant spécifiquement les muscles de la parole, tels que les muscles des lèvres.

Ces résultats doivent cependant être considérés en ayant en tête les limites de cette étude. Tout d'abord, la population étudiée n'était constituée que de femmes dans l'optique d'optimiser la procédure d'induction de rumination. Il se peut donc que nos observations ne soient pas applicables à une population masculine. Deuxièmement, dans le premier test, nous n'avons pas comparé nos résultats à ceux d'une population contrôle qui n'aurait pas subi l'induction de rumination. Il se peut donc que les résultats observés ne soient pas liés au phénomène de rumination mentale mais à d'autres processus cognitifs ou émotionnels qui auraient été induits

par ce paradigme. Troisièmement, des différences inter-individuelles notables ont été observées sur l'effet de l'induction de rumination sur l'activité EMG faciale, certaines participantes montrant peu d'augmentation du signal EMG entre l'état de repos initial et les ruminations. Ces différences de variation d'activité EMG peuvent s'expliquer par de plus ou moins grande tendance à montrer ou cacher ses émotions à l'expérimentateur. De semblables différences ont été observées dans l'étude de Jäncke *et al.* (1996). Enfin, bien que les ruminations se manifestent le plus souvent sous modalité verbale, elles peuvent parfois prendre la forme d'images visuelles. Nous ne pouvons exclure que certaines de nos participantes aient ruminé dans une modalité non verbale, ce qui expliquerait leur activité labiale plus faible que la moyenne. Nous envisageons donc par la suite de mieux décrire les modalités de rumination chez nos participants et de vérifier que nos résultats s'appliquent bien au cas des ruminations mentales verbales.

En conclusion, cette étude fournit de nouvelles données soutenant que le phénomène de rumination mentale, considéré comme un cas particulier de parole intérieure, est un processus incarné, requérant une simulation mentale des actions de parole. Même s'il est nécessaire de recueillir plus de données pour confirmer ces résultats préliminaires, nos observations semblent bien soutenir l'hypothèse de la *Simulation Motrice*, selon laquelle la parole intérieure correspond à de la parole à voix haute simulée articulatoirement et inhibée avant l'exécution motrice.

REMERCIEMENTS

Cette étude est financée par le projet ANR INNERSPEECH ANR-13-BSH2-0003-01 (<http://lpnc.univ-grenoble-alpes.fr/InnerSpeech>). Le deuxième auteur est titulaire d'une allocation de recherche doctorale de l'Université Grenoble Alpes et a bénéficié d'une bourse du *Pôle Grenoble Cognition*. Nous remercions Ernst Koster, Mircea Polosan, Lucile Rapin et Elsa Spinelli pour leurs conseils et leurs remarques constructives. Nous remercions Nathalie Vallet pour l'enregistrement des séances de relaxation. Nous remercions Marion Dohen et Maëva Garnier pour leur aide lors de l'enregistrement des stimuli audio dans la chambre sourde de GIPSA-lab, ainsi que Christophe Savariaux et Coriandre Vilain pour leurs conseils dans la synchronisation des données audio et EMG. Nous remercions également Marc Chanove et les membres de la MSH-Alpes de nous avoir permis d'utiliser la plateforme expérimentale *Screen*.

REFERENCES BIBLIOGRAPHIQUES

- Baars, B. J.; Motley, M. T. & MacKay, D. G. (1975). Output editing for lexical status in artificially elicited slips of the tongue. *Journal of verbal learning and verbal behavior*, 14, 382-391.
- Baciu, M. V., Rubin, C., Decorps, M. A., & Segebarth, C. M. (1999). fMRI assessment of hemispheric language dominance using a simple inner speech paradigm. *NMR in Biomedicine*, 12, 293-298.
- Baddeley, A., & Wilson, B. (1985). Phonological coding and short-term memory in patients without speech. *Journal of Memory and Language*; 24(4), 490-502.
- Basho, S., Palmer, E. D., Rubio, M. A., Wulfeck, B., & Müller, R. A. (2007). Effects of generation mode in fMRI adaptations of semantic fluency: paced production and overt speech. *Neuropsychologia*, 45(8), 1697-1706.
- Baeyens, C., Kornacka, M., Douilliez, C. (2012). De la dépression à une perspective transdiagnostique: Conceptualisation et interventions sur les ruminations mentales. *Revue francophone de clinique comportementale et cognitive*, 17, 66-85.
- Bergounioux, G. (2001). Endophasie et linguistique [Décomptes, quotes et squelette]. *Langue française*, N°132, *La parole intérieure*, sous la direction de Gabriel Bergounioux, 106-124.
- Bookheimer, S. Y., Zeffiro, T. A., Blaxton, T., Gaillard, W., & Theodore, W. H. (1995). Regional cerebral blood flow during object naming and word reading. *Human Brain Mapping*, 3, 93-106.
- Bottineau, D. (2014). Explorer l'iconicité des signifiants lexicaux et grammaticaux en langue française dans une perspective contrastive (anglais, arabe). *Le français moderne*, 82/2, *Formes de l'iconicité en langue française. Vers une linguistique analogique (Tome 2)*, Nobile, L. (Ed.), 243-270.
- Chomsky, N. (1965), *Aspects of the Theory of Syntax*, Cambridge, Mass.: The MIT Press
- Chomsky N. (1957). *Syntactic Structures*. Mouton.
- Clark, A. (2006). Language, embodiment, and the cognitive niche. *Trends in cognitive sciences*, 10, 370-374.

- Conrad, B., & Schönle, P. (1979). Speech and respiration. *Archiv für Psychiatrie und Nervenkrankheiten*, 226, 251–68.
- Corley, M., Brocklehurst, P. H., & Moat, H. S. (2011). Error biases in inner and overt speech: Evidence from tongue twisters. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37, 162–175.
- Davis, R. N., & Nolen-Hoeksema, S. (2000). Cognitive inflexibility among ruminators and nonruminators. *Cognitive Therapy and Research*, 24, 699-711.
- Decety, J., & Jeannerod, M. (1996). Mentally simulated movements in virtual reality: does Fitt's law hold in motor imagery? *Behavioral Brain Research*, 72, 127-134.
- Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review*, 93(3), 283-321.
- Dell, G. S., & Repka, R. J. (1992). *Errors in inner speech*. In *Experimental Slips and Human Error* (pp. 237-262). Springer US.
- Dostoïevski, F. (1866-1996). *Crime et Châtiment*. Traduction française de A. Markowicz, Ed. Actes Sud-Babel.
- Dronkers, N. F. (1996). A new brain region for coordinating speech articulation. *Nature*, 384(6605), 159–161.
- Egger, V. (1881). *La Parole intérieure. Essai de psychologie descriptive*. (G. Baillière, Ed.). Paris.
- Feinberg, I. (1978). Efference copy and corollary discharge: Implications for thinking and its disorders. *Schizophrenia Bulletin*, 4, 636–640.
- Fodor, J. A. (1875). *The language of thought*. Harvard University Press.
- Fuhrer, R., & Rouillon, F. (1989). La version française de l'échelle CES-D (Center for Epidemiologic Studies-Depression Scale). Description et traduction de l'échelle d'auto-évaluation. *European-Psychiatry*, 4(3), 163-166.
- Geva, S., Jones, P. S., Crinion, J. T., Price, C. J., Baron, J.-C., & Warburton, E. A. (2011). The neural correlates of inner speech defined by voxel-based lesion-symptom mapping. *Brain: A Journal of Neurology*, 134, 3071-3082.
- Grèzes, J., & Decety, J. (2001). Functional anatomy of execution, mental simulation, observation, and verb generation of actions: A meta-analysis. *Human Brain Mapping*, 12, 1-19.
- Huang, J., Carr, T. H., & Cao, Y. (2002). Comparing cortical activations for silent and overt speech using event-related fMRI. *Human Brain Mapping*, 15, 39-53.
- Jacobson, E. (1931). Electrical measurements of neuromuscular states during mental activities. V. Variation of specific muscles contracting during imagination. *American Journal of Physiology*, 96, 115-121.
- Jäncke, L., Vogt, J., Musial, F., Lutz, K., & Kalveram, K. T. (1996). Facial EMG responses to Auditory Stimuli. *International Journal of Psychophysiology*, 22, 85-96.
- Johnson, M. & Lakoff, G. (2002). Why cognitive linguistics requires embodied realism. *Cognitive Linguistics*, 13, 245-264.
- Jones, S. R., & Fernyhough, C. (2007). Thought as action: Inner speech, self-monitoring, and auditory verbal hallucinations. *Consciousness and Cognition*, 16, 391-399.
- LeMoult, J., & Joormann, J. (2014). Depressive rumination alters cortisol decline in Major Depressive Disorder. *Biological Psychology*, 100, 50–55.
- Levine, D. N., Calvanio, R., & Popovics, A. (1982). Language in the absence of inner speech. *Neuropsychologia*, 20(4), 391-409.
- Livesay, J., Liebke, A., Samaras, M., & Stanley, A. (1996). Covert speech behavior during a silent language recitation task. *Perception and Motor Skills*, 83, 1355–1362.
- MacKay, D. G. (1992). *Constraints of theories of inner speech*. In D. Reisberg (Ed.), *Auditory imagery* (pp. 121–142). Hillsdale, NJ: Erlbaum.
- Martin, L. L. & Tesser, A. (1996). Some ruminative thoughts. *Advances in Social Cognition*, 9, 1-47.
- Martin, S., Brunner, P., Holdgraf, C., Heinze, H.-J., Crone, N. E., Rieger, J., Schalk, G., Knight, R. T. & Pasley, B. N. (2014). Decoding spectrotemporal features of overt and covert speech from the human cortex. *Frontiers in Neuroengineering*, 7, 1-15.
- McGuigan, F. J., & Dollins, A. B. (1989). Patterns of covert speech behavior and phonetic coding. *Pavlovian Journal of Biological Science*, 24, 19–26.

- McGuire, P., Silbersweig, D., Murray, R., David, A., Frackowiak, R., & Frith, C. (1996). Functional anatomy of inner speech and auditory verbal imagery. *Psychological Medicine*, 26, 29–38.
- Merleau-Ponty, M. (1945). *Phénoménologie de la Perception*. Gallimard.
- Nalborczyk, L., Perrone-Bertolotti, M., Baeyens, C., Grandchamp, R., Polosan, M., Spinelli E., Koster E. & Løevenbruck, H. (2017). Orofacial electromyographic correlates of induced verbal rumination. *Biological Psychology*, 127, 53-63.
- Netsell, R., Ashley, E., & Bakker, K. (2010). The inner speech of persons who stutter. In: *Proceedings of the international motor speech conference*.
- Nolen-Hoeksema, S. (1987). Sex differences in unipolar depression: Evidence and theory. *Psychological Bulletin*, 101, 259–282.
- Nolen-Hoeksema, S. (1991). Responses to depression and their effects on the duration of depressive episodes. *Journal of Abnormal Psychology*, 100(4), 569–582.
- Nolen-Hoeksema, S., & Morrow, J. (1991). A prospective study of depression and posttraumatic stress symptoms after a natural disaster: the 1989 Loma Prieta Earthquake. *Journal of Personality and Social Psychology*, 61, 115–121.
- Oppenheim, G. M., & Dell, G. S. (2008). Inner speech slips exhibit lexical bias, but not the phonemic similarity effect. *Cognition*, 106, 528-537.
- Oppenheim, G. M., & Dell, G. S. (2010). Motor movement matters: the flexible abstractness of inner speech. *Memory & Cognition*, 38(8), 1147–1160.
- Owen, W. J., Borowsky, R., & Sarty, G. E. (2004). fMRI of two measures of phonological processing in visual word recognition: ecological validity matters. *Brain and Language*, 90(1), 40–46.
- Palmer, E. D., Rosen, H. J., Ojemann, J. G., Buckner, R. L., Kelley, W. M., & Petersen, S. E. (2001). An event-related fMRI study of overt and covert word stem completion. *Neuroimage*, 14(1), 182-193.
- Partovi, S., Konrad, F., Karimi, S., Rengier, F., Lyo, J. K., Zipp, L., ... & Stippich, C. (2012). Effects of covert and overt paradigms in clinical language fMRI. *Academic Radiology*, 19(5), 518-525.
- Shergill, S. S., Bullmore, E. T., Brammer, M. J., Williams, S. C., Murray, R. M., & McGuire, P. K. (2001). A functional study of auditory verbal imagery. *Psychological Medicine*, 31, 241-253.
- Shuster, L. I., & Lemieux, S. K. (2005). An fMRI investigation of covertly and overtly produced mono- and multisyllabic words. *Brain and Language*, 93, 20-31.
- Perrone-Bertolotti, M., Rapin, L., Lachaux, J. P., Baciú, M., & Løevenbruck, H. (2014). What is that little voice inside my head? Inner speech phenomenology, its role in cognitive performance, and its relation to self-monitoring. *Behavioural Brain Research*, 261, 220–39.
- Pinker, S. (1989). *Learnability and Cognition*, Cambridge, Mass.: The MIT Press.
- Postma, A., & Noordanus, C. (1996). The production and detection of speech errors in silent, mouthed, noise-masked, and normal auditory feedback speech. *Language and Speech*, 39, 375-392.
- Rapin, L., Dohen, M., Polosan, M., Perrier, P., & Løevenbruck, H. (2013). An EMG study of the lip muscles during covert auditory verbal hallucinations in schizophrenia. *Journal of Speech, Language, and Hearing Research*, 56(6), S1882–93.
- Rood, L., Roelofs, J., Bögels, S. M., & Arntz, A. (2012). The effects of experimentally induced rumination, positive reappraisal, acceptance, and distancing when thinking about a stressful event on affect states in adolescents. *Journal of Abnormal Child Psychology*, 40(1), 73-84.
- Shimizu, A., & Inoue, T. (1986). Dreamed speech and speech muscle activity. *Psychophysiology*, 23(2), 210-214.
- Smadja, S. (à paraître). *La Recherche en littérature : approches stylistiques du monologue intérieur*. Paris: Hermann.
- Smith, Scott; Brown, Hugh; Toman, James; Googman, L. (1947). The Lack of Cerebral Effects of d-Tubocurarine. *Anesthesiology*, 8(1), 1-14.
- Sokolov, A. N., Onischenko, G. T., & Lindsley, D. B. (1972). *Inner speech and thought*. (Plenum Press, Ed.). New York: Plenum Press.
- Stricker, S. (1885). *Du langage et de la musique*. (Traduit de l'allemand par Frédéric Schwiedland, Ed.). Paris: Alcan.
- Vallar, G., & Cappa, S. F. (1987). Articulation and verbal short-term memory: Evidence from anarthria. *Cognitive Neuropsychology*, 4(1), 55-77.

Version préliminaire produite par les auteurs.

In Lœvenbruck, H., Nalborczyk, L., Perrone-Bertolotti, M., Baeyens, C., Grandchamp, R. (2017). In Dokic, J. & Perrin, D. (eds.). *La cognition incarnée. Recherches sur la philosophie et le langage*, N°33, Vrin, 297-324, ISBN : 978-2-7116-8423-6 - **hal-01887636**

Vygotski L. S. (1934/1985). *Pensée et langage*. (Traduction de Françoise Sève, avant-propos de Lucien Sève, Collection « Terrains», Éditions Sociales, Paris, 1985).

Watson, J. B. (1919). *Psychology from the standpoint of a behaviorist*. (J. B. Lippincott, Ed.). Philadelphia.

Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin & Review*, 9, 625-636.

A cognitive neuroscience view of inner language: to predict and to hear, see, feel

Hélène Lœvenbruck¹, Romain Grandchamp¹, Lucile Rapin², Ladislav Nalborczyk^{1,3}, Marion Dohen⁴,
Pascal Perrier⁴, Monica Baciú¹ & Marcela Perrone-Bertolotti¹

1. Laboratoire de Psychologie et NeuroCognition, CNRS UMR 5105 & Université Grenoble Alpes, Grenoble, France

2. Douglas Mental Health University Institute, Department of Psychiatry, McGill University, Montreal, Canada

3. Department of Experimental Clinical and Health Psychology, Ghent University, Belgium

4. GIPSA-Lab, Département Parole et Cognition, CNRS UMR 5216 & Université Grenoble Alpes, Grenoble, France

ABSTRACT

The nature of inner language has long been under the scrutiny of humanities, through the practice of introspection. The use of experimental methods in cognitive neurosciences provides complementary insights. This chapter focuses on wilful expanded inner language, bearing in mind that other forms coexist. It first considers the abstract vs. concrete (or embodied) dimensions of inner language. In a second section, it argues that inner language should be considered as an action-perception phenomenon. In a third section, it proposes a revision of the « predictive control » account, fitting with our sensory-motor view. Inner language is considered as deriving from multisensory goals, generating multimodal acts (inner phonation, articulation, sign) with multisensory percepts (in the mind's ear, tact and eye). In the final section, it presents a landscape of the cerebral substrates of wilful inner verbalization, including multisensory and motor cortices as well as cognitive control networks.

Keywords: abstraction, simulation, embodiment, multisensory-motor, predictive control, agency

INTRODUCTION

Mental verbalization has long been under the scrutiny of writers, philosophers, literary scholars, psychoanalysts, psychologists and linguists, through the practice of thorough introspection, careful observation and reflection. Many terms have been used to describe it, including: inner language, inner speech, inner voice, covert speech, internal speech, silent speech, self-talk, internal monologue, internal dialogue, imagined speech, endophasia, private speech, verbal thought, subvocalisation, auditory imagery. The terms “inner speech” or “voice” are too restrictive, as mental verbalization is not always oral: consider deaf people who use sign language. We will therefore use the term inner language which captures the multimodal (auditory, somatosensory and visual) qualities of mental verbalisation.

The use of experimental methods and technology in neuroscience, psychology, psycholinguistics and psychiatry provides new insights into the nature of inner language. Inner language manifests in various ways. We often deliberately engage in inner language (e.g. when we count, make a list, schedule our objectives). This can be called “wilful/volitional inner language”. But sometimes, our internal monologue is less deliberate, and “more passive”. This latter form has been referred to as “verbal mind wandering” (Perrone-Bertolotti, Rapin, Lachaux, Baciú, & Lœvenbruck, 2014), and often occurs during “resting states” (mind wandering can also be non-verbal, as in visual imagery, hence the adjective “verbal”). Verbal mind wandering consists of flowing, spontaneous, stimulus-independent verbal thoughts. Whereas wilful inner language is an attention-demanding task, verbal mind wandering has been associated with the default mode network (Raichle, 2010), although it may also additionally activate executive regions (Christoff, Gordon, Smallwood, Smith & Schooler, 2009). Neural connectivity studies have shown that the attention and default mode networks fluctuate in an anticorrelated pattern (Ossandon *et al.*, 2011). Therefore, in addition to core language regions that they presumably share, these two modes of inner language may recruit distinct regions, related to attention vs. default mode networks.

Moreover, different levels of inner language have been identified, including condensed and expanded instances (Ferryhough, 2004).

In this chapter, we focus on the nature of wilful inner verbal production, in its expanded version, bearing in mind that other forms of inner language coexist. We first consider the abstract vs. concrete dimension of inner language. In a second section, we examine its sensory vs. motor dimension and argue that inner language should be considered as an action-perception phenomenon. We describe inner language as an act, spurring the mind's eye, ear, and tact. In a third section, we propose a revision of the "predictive control" account of inner speech, to fit with our sensory-motor view of inner language. In this integrated account, inner language is considered as deriving from multisensory goals, generating multimodal acts (inner phonation, articulation, sign) with multisensory percepts (in the mind's ear, tact and eye). In the final section, we present a landscape of the cerebral networks involved in wilful inner language production, including sensory and motor cortices as well as cognitive control networks.

1. THE ABSTRACT-CONCRETE DIMENSION OF INNER LANGUAGE

In many studies of language and cognition, an Abstraction view is taken in which inner language involves symbolic and abstract representations, divorced from bodily experience. Alternative approaches, such as the Motor Simulation view, posit that inner language is concrete and embodied, involving physical processes that unfold over time²⁹. These two views reflect different positions about internal processes, the first related to classical theories of mental architecture (Fodor & Pylyshyn, 1988; Newell & Simon, 1972) and the second, to the embodied cognition framework (Barsalou, 1999; Gallese & Lakoff, 2005; Pulvermüller & Fadiga, 2010).

1.1 Arguments for the abstractness and amodality of inner language

Introspective and psycholinguistic studies of inner language have led many scholars to view it as an abstraction, unconcerned with articulatory or auditory simulations. In the Abstraction view, inner speech is articulatorily impoverished and abstract (Oppenheim & Dell, 2010; Dell & Oppenheim, 2015). MacKay (1992, p.122) confidently stated that inner speech is amodal, i.e. nonarticulatory and nonauditory. According to him, articulatory movements 'are irrelevant to inner speech. Even the lowest level units for inner speech are highly abstract'.

A first argument in favour of the Abstraction view is condensation. Inner language is considered to be autonomous from perceptuo-motor processes and their operational details, condensing it, relative to overt speech, at different levels: articulation, phonology, lexicon and syntax. Its condensation would be manifest in the time course of its production, shorter than that of overt speech.

Introspective accounts of condensation are abundant. Although Egger (1881) provided many arguments for the embodied nature of inner speech, he was the first to clearly state why inner language may indeed be shorter. First, he listed physiological constraints. We cannot articulate overtly as quickly as covertly, the speed of our tongue movements being physiologically limited. Also, when we speak aloud, we need to take breath between speech fragments, as speech only occurs during expiration. Inner speech, not being subjected to these physiological constraints, can be accelerated. Secondly, Egger mentioned social constraints. In order to be understood, we need to articulate more clearly and slowly than in covert speech. Egger simply meant that the absence of physiological and social constraints shortens inner production. But drawing from similar durational observations, several psychologists have claimed that inner speech is even phonologically reduced, many phonemes being dropped and only the word-initial sounds being clearly produced (e.g. Vygotsky, 1934/1986). In this view, covert words lack the full phonological and articulatory specification they have overtly, making them more abstract and amodal. Furthermore, according to Egger, some of our mentally used expressions bear meanings that are explicit only to ourselves. To be understood by an addressee, we would need to supplement

²⁹ Abstract vs. concrete in the present paper relate to the format of the representation: symbolic and amodal vs. physical and modal. They do not refer to the semantic content of inner language, which may be abstract or concrete whatever its format.

them with contextual information. Therefore, condensation occurs not only at the phonological and articulatory levels, but also at the message level. Vygotsky (1934/1986) has further developed this notion of condensation. His theory is based on introspection, and on examination of children's private speech, in which children talk to themselves aloud, and which he claimed to be a precursor of adult inner speech (but see Perrone-Bertolotti *et al.*, 2014, for developmental data challenging this view). He asserted that important words or affixes may be dropped in inner language, the syntax of inner speech being "predicated". Bergounioux (2001) likewise claims that inner speech entails 'a generalised use of asyndeton, anaphora and an over-representation of predication' (p.120, our translation). Examples of such linguistic operations can be found in literary works associated with the "*monologue intérieur*" movement, initiated by Dujardin (1887, 1931; Smadja, in press). Hence, introspective observations have led to the speculation that inner language is impoverished, at the syntactic, lexical, phonological and articulation levels. Such condensation implies that modality-specific processes (e.g. articulatory planning) may be suppressed in inner language, making it abstract and amodal. Empirical evidence for the condensed quality of inner language has been searched for.

At the syntactic and lexical levels, evidence for condensation can be found in a study of the rate of spontaneous covert speech (Korba, 1990). Participants were asked to mentally solve short verbal problems. They reported the inner speech used to solve each problem, which gave an estimation of the number of elliptical words used. Then they delivered a full statement of their strategies, which provided an extended word count. The equivalent speaking rate of the extended statement exceeded 4000 words per minute, an unattainable rate in overt mode. These findings suggest that such inner verbalization is condensed at the syntactic and lexical levels. At the phonological level, the condensation hypothesis receives support from empirical studies showing that production is faster in covert mode, even when syntactic and lexical contents are kept equal, i.e. when participants are asked to recite the same sets of words in both modes (Anderson, 1982; MacKay, 1981; Marshall & Cartwright, 1978; Marshall & Cartwright, 1980). These studies could suggest that some of the phonological or articulatory processes involved in overt speech are absent in covert mode. An alternative interpretation, described in 1.2, is that inner speech involves the same operations as overt speech but that, as suggested by Egger (1881), the execution of articulator movements takes longer than their simulation.

A second argument for the Abstraction view is that inner speech would be deprived of some articulatory specification. Speech errors during inner recitation of tongue-twisters do display the lexical bias observed in overt production, but they do not show the phonemic similarity bias, which is based on articulatory representations (Oppenheim & Dell, 2008). This second bias is a tendency to exchange phonemes with common articulatory features (e.g. REEF slips more often to LEAF, with /r/ and /l/ sharing voicing and approximant features, than REEF to BEEF, with /r/ and /b/ only sharing voicing). Oppenheim and Dell (2008, 2010) argue that reciprocal activations between articulatory and phonological levels can explain this effect. They have only observed it in overt mode or with inner speech accompanied with mouthing, which has led them to claim that although inner speech is specified at the lexical level (because of the lexical bias), it is impoverished at lower (articulatory) levels. According to them, unarticulated inner speech is grounded on abstract linguistic representation and can emerge before any articulatory information is retrieved.

A third argument for the Abstraction view is that typical articulatory abilities are not required in inner speech. Patients with anarthria, who have motor cortex lesions disrupting articulatory abilities, may still have intact inner speech (Baddeley & Wilson, 1985; Vallar & Cappa, 1987). This could suggest that inner speech does not depend on articulation-specific processes. However, as discussed in 1.2, another explanation is that such lesions only affect speech execution, leaving earlier stages of speech planning (including articulatory specification) unaltered.

1.2 Arguments for the concreteness and multimodality of inner language

In contrast with the Abstraction view, it has been suggested that inner speech is concrete in nature, i.e. expressed in a modal format and fully specified, down to physical, motor processes. The earliest claims of the concreteness of inner speech probably date back to Erdmann (1851) and Geiger (1868), who, as cited by Stricker (1885), introspectively observed that inner speech is accompanied by feelings of tension in the speech musculature.

Stricker³⁰ explicitly associated inner speech with motor representations. He speculated that word representations consist in the awareness of impulses driven from cerebral speech centres to speech muscles. In that vein, Watson (1919) described inner speech as a weakened form of overt speech. He considered inner language as a 'highly integrated bodily activity' (p.325). Although Oppenheim & Dell (2010) have held that he went as far as claiming that movements of the articulators are part of inner speech, he merely suggested that inner speech may, in some individuals, be accompanied with articulatory movement. Whether he actually alleged that movements necessarily occur in inner speech, or whether, by the term "integrated activity" he simply meant simulated action, is debatable. The extreme view that inner speech requires actual movement has been refuted by Smith, Brown, Toman, & Googman (1947) who showed that temporary paralysis induced by curare did not prevent verbal thought, memory storage and presumably inner speech. Thus, this extreme version cannot be upheld. A more nuanced view, referred to as the *Motor Simulation* hypothesis, is that inner speech is a mental simulation of articulation, without actual movement. In this view, inner speech production is described as similar to overt speech production, except that motor execution is blocked (Grèzes & Decety, 2001; Postma & Noordanus, 1996). Under the Motor Simulation hypothesis, a continuum exists between overt and covert speech, in line with the continuum between imagined and actual actions proposed by Decety and Jeannerod (1996). This has led some authors to claim that inner speech should share features with speech motor actions (Feinberg, 1978; Jones & Fernyhough, 2007) and that it may be associated with concrete physiological correlates. The Motor Simulation hypothesis is supported by several findings, which we turn to now.

a. *Physiological correlates*

Physiological measurements suggest that inner speech is physically planned, in the same way that overt speech is. First, as concerns respiratory rate, Conrad and Schönle (1979) have showed that the respiratory cycle varies along a continuum. During rest, breathing is symmetrical, with inspiration and expiration phases displaying equal durations. In overt speech, the cycle is strongly asymmetrical with a short inspiration and a long expiration during which speech is emitted. Conrad and Schönle have shown that inner speech displays a slightly prolonged expiratory phase. They concluded that motor processes are at play during inner speech (see also Chapell, 1994).

Speaking rate findings are more debated. As mentioned in 1.1, silent recitation has been found to be faster than overt recitation by many researchers. Some studies have found similar rates for covert and overt recitation, however (Landauer, 1962; Weber & Bach, 1969; Weber & Castleman, 1970). This suggests that the difference might be tenuous. Netsell and colleagues have examined spontaneous sentence production in both covert and overt modes (Netsell, Kleinsasser, & Daniel, 2016). Participants generated full sentences by saying the first thing that came to their mind. The rate of inner productions was found to be slightly faster than that of overt speech. The fact that the difference was small suggests that speaking aloud only differs from inner speech by the longer time needed to overtly articulate, once the motor plan is designed, compared with simulated articulation (see Section 3).

Concerning muscular activity, Stricker's introspective observation that inner speech is accompanied with muscular sensation finds support from a few electromyographic (EMG) studies of inner speech. Using electrodes inserted in the tongue or lips of five participants, Jacobson (1931) was able to detect EMG activity during several tasks requiring inner speech, including silent recitation. Sokolov (1972) carried out EMG measurements of lip and tongue muscles during tasks requiring different degrees of inner verbalisation. He recorded intense muscle activation during complex tasks requiring substantial inner speech production (problem solving). Conversely, a decrease in muscle activity was observed for automatized tasks, with lesser need for inner verbalisation. Surface EMG recordings carried out by McGuigan & Dollins (1989) indicated that the lips were significantly active when silently reading the letter "P" (an instance of bilabial articulation), but not when reading "T" (alveolar articulation) or a nonlinguistic control stimulus. On the opposite, the tongue was significantly active when reading "T", but

³⁰ Stricker himself designed a clever introspective exercise to experience this orofacial activity: when one's mouth is positioned into the rounded shape required to pronounce 'o', if one tries to imagine uttering the phoneme 'm', a slight contraction is felt in the lip muscles, as if one was actually pressing lips for 'm'.

not when reading “P” or the control. The authors concluded that the speech musculature used for the overt production of specific phonemes is also selectively active when covertly reading the same phonemes. Livesay, Liebke, Samaras, & Stanley (1996) measured labial EMG activity in twenty participants during rest and mental tasks. They found a significant increase in EMG activity during silent recitation compared to rest, but no increase during the non-linguistic visualisation task. A study during dreamed speech, using inserted electrodes, suggests that the silent (non-phonated) speech that occurs in dream is associated with EMG activity in *orbicularis oris* and *mentalis* muscles (Shimizu & Inoue, 1986). Surface EMG activity has also been detected in *orbicularis oris inferior* during auditory verbal hallucination (which has been described as inner speech attributed to an external source, see Section 3) in patients with schizophrenia (Rapin, Dohen, Polosan, Perrier, & Løevenbruck, 2013). A study by Nalborczyk *et al.* (2017) on induced mental rumination, which can be viewed as a form of excessive negative inner speech, also shows an increase in labial EMG activity during rumination compared with relaxation. As concerns inner sign language, Max (1937) investigated activity in the *flexores digitorum*, a muscle in the forearm that flexes the fingers, in eighteen deaf participants during silent reading and mental verbal repetition. He observed that, compared to a baseline, these tasks were accompanied by an increase in EMG activity in the *flexores digitorum* in 84% of the cases. EMG activity in a control muscle did not vary as much. Overall, these results suggest that instances of inner speech or inner sign may be accompanied by activity in the orofacial or manual musculature.

b. Cerebral correlates

Several studies show that covert and overt speech production both recruit essential language areas in the left hemisphere, i.e. regions traditionally associated with speech production, such as motor and premotor cortex in the frontal lobe including Broca's area (or the left inferior frontal gyrus, LIFG), regions typically associated with speech perception, i.e. bilateral auditory areas and Wernicke's area in superior temporal gyrus (STG), and an associative region, the left inferior parietal lobule, including the left supramarginal gyrus (LSMG) (for a review, see Perrone-Bertolotti *et al.*, 2014, 2016). However, there are differences. Consistent with the Motor Simulation hypothesis and the notion of a continuum between covert and overt speech, overt speech is associated with stronger activity in motor and premotor cortices than inner speech (e.g., Palmer *et al.*, 2001). This can be related to the suppression of articulatory movements during inner verbal production. Moreover, overt speech recruits sensory areas more strongly than covert speech (Shuster & Lemieux, 2005). Overt speech is therefore not just inner speech with added motor processes, but it involves greater sensory activation, associated with the processing of one's speech. Reciprocally, inner speech involves cerebral areas that are not recruited during overt speech (Basho, Palmer, Rubio, Wulfeck, & Müller, 2007), such as those underlying the inhibition of overt response (cingulate gyrus, left middle frontal gyrus). Overall, these findings support the claim that inner speech is a motor simulation of speech, including motor planning, but excluding motor execution. The processes involved in overt speech therefore include those required for inner speech (except for inhibition). Lesion studies corroborate this conclusion: when overt speech is impaired, inner speech is either intact or altered, depending on the processes impacted. Several studies of brain-lesioned patients with aphasia have shown that the overt speech loss can be associated with an impairment in inner speech (e.g., Levine, Calvanio & Popovics, 1982; Martin & Caramazza, 1982). s.

Geva, Bennett, Warburton, & Patterson (2011a) have reported a dissociation that challenges this view, however. In three patients with chronic post-stroke aphasia (out of twenty-seven patients tested)³¹, poorer homophone and rhyme judgement performance was observed in covert compared with overt mode. Drawing on accounts of speech production that include a speech comprehension system, such as Levelt, Roelofs & Meyer's (1999) model, Geva and colleagues suggested that inner speech relies on a connection between the production and comprehension systems, the latter being used to monitor internal representations. A damage in this connection could selectively impact inner speech while preserving overt speech. A limitation of this study, however, is that the task was to detect rhymes in written words. The deficit could have been induced by silent reading difficulties. To overcome this limitation, Langland-Hassan, Faries, Richardson, & Dietz (2015) have tested aphasia patients

³¹ The other patients were similarly impaired in both inner and overt speech, or had an impairment with overt speech only, resulting from motor deficits or from articulatory encoding difficulties.

with a rhyming task using pictures rather than written words. The performance of patients on covert rhyming was poorer than that of controls, but many patients were unimpaired at overtly naming objects. The authors therefore suggested the deficit could be due to a specific inability to generate words in inner mode. Since the deficit was not due to an impairment in rhyme judgment (patients could judge whether words spoken to them rhymed) and since patients were also impaired in a generative naming task, the authors attributed the deficit in covert rhyming to a difficulty in generating multiple names for the same object to find a word rhyming with the companion picture. The authors left open the possibility that generating speech may be more cognitively and linguistically demanding in covert mode, and that inner speech may be a distinct ability, with specific neural substrates.

We suggest an alternative interpretation of this dissociation. First, the disconnection between production and comprehension systems invoked by Geva and colleagues would also impair overt speech, since the monitoring loop is recruited for overt speech, allowing for the repair of speech errors. Therefore, such a disconnection cannot explain these findings. Secondly, the specificity of inner speech defended by Langland-Hassan and colleagues is hard to reconcile with the fact that in Geva *et al.*'s study, correlations between inner and overt speech were significant. The lack of a comparable task in overt mode in Langland-Hassan *et al.*'s study makes it difficult to conclude along that line. According to our view, rhyme judgement relies on auditory representations of the stimuli (e.g. Paulesu, Frith, & Frackowiak, 1993). Overt speech generates a strong acoustic output, through the ear as well as through bone conduction, which is fed back to the auditory cortex and can be used to monitor speech. In the covert mode, the auditory information is the mentally simulated signal which is not as salient. White noise has been reported to interfere with rhyme judgments (Wilding & White, 1985), which confirms that inner auditory sensations are weak. The fact that even the control participants in Langland-Hassan *et al.*'s study did not reach perfect scores in the silent rhyming task supports this interpretation. In patients with aphasia, the weakness of auditory sensations may be accentuated for two reasons: first, because of an impairment in the final stages of articulatory simulation, and second, because of associated auditory deficits. Interestingly, one of the three patients in Geva *et al.*'s study had auditory comprehension deficits. Therefore, we speculate that the dissociation is due to an amplified lower saliency of the auditory sensations evoked during inner speech.

c. *Articulatory specification*

Another argument for the concreteness of inner speech comes from behavioural evidence of articulatory effects. Advocates of the Abstraction view have suggested that inner speech is impoverished at the articulatory level. This claim is still debated however, since a phonemic similarity bias has in fact been found by Corley, Brocklehurst & Moat (2011) during tongue-twister production, even in a covert mode. Moreover, Scott, Yeung, Gick & Werker (2013) have examined the influence of concurrent inner speech production on speech perception. They showed that the content of inner speech orients the perception of ambiguous syllables. They found that this influence even operates at the articulatory level: the inner production of /a'fa/ vs. /a'pa/ specifically biased perception towards /a'va/ vs. /a'ba/, respectively. A recent fMRI study suggests that inner speech during reading codes detail as fine as voicing (Kell *et al.*, 2017). In this study, the number of voiceless and voiced consonants in the silently read sentences was systematically varied. Increased voicing modulated voice-selective regions in auditory cortex. Overall, these data suggest that inner speech may indeed be specified at the articulatory level.

Moreover, studies on articulatory difficulty also reveal articulatory effects during inner speech. Smith, Hillenbrand, Wasowicz, & Preston (1986) had participants repeat bisyllabic stimuli in both overt and covert modes. The stimuli covered a range of "production difficulty". An important durational range was found across stimuli, in both modes. Words which took longer to be (covertly and overtly) produced involved alternations in similar phonemes in the same syllable position. They concluded that production difficulty (reflected by duration) is not solely due to execution but also to planning. We add that the finding that 'wristwatch' takes longer than 'wristband' in both modes suggests that articulatory specification does occur in inner speech. The labio-velar glide /w/ is articulated with lip rounding and protrusion, and so is the retroflex /r/ (Johnson, 1997). Both require precise control of the lip configuration. This is different from /b/ which involves a ballistic lip closing gesture. The phonemes in the /r/-/w/ alternation are therefore more similar articulatorily than those in /r/-/b/. Motor control studies show that alternating between the movements of two effectors is faster than repetition of a single

effector movement, because in the first case, the motion of one effector can be anticipated during the movement of the other one (Rochet-Capellan & Schwartz, 2007). This explains why ‘wristwatch’ is longer to pronounce overtly than ‘wristband’. The fact that it is also longer covertly suggests that articulatory coordination does take place in inner speech.

d. Gestural representation in covert sign language

Another line of reasoning for the modal nature of inner language comes from the examination of inner language in deaf signers. Behavioural studies have shown that the equivalent of inner speech in deaf signers involves internal representations of signs instead of auditory representations. In a verbal short term memory task, Bellugi, Klima, & Siple (1975) showed that errors made by hearing subjects were mainly sound based, and conform to previous experiments (e.g. ‘vote’ misrecalled as ‘boat’). This suggests that hearing subjects were coding and remembering words in terms of their phonological properties. In deaf signing subjects, substitution errors reflected the visual configurational properties of the signs (e.g. ‘noon’ replaced by ‘tree’, both featuring the same arm position in American Sign Language). Other studies of the properties of verbal working memory in deaf signers reflect a transfer from the auditory to the visual modality, with a sign length effect instead of the auditory word-length effect in spoken language, or a manual suppression effect replacing articulatory suppression (Wilson & Emmorey, 1998). Such studies suggest that sign language is stored in terms of visual percepts as well as manuo-articulatory representations, just like speech is presumably stored in both auditory and oro-articulatory formats³². Therefore, inner language in deaf signers presumably involves an internal representation of signs. As reviewed in MacSweeney, Capek, Campbell & Woll (2008), lesion and neuroimaging studies corroborate these data: like inner speech, inner signing involves a predominantly left-lateralized perisylvian network. Differences exist between the networks supporting signed and spoken languages, reflecting specificities in the early stages of sensory processing (auditory vs. visual) or in higher-level language characteristics (e.g. referential use of space in sign language). Yet, inner language recruits a common core of regions, independent of the modality in which it is expressed.

As mentioned above, and as detailed in Section 3, auditory verbal hallucination (AVH) can be considered as a form of inner speech, which is attributed to an external source. Admittedly, because they often occur in delusional situations, AVH cannot be taken to be fully representative of inner speech. Yet they can be viewed as a specific case of inner speech, worth considering. The descriptions of AVH in deaf patients further illustrates the modality-specific qualities of inner language. Atkinson, Gleeson, Cromwell, & O’Rourke (2007) showed that the hallucinatory phenomenon in deaf schizophrenia patients depended on their auditory experience. Patients born profoundly deaf reported that the “voices” they experienced were nonauditory. They reported seeing a moving image communicating with them through sign, lip motion or fingerspelling. Deaf patients with experience of hearing speech, due to residual hearing or predeafness experience, reported auditory features or uncertainty about mode of perception.

To summarize, behavioural measurements seem to indicate that phonatory-articulatory-gestural planning is at play during inner language and that inner language may be accompanied with activity in the speech and sign musculature. In terms of brain activity, overt and covert language seem to share common core neural correlates, with overt language recruiting motor and sensory areas more than inner language, and inner language recruiting inhibition circuits more than overt language. Therefore, contrary to the Abstraction view, some instances of inner language seem fully physically planned, including concrete articulatory (laryngeal, orofacial and manual) specifications that are coordinated, just like in overt language, but that are inhibited and not executed.

1.3 Coexistence of abstract-amodal and concrete-multimodal forms

The seemingly opposite views of Abstraction and Motor Simulation are not mutually exclusive, however. As explained in Fernyhough (2004), Alderson-Day and Fernyhough (2015) or Geva *et al.* (2011b), at least two levels

³² More precisely, signs are expressed through movements of the arms, hands and also face; speech is expressed through movements of the larynx, tongue, mouth, face and is often accompanied with hand gestures; so both modalities are presumably stored in a brachio-manuo-oro-facial articulatory format (see 2.2).

of inner speech can be distinguished. The first one, condensed inner speech, is argued to correspond to Vygotsky's (1934) description: "inner speech is to a large extent thinking in pure meanings" (p.249). In Vygotsky's view, inner speech has lost most of the acoustic and structural qualities of external speech. As Vygotsky wrote, "[Th]e development of verbal thought takes the [following] course: from the motive that engenders a thought to the shaping of the thought, first in inner speech, then in meanings of words, and finally in words" (p.253). This level of inner speech can indeed be considered as abstract in format. Expanded inner speech, on the other hand, retains many of the phonological properties of external dialogue, and can be viewed as concrete in format. Fernyhough (2004) has suggested that inner speech varies with cognitive and emotional conditions between these two (or more) forms. We consider the expanded form as an outcome of the condensed form. The condensed form, we conjecture, is the conceptual message cast in a preliminary linguistic form, that involves lemmas³³, linearly ordered, but that does not yet have the full phonological (articulatory, gestural, acoustic) specification that expanded inner language has. A similar position is taken and defended in detail in Vicente & Martínez-Manrique (2016). Inner language can be defined as truncated overt verbalisation, but the level at which the production process is interrupted (abstract linguistic representation vs. articulatory/gestural representation) depends on which variant of inner language is at play. In the rest of this chapter, we will focus on expanded inner language.

2. The sensory-motor dimension of inner language

If we accept the concrete nature of inner language, at least in its expanded version, then we are still faced with another question related to its nature: is inner language motor or sensory? Are inner speaking (or signing) and inner hearing (or viewing) different phenomena or are they two sides of the same coin?

2.1 Arguments for a motor or enactive nature

As explained in 1.2, the Motor Simulation view, also referred to as the 'Action' view (Jones & Fernyhough, 2007) or the 'Activity' view (Martínez-Manrique & Vicente, 2015) holds that inner language is an act, with a prior intention to express a certain thought, which is transformed into orofacial and/or manual motor commands. This view is grounded both on introspective experiments and empirical data (physiological recordings, behavioural measures as well as neuroimaging and brain lesion data) described above.

Inner language therefore seems to involve motor acts that are inhibited. If inhibition prevents motor acts from actually being executed, then the neurophysiological activity measured in peripheral muscles must be explained. We suggest that motor commands might be emitted, together with inhibitory signals blocking articulatory movement. This speculation is in line with Jeannerod & Decety's (1995) description of action imagery. According to them, during mental simulation of an action, "it is likely that the excitatory motor output generated for executing the action is counterbalanced by another parallel inhibitory output. The competition between two opposite outputs would account for the partial block of the motoneurons, as shown by residual EMG recordings and increased reflex excitability" (p.728).

Inner language therefore seems to involve the production of imaginary motor acts; be they articulatory, facial or manual. In a predictive control account, these imaginary motor acts can be viewed as the predicted actions that result from a copy of inhibited motor commands (see Section 3). They can be posited to correspond to the activations observed in premotor cortex and inferior frontal regions. They seem to have physiological sequels in orofacial muscles and in respiratory patterns. It could therefore be concluded from empirical data that inner language is fundamentally of a motor or enactive nature.

Yet, as explained in the preceding section, these imaginary motor acts give rise to sensory percepts, feelings in our muscles (Stricker, 1885) but also sounds in our heads. Taine (1870) himself was a precursor when he recognized the motor and sensory qualities of verbal thought: 'In normal state, we silently think with words that

³³ The term lemma in Levelt and colleagues' terminology refers to the word's syntax, see Levelt et al. (1999). It is different from the lexeme which denotes the word's phonological features and from the lexical concept which refers to the word's semantics.

are mentally heard, read or uttered, and what is inside of us is the image of such sounds, letters, or of such muscular and tactile sensations in the throat, tongue and lips' (p.25-26, our translation). The sensory qualities of inner language are examined in the following section.

2.2 Arguments for a sensory nature

Early introspective works have claimed that inner speech is endowed with auditory qualities. Egger (1881) and Ballet (1886) claimed that rhythm, pitch, intensity and even timbre can be found in inner speech. The concept of an inner ear (or mind's ear) finds support in recent data.

The 'Verbal Transformation Effect' (VTE) refers to the perceptual phenomenon in which listeners report hearing a new percept when an ambiguous stimulus is repeated rapidly (Warren, 1961). Rapid repetitions of the word 'life', for example, produce a soundstream that is fully compatible with segmentations into 'life' or 'fly'. Reisberg, Smith, Baxter, & Sonenshine (1989) examined the imagery analogue of the VTE. Participants were instructed to imagine the word "stress" being repeated by a friend's voice. The VTE was observed (subjects detected the compatible word "dress") showing that subjects are able to imagine an ambiguous soundstream, to parse it and find alternative construal of it. Smith, Wilson, & Reisberg (1995) further studied the VTE in a covert mode, using Baddeley's distinction between two components of the phonological loop involved in verbal short-term memory. According to Baddeley, the phonological loop relies on the "inner ear" – the phonological store –, and on the "inner voice"³⁴ – the articulatory rehearsal process. Smith and colleagues examined the VTE in a covert mode, asking whether the imagery judgement and reconstrual is based on the inner voice, the inner ear or both. Participants were instructed to imagine a friend repeating the word "stress" and to report any transformation. Repetition imagery was executed in three conditions: no-interference, articulatory suppression and irrelevant speech perception. A more important VTE was found in the no-interference condition than in the suppression and irrelevant-speech conditions. The disruptive impact of articulatory suppression was interpreted as a role for the inner voice in the VTE: to discern the transformations, subjects need to subvocally rehearse the material. The impact of irrelevant speech was taken to suggest that the VTE also depends on the inner ear. It was concluded that subjects seem to reinterpret ambiguous verbal images by using both components of the phonological loop, the inner voice and ear.

The neural correlates of the VTE have been examined by Sato *et al.* (2004). Participants were asked to silently repeat pseudo-words such as /psə/. In the baseline condition, participants were asked to covertly repeat a pseudo-word over and over. In the verbal transformation condition, they additionally had to actively search for a transformation (from /psə/ to /səp/ for instance). When compared with the baseline condition, active search for verbal transformation correlated with stronger activation in the left inferior frontal gyrus, left supramarginal gyrus, bilateral cerebellum as well as left superior temporal gyrus: when inner speech involves consciously attending to mental production, speech production as well as perception regions are more strongly activated. These results therefore corroborate the hypothesis of a close partnership between inner production and perception in the VTE.

Findings of error detection during covert tongue-twister repetition also seem to indicate that inner verbal production has sensory qualities that can be attended to. As mentioned in Section 1, several studies (reviewed in Dell & Oppenheim, 2015) show that participants are able to report the errors that they mentally hear. This can be interpreted as a role for the mind's ear in inner speech monitoring. A recent fMRI study of slip detection provides contradictory findings, however. Gauvin, Baene, Brass, & Hartsuiker (2016) investigated whether internal verbal monitoring takes place through the speech perception system. In a production condition, they had participants produce tongue-twisters overtly and judge whether their production was correct or incorrect, while white noise was presented via headphones to mask auditory feedback. Adding noise was meant to induce internal verbal monitoring, as participants could not hear their auditory feedback, while ensuring that the experimenter could judge repetition correctness. In a perception condition, participants simply heard the

³⁴ 'Inner voice' is taken here as the imaginary motor act (articulation and phonation). In the rest of the chapter, it refers to the result of that act, i.e. the auditory stimulus heard in the mind's ear.

tonguetwister and made a correctness judgment. The superior temporal areas were found to be activated by error detection during the perception condition but not during production. The authors concluded that internal monitoring occurs independently of speech perception systems. The fact that no activation was found in speech perception areas during the production condition could be due to the use of noise masking, however. Adding noise saturates the auditory system, which could mask subtle differences between contrasts. The examined contrasts were between erroneous and correct trials. Erroneous trials were instances in which an error was detected, which could indeed augment the activation of the auditory system, relative to a correct trial, but quite subtly. Since in both types of trials, the noise level was high, this subtle difference might have been undetectable. Therefore, we do not think that these results are conclusive.

Neuroimaging studies of covert speech production themselves reveal auditory cortex, and specifically superior temporal gyrus, activation (Perrone-Bertolotti *et al.*, 2014 for a review). Although this activation is lesser than the one observed in overt speech, it entails that an auditory experience accompanies inner speech. An interesting study suggests that inner speech can be disrupted during abnormal activity in the temporal lobe. Vercueil & Perrone-Bertolotti (2013) described the case of a woman who reported experiencing inner speech jargon (incomprehension of her own inner language) during her epileptic seizures which involved sharp theta waves in the left temporal regions.

Evidence for auditory sensations during reading has been provided by experimental psychology. Several studies suggest that silent reading is modulated by the knowledge of the author's speaking speed (Alexander & Nygaard, 2008), the talker's voice familiarity (Kurby, Magliano, & Rapp, 2009) or the reader's regional accent (Filik & Barber, 2011). The involvement of the mind's ear during silent reading has been recently confirmed by fMRI experiments (Yao, Belin, & Scheepers, 2011, 2012). Several areas in the auditory cortex, called temporal voice area (TVA), are selectively involved during human voice perception (Belin, Zatorre, Lafaille, Ahad, & Pike, 2000). Yao and colleagues contrasted silent reading of direct (e.g., Mary said: "I'm hungry") and indirect speech (e.g., Mary said that she was hungry) sentences. The direct speech condition induced greater activation of the right TVA than the indirect speech condition, which suggest that voice-related perceptual representations are more engaged when silently reading direct speech statements. Further support for the assumption that silent reading involves the mind's ear comes from an fMRI study by Lœvenbruck, Baciú, Segebarth & Abry (2005). In the baseline condition, participants silently read a sentence in French, with a neutral prosody (*Madeleine m'amena*, "Madeleine brought me around"). In the prosodic focus condition, they silently read the same sentence, adding contrastive focus on the subject. In an overt mode, this would correspond to higher pitch and longer duration on the focused subject, followed by pitch compression on the post-focal constituents (*MADELEINE_F m'amena*). When compared with the baseline, the silent prosodic focus condition yielded greater activity in the left inferior frontal gyrus, insula, supramarginal gyrus as well as in Wernicke's area. These results suggest that when we silently read, we can use specific prosodic contours, with distinctive auditory qualities. These auditory variations correspond to objectively measurable cerebral correlates. Further evidence for TVA activation during silent reading comes from intracranial EEG recording of TVA in four epileptic patients (Perrone-Bertolotti, Kujala, Vidal *et al.*, 2012). Patients were instructed to perform a silent reading task in which attention was manipulated: they were asked to only attend to the words written in grey, ignoring the white words. Consecutive grey words formed a story, about which they were questioned after the experiment. The results not only showed that silent reading activate the TVA, but also, that the neural response to written words was increased during attended compared to unattended words. This suggest that TVA activity increase is under top-down attentional control. It must be noted however, that reading is not systematically associated with inner speech, even when attention is high. A few aphasia case reports suggest that some reading abilities may be maintained even when inner speech is impaired (Levine *et al.*, 1982; Saffran & Marin, 1977). This can be explained by the fact that silent reading of frequent words may take a direct route from orthography to meaning, without necessarily recurring to inner speech (Coltheart, 2005). To sum up, behavioural and neuroimaging data suggest that auditory sensations are often present during silent reading.

The concept of a mind's ear is appropriate, but it is insufficient. As we have argued, the imaginary sensory consequences of imaginary motor acts may be multimodal: they may lead to sounds in our heads and, as hinted by Taine (1870) or Stricker (1885), to imaginary proprioceptive and tactile sensations. Paulhan (1886) claimed

that inner speech involves visual, auditory and motor images. By visual images he meant the form, shape and colour of the letters that compose written words. He stated that these were rare. He qualified auditory images as dominant in inner speech. He defined motor images as the sensations in the speech organs that sometimes accompany inner speech. Contrary to Stricker who considered inner speech as purely motor, he claimed that motor images cannot be isolated from auditory images, whereas the reverse is possible.

A few terminology precautions are necessary here. It is not always clear what the nineteenth century authors meant by “motor” and “articulatory” representations. Even nowadays, “articulatory” is often opposed to “auditory” or “acoustic”, with some confusion. Sometimes, the process is targeted: what is meant by “articulatory” is motion (action), in contrast to audition (perception). Sometimes, modality is at play: “articulatory” refers to somatosensory sensations, in contrast to auditory percepts. Bearing in mind this confusion, we use the term “motor” to refer to action and “somatosensory” to describe bodily sensations. Although Stricker clearly claimed that inner speech consisted of imagined actions, Paulhan’s intuitive notion of “motor images” are related to somatosensory percepts, i.e. to the evocation of sensations, rather than to the simulation of actual speech movements.

Nevertheless, inner speech seems indeed to involve somatosensory sensations, which include proprioception and tactile sensations. Proprioception provides information about articulator location and movement and is sent by receptors in the muscles, joints and skin. Tactile information corresponds to the touch sense from mechanoreceptors that report contact (e.g. between tongue and palate). According to Lackner & Tuller (1979), speech errors can be detected by means of proprioceptive and tactile information and it has been claimed that proprioceptive and tactile feedback play a role in speech motor control (Levelt, 1989; Postma, 2000; Gick, 2015). We speculate that imagined proprioceptive and tactile feedback are part of inner speech: in addition to the mind’s ear, the mind’s ‘tact’ should also be considered. Moreover, the fact that, as explained above, motor commands may reach muscles during inner speech, could explain the actual (not imagined) sensations in the speech muscles introspectively reported by Stricker and Paulhan. We will further address the co-existence of motor, auditory and proprioceptive representations in Section 3.

Finally, the ‘mind’s eye’ certainly plays a role in inner language. As mentioned earlier, inner language representations in deaf signers include visual information. Gestures are not only used in the deaf population. They accompany speech in normal hearers and play a fundamental role in thought and speech (De Ruiter, 2007). Moreover, speech is audiovisual: lip reading enhances speech comprehension when the acoustic signal is degraded by noise (Sumbly & Pollack, 1954). Lip reading occurs even with nondegraded acoustic signals, as the McGurk effect shows (McGurk & MacDonald, 1976). Auditory and visual speech information include common stages of processing (Nahorna, Berthommier, & Schwartz, 2015). These findings suggest that visual information (facial and manual) could be involved in inner speech, even in hearing subjects. A preliminary work by Arnaud, Schwartz, Løevenbruck, & Savariaux (2008) provides tentative suggestions that speakers can have visual representations of their own lip movements. Furthermore, as suggested by Paulhan, visual written representations may occur during inner speech. More research is needed to confirm that inner language involves visual (labial, facial, manual, written) representations, even in the hearing population.

Inner verbalizing therefore involves the reception of imaginary sensory signals, presumably including auditory, proprioceptive, tactile and visual elements, handled by the mind’s ear, tact and eye.

To wrap up, the nature of inner language is both motor and sensory. One can conceive that imaginary acts give rise to multisensory percepts. But these acts themselves could stem from prior sensory goals, as Paulhan hinted in 1886. The precedence of some sensory representations over motor ones in wilful inner verbalization will now be discussed, in a motor control framework.

3. Integrating the sensory-motor nature of inner language into the ‘Predictive Control’ account

The ‘sensory-motor’ nature of inner verbalization can be accounted for in a motor control perspective in which intended sensory goals can give rise to motor acts which themselves generate sensory percepts. The ‘predictive control’ account of inner speech, also called ‘comparator model’, pertains to this perspective. This account is

based on the hypothesis that action control uses internal models, i.e. systems that simulate the behaviour of a natural process (Kawato, Furukawa, & Suzuki, 1987; Jordan & Rumelhart, 1992). Two kinds of internal models, forward and inverse models, are supposed to be coupled and regulated through several comparators. A forward model is an internal representation of the system (body, limb, organ) that captures the forward or causal relationship between the inputs to the system (motor commands) and the outputs (Wolpert & Kawato, 1998). An inverse model performs the inverse computation, i.e. provides motor commands from desired sensory states. During the execution of a goal-directed motor task, an inverse model computes motor commands from the specification of desired changes in the sensory state of the motor apparatus. A copy of the motor commands, called “efference copy”, is fed to a forward model that, given the current state of the apparatus, generates a prediction of the upcoming sensory consequences of the action. Thanks to its negligible delay, this sensory prediction, also called “*internal feedback*”, ensures a stable feedback control of actions (Miall, Weir, Wolpert, & Stein, 1993; Miall & Wolpert, 1996; Wolpert & Kawato, 1998). The propagation of the *actual* feedback to the central nervous system is indeed delayed, due to axon transmission and synaptic delays (during speech production, the delay between auditory feedback perturbation and motor command adaptation is about 200 ms, i.e. the duration of one syllable, Houde, Nagarajan, Sekihara, & Merzenich, 2002). Because of these delays, a control based on actual feedback would either require very slow execution or be unstable. Forward models, by providing an internal feedback that occurs earlier than the actual experience, can trigger early error correction and allow for stable action control.

The efference copy mechanism is not only crucial to smooth motor control. It is also considered to play a role in the awareness of action. It has been hypothesized that disruptions in the predictive mechanism could lead to delusions of control and, in the case of speech, to auditory hallucination (e.g., Frith, 1992; Frith, Blakemore, & Wolpert, 2000). A model is presented in Figure 1, that explains this hypothesis in the context of overt speech (and that includes an adaptation to inner speech, detailed below). To make things clear, we take as example the goal of ‘uttering vowel /i/’, although it is debatable whether such a mechanism is necessary in isolated vowel production. In Frith and colleagues’ view, the goal is associated with a desired³⁵ multisensory state, which can be expressed in terms of articulatory properties (anterior elevated tongue position, lip spreading, phonation) as well as acoustic properties (first two spectral formants spread apart). An inverse model transforms the desired sensory state into motor commands, which are sent to the articulatory-phonatory motor system. This leads to the production of labial, lingual, and laryngeal movements, and to an acoustic signal. In turn, these movements and sound generate long-delay somatosensory and auditory feedbacks, the actual sensory experience. An efference copy of the motor commands is also sent to a forward model, which generates predicted somatosensory and auditory feedbacks. A delay is applied to the internal feedback signals (which become the “*corollary discharge*”) so that they are synchronized with the actual feedbacks. The efference copy mechanism is depicted in dashed line in Figure 1. The predictive model includes three state comparisons which have each a specific role in overt speech production.

Insert Figure 1 about here

The first comparison (referred to as C1 in Figure 1) takes place between the actual sensory feedback and the desired sensory state. If a discrepancy results from C1, the inverse model receives an error signal and the motor commands are adjusted. C1 is irrelevant in ongoing actions, as the time necessary for the actual feedback to reach the central nervous system is of about one syllable. This would lead to utterly slow speech production. C1 is instead supposed to play a role in speech learning, by tuning the inverse model to produce motor commands that are best adapted to new goals. Moreover, it has been suggested that C1 contributes to the sense of body ownership, the pre-reflective experience that it is our own body that is currently moving, voluntarily or not (Gallagher, 2000; Franck & Thibaut, 2003; Tsakiris, Schütz-Bosbach, & Gallagher, 2007).

³⁵ We use the term « desired » rather than « intended », to allow for unintended action to be monitored via this mechanism (see below, on unbidden thoughts).

The second comparison (C2) is the one involved in the stable control of actions, using internal instead of actual feedbacks. It compares desired and predicted states. Via C2, errors can be detected in the motor commands, and be corrected, before actual feedback reaches the central nervous system³⁶.

A third comparison (C3) is involved, between the actual sensory state and the delayed prediction (corollary discharge). If the afferent sensory feedback and the corollary discharge do not match, the forward model is adjusted. This forward model updating, together with the inverse model tuning via C1, are claimed to improve performance when learning new actions, by generalizing the tuning for future productions (but see Tremblay, Houle, & Ostry, 2008; Rochet-Capellan, Richer, & Ostry, 2012, who have only observed limited generalization to future actions). It has been suggested that C3 could also play a role in self-monitoring (Wolpert, Ghahramani, & Jordan, 1995; Wolpert, 1997). If the actual sensory feedback matches the predicted sensory signal, then the sensory cortex could be informed that the perceived stimuli are self-generated, which would provide a sense of agency. Frith (1992) posited that a defective predictive system could explain why a self-initiated action may be experienced as externally controlled in delusions of control: if the predicted and actual sensory feedbacks do not match, then some external influence must have taken place.

In line with agency, another advantage of the predictive model is that it explains the observed modulation of sensory cortex activity during self-initiated actions. If subjects can predict the sensations they are going to feel, then these are not informative and can be attenuated, relative to externally caused sensations which need to be attended. When actual and predicted feedbacks match, the sensory consequence of the motor act is thus attenuated, compared with the same stimulation produced by an external agent (Blakemore, 2003; Blakemore, Frith, & Wolpert, 1999; Frith, 2002). This mechanism has been invoked to explain why we cannot tickle ourselves (Blakemore, Wolpert, & Frith, 2000).

According to some authors, the mere presence of a predicted signal (even before C3 takes place) could itself contribute to the awareness of initiating a movement, to feeling in control (Blakemore, Wolpert, & Frith, 2002; Frith, 2002). Temporal measurements by Libet, Gleason, Wright, & Pearl (1983) or Haggard, Newman, & Magno (1999) indeed indicate that subjects are aware of initiating a movement about 80 ms before the actual movement occurs.

In sum, predictive control seems to play an important role in self-monitoring. It is claimed that it provides senses of ownership and agency, that are essential components of self-awareness, via C1 and C3 comparisons, involving desired, predicted and actual states.

The predictive control framework has been fruitful in the speech domain (Guenther, Ghosh, & Tourville, 2006; Houde & Nagarajan, 2011; Postma, 2000). It has even been applied to covert speech production (Feinberg, 1978; Frith, 1992). Several researchers, including Frith (1992), Jones & Fernyhough (2007), Seal, Aleman, & McGuire (2004), have claimed that disruptions in the predictive control mechanism explain auditory verbal hallucination (AVH). According to their view, if the prediction is faulty, the actual sensory consequences of inner speech are not attenuated and agency is not felt. Either because of attributional biases (Seal *et al.*, 2004) or simply because self-authorship is not felt (Jones & Fernyhough, 2007), inner speech would then be experienced as other-generated.

The involvement of a corollary discharge in inner speech control is supported by several studies. Dampening or delaying of auditory cortex responsivity has been observed during inner speech (with EEG: Ford & Mathalon, 2004; with MEG: Numminen & Curio, 1999) and interpreted as a modulatory influence of frontal speech production areas on temporal speech reception areas. This finding should be interpreted with caution, however. In Ford & Mathalon's EEG study, the inner production was preceded by an auditory stimulus, which could have dampened the subsequent auditory response (via auditory suppression). Nevertheless, in an fMRI study, Shergill

³⁶ Discrepancies between desired and predicted feedbacks could also be due to an inaccurate forward model. C2 could therefore also contribute to adjust the forward model, not just the inverse model. But this would require an additional mechanism by which the discrepancy would be sent either to the inverse or the forward model. In the absence of evidence for such a mechanism, we stick to the classical view, with C2 only affecting the inverse model.

et al. (2002) did find increased fronto-temporal connectivity during inner speech, associated with the increase in inner speaking rate. Tian, Zarate, & Poeppel (2016) also found temporal cortex activation during inner speech, which they related to the presence of a corollary discharge. Scott (2013) provided behavioural evidence for auditory attenuation in inner speech. The “Mann effect” refers to the influence of contextual speech sounds on the perception of subsequent speech sounds. Scott showed that this effect was specifically weakened when the contextual sound was played during matching speech imagery, suggesting that the impact of the auditory stimuli was only attenuated when inner speech matched.

Some researchers have questioned the functional relevance of a monitoring system in inner speech, however. MacKay (1992) specifically asked “*why speakers must independently ‘listen to’ the meaning and sound of what they are saying internally when they know all along the meaning and sound of what they are saying*” (p.140). Stephens & Graham (2000), Gallagher (2004) and Langland-Hassan (2008) also argue that the predictive mechanism is redundant in inner speech. This critique has been addressed by Jones & Fernyhough (2007) who inscribe the necessity for self-monitoring in a Vygotskian developmental perspective. According to them, children start off by overt “private speech”, simulating dialogues with interlocutors. Verbal thought would only become covert after several years, through a gradual process of internalization. During this process, it is crucial for children to be able to label the received auditory stimuli as self- or other-generated. This means that the efference copy is not an *ad hoc* mechanism solely invoked to explain delusions of thought insertion, but it is ontogenetically necessary for inner speech to develop from private speech. We further claim that distinguishing self-generated from other-generated voices remains compelling in adult inner speech. As argued in Section 2, we can hear our inner voice, its timbre, and its intonational variations, we can even detect inner speech errors. We can have imaginary dialogues, involving various voices. We claim that it is through self-monitoring that we do not mistake these internal voices for external voices, and that we are aware that we have imagined them. A broader role for the predictive mechanism will be discussed below, related to awareness and distinguishing wilful inner speech from unbidden thoughts. But before this, we need to address a further critique, stemming from the direct application of the predictive control model to inner speech. Although Frith (1992, see also Feinberg, 1978) was one of the first to suggest that inner speech control could rely on such a model, he himself questioned the notion of actual sensory feedback during inner speech, which by essence is silent and motionless (Frith, 2012). In the case of inner speech, C3 is irrelevant, as it would compare a predicted sensory signal with an absent actual feedback. Rapin *et al.* (2013) have offered two alternative accounts (see also Rapin, Dohen, & Løevenbruck, 2016).

The first account relies on the hypothesis that, as argued in 2.1, inhibitory signals may be sent to prevent motor command amplitude from reaching a sufficient threshold for speech movement to occur. But even though the speech apparatus may not move, the motor commands could slightly increase muscle tension. The actual sensory feedback during inner speech would thus consist of some residual proprioceptive feedback (rather than auditory). During AVHs, this residual signal could be the sensory feedback that does not match the faulty prediction and that leads to self-generated signals being interpreted as external.

In the second account, the relevant comparison for agency-monitoring during inner speech is not C3 because the actual feedback is silent and motionless. Agency during inner speech, which is faulty during AVH, cannot come either from the mere presence of a prediction (see above), because we claim that the predicted signal is precisely what becomes identified as an external voice (or manual/ facial gestures in deaf subjects). Instead, we suggest that agency comes from C2, the comparison between desired and predicted states (a distinction between predicted state and predicted experience is made below). The AVH symptoms could be explained as follows: if the prediction is defective, then there is no match between predicted and desired states, agency is not felt and the inner voice or gesture (predicted experience) could feel alien. In addition, C2 would signal a discrepancy, which would abort the perceptual attenuation and which would reinforce the saliency of the inner voice (or gesture), accentuating its alien character. It must be reminded that C2 was originally introduced to explain stable feedback control of action, by early tuning of the motor commands when the predicted state does not match the goal. C2 should therefore still issue a sense of agency in case of goal unattainment, i.e. when the prediction and the goal are only slightly discrepant.

The first account, which entails that some proprioceptive sensations could subsist in muscles during inner speech, is supported by introspective experiments (Stricker, 1885). The second account has been concurrently formulated by Tian & Poeppel (2012) as well as Swiney & Sousa (2014) who have similarly proposed that C2 is the suitable comparison for agency and perceptual attenuation in inner speech. This was even proposed by Frith (2005) himself. It can also be found, incidentally, in Gallagher (2000). These two accounts are compatible and are integrated in Figure 1. The lines and boxes shaded in light grey are irrelevant in inner speech and only apply to overt speech. The red arrow corresponds to inhibitory signals sent in parallel with the goal of inner speaking (this arrow is irrelevant in overt speech).

We have added to Figure 1 the concept of “inner language percepts”, at the level of the predicted experience. Tian & Poeppel, Swiney & Sousa, as well as Scott (2013) and Scott *et al.* (2013), argue, like us, that the voice perceived during inner speech precisely consist of the predicted signal. In other domains, researchers have claimed that the forward model could be used during mental training, to predict the sensory consequences of an action without having to execute it. This could tune the inverse model for future actions (Jeannerod & Pacherie, 2004; Pacherie, 2008). The predicted signal would thus correspond to the subjective feeling in mental imagery (Grush, 2004). During inner speech, the predicted signal would thus equate to the voice mentally heard (and the somatosensory sensations felt), or the sign/lip gesture internally seen. As explained above, this simulated signal occurs earlier than the actual experience would, which explains why inner speech may be shorter than overt speech.

We also include in Figure 1 two perceptual attenuation mechanisms, at the level of C3 for external sensory signals and at C2 for internally generated signals. During overt speech control, we speculate that agency could result from both C2 and C3 and would therefore be stronger than during inner speech. C2 would attenuate the predicted sensory signal (the inner voice) and C3 would dampen the external feedback.

According to our view, the efference copy is therefore more than a mechanism that identifies self vs. other-generated voices. It is what makes our own verbal thoughts come to awareness (Frith, 2010). As mentioned above, it has been argued that the desired state, used by the inverse model to derive the efference copy, could itself be the inner voice consciously heard and felt as our own, with no need for a prediction (Gallagher, 2004; Langland-Hassan, 2008). In Langland-Hassan’s ‘filter model’, the mere existence of an efference copy, without computing a prediction, and without using a comparator, could act as a filter during inner speech. This filter would itself endow our actions with a sense of agency. In this alternative model, however, the comparator mechanism is argued to be necessary for other somatosensory modalities than those involved in inner speech. This entails that different mechanisms would be required depending on the modality, which does not seem parsimonious. We add that the desired state itself cannot be experienced as a voice. In many motor control theories, the comparisons take place between end sensory states, not between ongoing experiences. We speculate that the desired state, being expressed in terms of goals in acoustic and articulatory spaces, is a coarse plan, not a full speech experience, with the unfolding of speech muscle movements and sounds over time. In Figure 1, we have included our speculated distinction between predicted experience and predicted end state. The predicted experience, developing over time, is the inner voice. The predicted state is the end sensory product, compared with the desired goal. The inner voice is not thoroughly felt until it is fully simulated over time, through the efference copy. And it is not felt as self-intended before its end product, the predicted state, is compared with the desired state. We further speculate that top-down executive signals presumably control for the generation of a prediction. Three types of verbal thought can then be explained. First, unbidden thoughts, i.e. verbal thoughts without a feeling of agency (Gallagher, 2004), can be viewed as desired states with no corresponding predicted states. They sound evanescent and muffled, because they are not fully specified over time. They do not feel alien, because no comparison is made at all, presumably because no top-down signal has launched the generation of a prediction. In Figure 1, we have added “unbidden thoughts”, at the level of the desired state. A second type is wilful inner speech, in which top-down signals initiate the generation of a prediction. A sensory experience unfolds over time and an inner voice is distinctly heard. The desired and predicted states match (even only slightly): agency is felt. A third type is AVH, in which top-down signals initiate the generation of a prediction, but, due to a dysfunction, the desired and predicted states do not match at all and the prediction feels alien. The alien voice is vividly heard, as the absence of perceptual attenuation (due to

the discrepant comparison) makes the predicted experience more salient than an ordinary inner voice. The efference copy mechanism therefore contributes to creating the rich sensory qualities of inner speech, as well as the feeling of agency, of awareness of our thoughts.

In summary, in our adapted version of the predictive control account, wilful inner language is seen as a process in which verbal goals are converted to motor commands. These motor commands are inhibited but still transmitted to the orofacial and manual systems, giving rise to residual proprioceptive feedback. This residual feedback may provide a sense of ownership and is probably felt by some individuals. In parallel, a copy of the motor commands is sent to a forward model which computes multisensory inner language percepts associated with the simulated acts: sounds – the inner voice heard by our mind’s ear –, proprioceptive sensations in the orofacial and manual musculature – the inner movements perceived by our mind’s tact –, and visible manual/lip gestures – the inner signs potentially seen by our mind’s eye.

We note here that this account does not explain how visual information about one’s manual and facial movements may be derived from motor commands. The role of forward models is to map motor commands onto resulting sensory percepts, through a simulation of the motor and sensory systems. During overt speech, the sensory percepts resulting from the motor commands sent to the speech musculature to pronounce an /i/, for instance, correspond to the audition of the /i/ sound and the associated proprioceptive sensation. But are they linked with visual information about the associated facial configuration? In sign language, the sensory percepts associated with the motor commands sent to form the sign for ‘tree’, for instance, correspond to the proprioceptive sensation of a raised arm and hand as well as to the vision of a straight arm and extended hand, in an egocentric perspective. Lip-reading being so important in deafness, is visual information about one’s arm linked with information about one’s lips, in an allocentric perspective? If future research shows that visual information about one’s face indeed plays a role in language production, even in hearing subjects (as suggested in Section 2), then an additional mechanism needs to be included to handle the presence of predicted allocentric facial visual information in addition to the predicted egocentric visual feedback about the arm and hand.

4. A cerebral landscape

We will now sketch a landscape of the cerebral regions involved in wilful inner language production. Our sketch is based on findings and theoretical assumptions in linguistics, psycholinguistics, and neurolinguistics described in the previous sections. It shares some of the hypotheses described in the functional anatomic models of overt speech production by Guenther & Vladusich (2012), Hickok (2012) or Tian & Poeppel (2013), but it differs in specific points. It is displayed in Figure 2. The diagram in Figure 1 is also complemented with anatomical locations corresponding with this sketch. In both figures, the efference copy mechanism is depicted in dashed lines.

A few words of caution are first needed. The functional anatomic sketch proposed here specifically describes volitional inner language production. Additional regions, or perhaps a different network altogether, may be at play for the more evanescent and less wilful form of inner speech that corresponds to verbal mind wandering (see Introduction). Moreover, Hurlburt (2011) makes a phenomenological distinction between “inner hearing” and “inner speaking”, which does not coincide with our view of inner hearing as the sensory prediction elicited by the act of inner speaking. Hurlburt claims that another form of inner hearing exists, in which subjects feel as the recipient of the voice, not their creator. A different network might mediate this particular phenomenon of “inner hearing”. A neuroimaging study seems to confirm this intuition (Hurlburt, Alderson-Day, Kühn, & Fernyhough, 2016), although we think the inner hearing phenomenon elicited was in fact related to verbal mind wandering. Hurlburt’s inner hearing could be close to verbal mind wandering, or it could be akin to auditory verbal hallucination. Our sketch is restricted to Hurlburt’s “inner speaking”. Moreover, we only consider the very last concrete stages of inner language production, sometimes referred to as “expanded inner speech”. As argued in 1.3, condensed inner speech may correspond to the initial stages of inner language. These include two of the stages described in Levelt (1989): the conceptualizer, the output of which is a linearized preverbal message, and grammatical encoding, which consists in selecting the appropriate lemmas from the lexicon and arranging them in a syntactic order. We start our sketch of inner language production precisely where condensed inner speech may most certainly stop, *i.e.* once lemmas have been retrieved and arranged. To simplify things, we restrict the

landscape to the production of single words with no consideration of syntax or prosody. Inner speech can be produced with one's own or someone else's voice (e.g. Geiselman & Glenny, 1977). The cerebral networks underlying the monitoring of different voices remain to be described (but see Grandchamp *et al.*, 2016), so our sketch is limited to own-voice inner speech. Finally, current data on inner sign are too scarce, but we speculate that the auditory processes invoked in our sketch may be replaced with visual processes in inner sign. Therefore, our sketch only applies to wilful spoken inner production of isolated words with one own's voice.

Insert Figure 2 about here

So, if we skip conceptual preparation and grammatical encoding, we can start off with a lemma being retrieved. The meta-analyses by Indefrey & Levelt (2004) and Indefrey (2011) suggest that lemma retrieval is handled by the mid-section of the left middle temporal gyrus (Brodmann Area (BA) 21). Tian & Poeppel (2013) locate this process in the left posterior middle temporal gyrus. Until more research decides between these two proposals, we broadly associate lemma retrieval with the left middle temporal gyrus. So, inner word production presumably starts with an activated left middle temporal gyrus mediating lemma retrieval.

According to Levelt *et al.* (1999), the next stage is phonological code retrieval, which generates the lexeme. It is not clear whether Levelt and colleagues think it implies sound as well as articulation. Although Levelt (1994) states that phonological encoding generates "an articulatory or phonetic shape for all words" (p.91), Indefrey & Levelt (2004) in fact reduce this stage to activations in Wernicke's area. Tian & Poeppel (2013)'s model also limits this stage to auditory specification. We see things differently. In our revised predictive control account, the inner language goal (or the lemma) is associated with a desired state, expressed in a multisensory format. Because more research is needed to confirm the role of visual information, we restrict the sketch to auditory and somatosensory information. The sketch remains fully compatible with the inclusion of visual information, via visual cortex activation, however (and is also compatible with inner sign production). According to us, the lemma is converted to a lexeme in a multisensory format, through two pathways, one for auditory and one for somatosensory representations. These two pathways are presumably parallel, but auditory specification may in fact be sequentially followed by somatosensory specification, or the reverse. A similar view is proposed in Hickok (2012)'s model. Hickok makes the additional claim that these two pathways correspond to two hierarchical levels. The higher level codes speech information at the syllable level and involves auditory goals, whereas the lower level deals with articulatory feature clusters, roughly corresponding to phonemes, and involves somatosensory goals. It is not clear to us whether the auditory and somatosensory pathways are reserved to one level each. Further research will help better specifying this stage. Meanwhile, we remain agnostic as to whether a parallel or a sequential scheme applies, and as to whether each pathway is linked to a specific speech level or not. In Figure 2, these two pathways are simply labelled as 'a' and 'b', for auditory and somatosensory, respectively. Following suggestions by Indefrey & Levelt (2004), Guenther *et al.* (2006), Hickok (2012) as well as Tian & Poeppel (2013), we posit that the auditory specification activates the left posterior superior temporal gyrus (pSTG) and the superior temporal sulcus (STS), arrow 1a. Following Guenther *et al.* (2006) or Hickok (2012) we further suggest that the parallel somatosensory pathway activates the anterior supramarginal gyrus (aSMG) and the primary somatosensory cortex (S1), arrow 1b.

The next stage in Levelt and colleagues' model is syllabification, which is supposed to operate directly from the phonological code (the desired state) and to be mediated by the left inferior frontal gyrus (Indefrey, 2011). In line with the predictive control account, we suggest that a transformation is first needed, from the desired state expressed in a multisensory format, to commands, expressed in a motor format. This inverse model transformation involves two pathways. The auditory specification is fed to the temporo-parietal junction (TPJ, arrow 2a). The somatosensory specification is sent to the cerebellum (arrow 2b). Activities in the cerebellum has indeed been observed during the execution of a motor task and has been related to the generation of motor commands (Gomi *et al.*, 1998; Grush, 2004; Imamizu & Kawato, 2009; Kawato *et al.*, 1987; Wolpert, Miall, & Kawato, 1998).

This transformation will make it possible for motor programs to be specified, again following two pathways, as in Hickok (2012): the transformed auditory goals are sent from the TPJ to the left inferior frontal gyrus (LIFG) and to the premotor cortex (ventral BA6), arrow 3a; the transformed somatosensory goals are sent from the cerebellum to the lower primary motor cortex (M1), arrow 3b. We add to Hickok (2012) the speculation that the motor programs issued by LIFG are themselves sent to M1 (arrow 4) presumably leading to a unique motor plan, specified in M1, and integrating the two motor programs, from the auditory and the somatosensory pathways.

Articulation is then inhibited, via a signal presumably emitted when the desired state was specified (or even at an earlier stage, during lemma retrieval or conceptual preparation). This signal is probably issued in regions involved in inhibitory control, i.e. in rostral prefrontal cortex (BA 10) and anterior cingulate gyrus (BA 32) (Basho *et al.*, 2007). It may be sent to M1 only (if the assumption that a unique motor program is specified in M1 is correct) or to both LIFG and M1, as suggested in Figure 2. A residual somatosensory feedback may be felt, resulting from attenuated motor commands being sent to the motor system. This may activate the aSMG and S1.

Once motor commands are computed, an efference copy is used by the forward model to simulate a predicted state. We assume that this transformation involves the inverse pathways from the ones used in the transformation from sensory states to motor commands. A similar step is taken by Tian *et al.* (2016). But whereas they assume a sequential pathway, from motor representations in the LIFG to auditory consequence in pSTG and STS via somatosensory consequences in SMG, we stick to the two-pathway scheme. The efference copy mediated by LIFG is sent to the TPJ (arrow 4a) and is transformed into a predicted auditory signal that activates pSTG and STS (arrow 5a). The other copy, in M1, is sent to the cerebellum (arrow 4b) and is transformed into a predicted somatosensory signal that activates aSMG and S1 (arrow 5b). We conjecture that C2 (between predicted and desired states) takes place at two sites, in auditory and somatosensory cortices. This comparison is presumably under the supervision of cognitive control regions, but too little research has been carried out in this field to make any speculation.

Conclusion

Although we are still far from having a complete picture of the nature of inner language, we argue that our integrated approach, in which inner language is conceived of as multimodal acts with multisensory percepts, stemming from coarse multisensory goals, is backed up by data in linguistics, psycholinguistics, and neurolinguistics. Many issues still need to be resolved. First, the dynamics of cerebral activation proposed here still needs empirical evidence. We have claimed for instance that inner language stems from a coarse desired multisensory state that originates from temporal and parietal regions, and is converted into motor commands in the frontal regions. A copy of these motor commands is itself converted back into a predicted multisensory signal, with activations in temporal and parietal regions. A temporo-parieto-fronto-temporo-parietal loop is therefore hypothesized and should be demonstrated. Further research is needed to assess the dynamic pattern of activation and connectivity of the cerebral regions involved in inner word production. Second, we have limited ourselves to word-level production. Further research should examine the additional processes involved in full sentence generation. Third, we have only focused on the later stages of inner language, once conceptual preparation and grammatical encoding have taken place. These early stages should be examined for a full picture of inner language. Fourth, we have mainly focused on wilful inner speech, yet verbal mind wandering is a very frequent inner language instance, which may be related to the experience of inner hearing without feeling in control (Hurlburt, Alderson-Day, Kühn, & Fernyhough, 2016). Better understanding its mechanism would provide important insights into the origin of auditory verbal hallucination. Further examining the fluctuations between involuntary and wilful inner language could also help explaining verbal rumination, an excessive form of negative inner speech, during which supervisory mechanisms seem faulty. Finally, current theories do not provide satisfactory accounts of how cognitive control is unfolded during inner language. Although many of the subcomponents of inner language processes can be associated with specific regions or networks, several stages remain unknown. In particular, it is still unclear which regions process the results of the comparisons supposed to occur in the predictive control account and how cognitive control integrates these outcomes. We are currently carrying out research to explore these issues.

Acknowledgements

This research was funded by the ANR project INNERSPEECH [grant number ANR-13-BSH2-0003-01], <http://lpnc.univ-grenoble-alpes.fr/InnerSpeech>. We thank Anne Vilain, Maëva Garnier, Jean-Philippe Lachaux, Luciano Fadiga, Christopher Moulin, Cédric Pichat, Yanica Klein, Laurent Lamalle, Jean-Luc Schwartz, Irène Troprès for helpful advice. We are grateful to the editors Agustín Vicente and Peter Langland-Hassan for their insightful comments and suggestions on an earlier version of this chapter.

References

- Alderson-Day, B., & Fernyhough, C. (2015). Inner Speech: Development, Cognitive Functions, Phenomenology, and Neurobiology. *Psychological Bulletin*, 141(5), 931-965.
- Alexander, J. D., & Nygaard, L. C. (2008). Reading voices and hearing text: talker-specific auditory imagery in reading. *Journal of Experimental Psychology: Human Perception and Performance*, 34(2), 446-459.
- Anderson, R. E. (1982). Speech imagery is not always faster than visual imagery. *Memory & Cognition*, 10, 371-380.
- Arnaud, L., Schwartz, J.-L., Lœvenbruck, H., & Savariaux, C. (2008). Perception as a (Shaped) Mirror of Action: It Seems Easier to Lipread One's Own Speech Gestures than those of Somebody Else. *Workshop on Speech and Face to Face Communication*, Grenoble, 27-29 Oct. 2008.
- Atkinson, J. R., Gleeson, K., Cromwell, J., & O'Rourke, S. (2007). Exploring the perceptual characteristics of voice-hallucinations in deaf people. *Cognitive neuropsychiatry*, 12, 339-361.
- Baddeley, A., & Wilson, B. (1985). Phonological coding and short-term memory in patients without speech. *Journal of Memory and Language*, 24, 490-502.
- Ballet, G. (1886). Le langage intérieur et les diverses formes de l'aphasie. *Félix Alcan, éditeur, Paris*.
- Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral & Brain Sciences*, 22, 577-609.
- Basho, S., Palmer, E. D., Rubio, M. A., Wulfeck, B., & Müller, R.-A. (2007). Effects of generation mode in fMRI adaptations of semantic fluency: paced production and overt speech. *Neuropsychologia*, 45(8), 1697-1706.
- Belin, P., Zatorre, R., Lafaille, P., Ahad, P., & Pike, B. (2000). Voice-selective areas in human auditory cortex. *Nature, Nature Publishing Group*, 403, 309-312.
- Bellugi, U., Klima, E., & Siple, P. (1975). Remembering in signs. *Cognition*, 3 (2), 93-125.
- Bergounioux, G. (2001). Endophasie et linguistique [Décomptes, quotes et squelette]. *Langue française*, 132, 106-124.
- Blakemore, S. (2003). Deluding the motor system. *Consciousness and Cognition*, 12 (4), 647-655.
- Blakemore, S., Frith, C., & Wolpert, D. (1999). Spatio-temporal prediction modulates the perception of self-produced stimuli. *Journal of Cognitive Neuroscience*, 11(5), 551-559.
- Blakemore, S.-J., Wolpert, D., & Frith, C. (2000). Why can't you tickle yourself? *Neuroreport*, 11(11lemma), R11-R16.
- Blakemore, S.-J., Wolpert, D. M., & Frith, C. D. (2002). Abnormalities in the awareness of action. *Trends In Cognitive Sciences*, 6, 237-242.
- Chapell, M. S. (1994). Inner Speech and Respiration: Toward a Possible Mechanism of Stress Reduction. *Perceptual and Motor Skills*, 79, 803-811.
- Christoff, K., Gordon, A. M., Smallwood, J., Smith, R. & Schooler, J. W. (2009). Experience sampling during fMRI reveals default network and executive system contributions to mind wandering. *Proceedings of the National Academy of Sciences*, 106, 8719-8724.
- Coltheart, M. (2005). Modeling Reading: The Dual-Route Approach. *The science of Reading: a Handbook*, Oxford: Blackwell Publishing, 6-23.
- Conrad, B., & Schönle, P. (1979). Speech and Respiration. *Archiv Für Psychiatrie Und Nervenkrankheiten*, 226, 251-268.
- Corley, M., Brocklehurst, P. H., & Moat, H. S. (2011). Error biases in inner and overt speech: Evidence from tongue twisters. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37, 162-175.
- Decety, J., & Jeannerod, M. (1996). Mentally simulated movements in virtual reality: does Fitt's law hold in motor imagery? *Behavioral Brain Research*, 72, 127-134.

- Dell, G., & Oppenheim, G. M. (2015). Insights for Speech Production Planning from Errors in Inner Speech. *The Handbook of Speech Production*, Redford, M. (Ed.), John Wiley & Sons, West Sussex, UK, 404-418.
- De Ruiter, J. P. (2007). Postcards from the mind: The relationship between speech, imagistic gesture, and thought. *Gesture*, 7(1), 21-38
- Dujardin, E. (1887). Les lauriers sont coupés. Published in instalments in *Revue indépendante*, May-August 1887, Paris (subsequent one-volume edition Paris : éditions Messéin, 1925).
- Dujardin, E. (1931). Le Monologue intérieur, son apparition, ses origines, sa place dans l'œuvre de James Joyce et dans le roman contemporain. Paris : Messéin.
- Egger, V. (1881). La Parole intérieure. Essai de psychologie descriptive. G. Baillièrre, (Ed.), Paris.
- Erdmann, J. E. (1851). *Psychologische Briefe*, Aufl. Leipzig: Reichardt.
- Feinberg, I. (1978). Efference copy and corollary discharge: Implications for thinking and its disorders. *Schizophrenia Bulletin*, 4, 636-640.
- Fernyhough, C. (2004). Alien voices and inner dialogue: towards a developmental account of auditory verbal hallucinations. *New Ideas in Psychology*, 22, 49-68.
- Filik, R., & Barber, E. (2011). Inner speech during silent reading reflects the reader's regional accent. *PLoS One*, 6(10), e25782.
- Fodor, J. A., & Pylyshyn, Z. W. (1988). Connectionism and cognitive architecture: A critical analysis. *Cognition*, 28, 3-7.
- Ford, J., & Mathalon, D. (2004). Electrophysiological evidence of corollary discharge dysfunction in schizophrenia during talking and thinking. *Journal of Psychiatric Research*, 38(1), 37-46.
- Franck, N., & Thibaut, F. (2003). Les hallucinations. *Encyclopédie Médico-Chirurgicale*, 37120 A10.
- Frith, C., Blakemore, S., & Wolpert, D. (2000). Explaining the symptoms of schizophrenia: abnormalities in the awareness of action. *Brain Research Reviews*, 31, 357-363.
- Frith, C. D. (1992). The cognitive neuropsychology of schizophrenia. *Psychology Press*.
- Frith, C. D. (2002). Attention to action and awareness of other minds. *Consciousness and Cognition*, 11(4), 481-487.
- Frith, C. D. (2005). The self in action: Lessons from delusions of control. *Consciousness and Cognition*, 14(4), 752-770.
- Frith, C. D. (2010). What is consciousness for? *Pragmatics & Cognition*, 18(3), 497-551.
- Frith, C. D. (2012). Explaining delusions of control: The comparator model 20 years on. *Consciousness and Cognition*, 21(1), 52-54.
- Gallagher, S. (2000). Philosophical conceptions of the self: implications for cognitive science. *Trends in cognitive sciences*, 4(1), 14-21.
- Gallagher, S. (2004). Neurocognitive models of schizophrenia: a neurophenomenological critique. *Psychopathology*, 37(1), 8-19.
- Gallese, V., & Lakoff, G. (2005). The brain's concepts: The role of the sensory-motor system in conceptual knowledge. *Cognitive neuropsychology*, 22, 455-479.
- Gauvin, H. S., Baene, W. D., Brass, M., & Hartsuiker, R. J. (2016). Conflict monitoring in speech processing: An fMRI study of error detection in speech production and perception. *NeuroImage*, 126, 96-105.
- Geiger, L. (1868). *Ursprung und Entwicklung der menschlichen Sprache und Vernunft*, Verlag der J. G. Cotta'schen Buchhandlung, Stuttgart.
- Geiselman, R., & Glenny, J. (1977). Effects of imagining speakers' voices on the retention of words presented visually. *Memory & Cognition*, 5, 499-504.
- Geva, S., Bennett, S., Warburton, E. A., & Patterson, K. (2011a). Discrepancy between inner and overt speech: Implications for post-stroke aphasia and normal language processing. *Aphasiology*, 25, 323-343.
- Geva, S., Jones, P. S., Crinion, J. T., Price, C. J., Baron, J.-C., & Warburton, E. A. (2011b). The neural correlates of inner speech defined by voxel-based lesion-symptom mapping. *Brain*, 134(10), 3071-3082.
- Gick, B. (2015). Speech production and articulatory phonetics. *Proceedings of the 18th ICPhS, Glasgow, University of Glasgow*.

- Gomi, H., Shidara, M., Takemura, A., Inoue, Y., Kawano, K., & Kawato, M. (1998). Temporal firing patterns of Purkinje cells in the cerebellar ventral paraflocculus duringocular following responses in monkeys I. Simple spikes. *Journal of Neurophysiology*, 80, 818–831.
- Grandchamp, R., Rapin, L., Lœvenbruck, H., Perrone-Bertolotti, M., Pichat, C., Lachaux, J. P., & Baciú, M. (2016). Inner Speech with your own or someone else's voice. Cerebral correlates assessed with fMRI. *Society for the Neurobiology of Language (SNL) Conference 2016*, London 17-20 August 2016.
- Grèzes, J., & Decety, J. (2001). Functional anatomy of execution, mental simulation, observation, and verb generation of actions: A meta-analysis. *Human Brain Mapping*, 12, 1-19.
- Grush, R. (2004). The emulation theory of representation: motor control, imagery, and perception. *Behavioral and brain sciences*, 27(3), 377-396.
- Guenther, F. H., Ghosh, S. S., & Tourville, J. A. (2006). Neural modeling and imaging of the cortical interactions underlying syllable production. *Brain and Language*, 96(3), 280-301.
- Guenther, F. H., & Vladusich, T. (2012). A neural theory of speech acquisition and production. *Journal of Neurolinguistics*, 25, 408 - 422.
- Haggard, P., Newman, C., & Magno, E. (1999). On the perceived time of voluntary actions. *British Journal of Psychology*, 90(2), 291-303.
- Hickok, G. (2012). Computational neuroanatomy of speech production. *Nature Reviews Neuroscience*, 13(2), 135-145.
- Houde, J., & Nagarajan, S. (2011). Speech production as state feedback control. *Frontiers in Human Neuroscience*, 5, 82.
- Houde, J., Nagarajan, S., Sekihara, K., & Merzenich, M. (2002). Modulation of the auditory cortex during speech: An MEG study. *Journal of Cognitive Neuroscience*, 14(8), 1125-1138.
- Hurlburt, R. T. (2011). *Investigating pristine inner experience: Moments of truth*. (Cambridge University Press).
- Hurlburt, R. T., Alderson-Day, B., Kühn, S., & Fernyhough, C. (2016). Exploring the Ecological Validity of Thinking on Demand: Neural Correlates of Elicited vs. Spontaneously Occurring Inner Speech. *Plos One*, 11(2), e0147932. <http://doi.org/10.1371/journal.pone.0147932>
- Imamizu, H., & Kawato, M. (2009). Brain mechanisms for predictive control by switching internal models: implications for higher-order cognitive functions. *Psychological Research PRPF*, 73(4), 527-544.
- Indefrey, P. (2011). The spatial and temporal signatures of word production components: a critical update. *Frontiers in psychology*, 2, 255.
- Indefrey, P., & Levelt, W. (2004). The spatial and temporal signatures of word production components. *Cognition*, 92, 101-144.
- Jacobson, E. (1931). Electrical measurements of neuromuscular states during mental activities. VII. Imagination, recollection, and abstract thinking involving the speech musculature. *American Journal of Physiology*, 97, 200-209.
- Jeannerod, M., & Decety, J. (1995). Mental motor imagery: a window into the representational stages of action. *Current Opinion in Neurobiology*, 5, 727-732.
- Jeannerod, M., & Pacherie, E. (2004). Agency, simulation and self-identification. *Mind & Language*, 19(2), 113-146.
- Johnson, K. (1997). *Acoustic and auditory phonetics*. Blackwell, Oxford.
- Jones, S. R., & Fernyhough, C. (2007). Thought as action: Inner speech, self-monitoring, and auditory verbal hallucinations. *Consciousness and Cognition*, 16, 391-399.
- Jordan, M. I., & Rumelhart, D. E. (1992). Forward models: Supervised learning with a distal teacher. *Cognitive science*, 16(3), 307-354.
- Kawato, M., Furukawa, K., & Suzuki, R. (1987). Hierarchical neural-network model for control and learning of voluntary movement. *Biological cybernetics*, 57 (3), 169-185.
- Kell, C. A., Darquea, M., Behrens, M., Cordani, L., Keller, C., & Fuchs, S. (2017). Phonetic detail and lateralization of reading-related inner speech and of auditory and somatosensory feedback processing during overt reading. *Human Brain Mapping*, 38, 493-508.
- Korba, R. J. (1990). The Rate of Inner Speech. *Perceptual and Motor Skills*, 71, 1043-1052.

- Kurby, C. A., Magliano, J. P., & Rapp, D. N. (2009). Those voices in your head: Activation of auditory images during reading. *Cognition, Elsevier*, 112 (3), 457-461.
- Lackner, J. R., & Tuller, B. H. (1979). Role of efference monitoring in the detection of self-produced speech errors. In W. E. Cooper, & E. C. T. Walker (Eds.), *Sentence processing: psycholinguistic studies presented to Merrill Garret*. Hillsdale, NJ: Lawrence Erlbaum.
- Landauer, T. K. (1962). Rate of implicit speech. *Perceptual and motor skills*, 15, 646-646.
- Langland-Hassan, P. (2008). Fractured phenomenologies: Thought insertion, inner speech, and the puzzle of extraneity. *Mind & Language*, 23(4), 369-401.
- Langland-Hassan, P., Faries, F. R., Richardson, M. J., & Dietz, A. (2015). Inner speech deficits in people with aphasia. *Frontiers in Psychology*, 6, 1-10.
- Levelt, W. J. M. (1989). *Speaking: from intention to articulation*. Cambridge, MA: MIT Press.
- Levelt, W. J. M. (1994). The skill of speaking. In Bertelson P., Eelen P., d'Ydewalle G. (Eds.), *International Perspectives on Psychological Science Vol. 1: Leading themes*, Hillsdale, NJ: Lawrence Erlbaum Associates, 89-103.
- Levelt, W. J. M., Roelofs, A., & Meyer, A. (1999). A theory of lexical access in speech production. *Behavioral and brain sciences*, 22, 1-38.
- Levine, D. N., Calvanio, R., & Popovics, A. (1982). Language in the absence of inner speech. *Neuropsychologia*, 20(4), 391-409.
- Libet, B., Gleason, C. A., Wright, E. W., & Pearl, D. K. (1983). Time of conscious intention to act in relation to onset of cerebral activity (readiness potential): The unconscious initiation of a freely voluntary act. *Brain*, 106(3), 623-642.
- Livesay, J., Liebke, A., Samaras, M., & Stanley, A. (1996). Covert speech behavior during a silent language recitation task. *Perception and Motor Skills*, 83, 1355-1362.
- Lœvenbruck, H., Baciú, M., Segebarth, C., & Abry, C. (2005). The left inferior frontal gyrus under focus: an fMRI study of the production of deixis via syntactic extraction and prosodic focus. *Journal of Neurolinguistics*, 18(3), 237-258.
- MacKay, D. G. (1981). The problem of rehearsal or mental practice. *Journal of motor behavior*, 13, 274-285.
- MacKay, D. G. (1992). Constraints on theories of inner speech. In D. Reisberg (Ed.), *Auditory Imagery*, NJ/England: Erlbaum, 121-49.
- MacSweeney, M., Capek, C. M., Campbell, R., & Woll, B. (2008). The signing brain: the neurobiology of sign language. *Trends in Cognitive Sciences*, 12, 432 - 440.
- Marshall, P. H., & Cartwright, S. A. (1978). Failure to replicate a reported implicit-explicit speech equivalence. *Perceptual and Motor Skills*, 46, 1197-1198.
- Marshall, P. H., & Cartwright, S. A. (1980). A final (?) note on implicit/explicit speech equivalence. *Bulletin of the Psychonomic Society*, 15, 409-409.
- Marshall, R., Rappaport, B., & Garcia-Bunuel, L. (1985). Self-monitoring behavior in a case of severe auditory agnosia with aphasia. *Brain and Language*, 24, 297-313.
- Martin, R. C., & Caramazza, A. (1982). Short-term memory performance in the absence of phonological coding. *Brain and Cognition*, 1, 50 - 70.
- Martínez-Manrique, F., & Vicente, A. (2015). The activity view of inner speech. *Frontiers in psychology*, 6, 232, doi: 10.3389/fpsyg.2015.00232
- Max, L. W. (1937). Experimental study of the motor theory of consciousness. IV. Action-current responses in the deaf during awakening, kinaesthetic imagery and abstract thinking. *Journal of Comparative Psychology*, 24, 301.
- McGuigan, F. J., & Dollins, A. B. (1989). Patterns of covert speech behavior and phonetic coding. *Pavlovian Journal of Biological Science*, 24, 19-26.
- McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, 264, 746-748.
- Miall, R., & Wolpert, D. (1996). Forward models for physiological motor control. *Neural networks*, 9(8), 1265-1279.
- Miall, R. C., Weir, D. J., Wolpert, D. M., & Stein, J. F. (1993). Is the cerebellum a Smith Predictor? *Journal of Motor Behaviour*, 25, 203-216.

- Nahorna, O., Berthommier, F., & Schwartz, J.-L. (2015). Audio-visual speech scene analysis: characterization of the dynamics of unbinding and rebinding the McGurk effect. *The Journal of the Acoustical Society of America*, 137 (1), 362-377
- Nalborczyk, L., Perrone-Bertolotti, M., Baeyens, C., Grandchamp, R., Polosan, M., Spinelli, E., ... & Løevenbruck, H. (2017). Orofacial electromyographic correlates of induced verbal rumination. *Biological Psychology*, 127, 53-63.
- Netsell, R., Kleinsasser, S., & Daniel, T. (2016). The Rate of Expanded Inner Speech During Spontaneous Sentence Productions. *Perceptual and Motor Skills*, 123, 383–393.
- Newell, A., & Simon, H. A. (1972). *Human problem solving*. New-York: Prentice-hall.
- Numminen, J., & Curio, G. (1999). Differential effects of overt, covert and replayed speech on vowel-evoked responses of the human auditory cortex. *Neuroscience Letters*, 272(1), 29 – 32.
- Oppenheim, G. M., & Dell, G. S. (2008). Inner speech slips exhibit lexical bias, but not the phonemic similarity effect. *Cognition*, 106, 528-537.
- Oppenheim, G. M., & Dell, G. S. (2010). Motor movement matters: the flexible abstractness of inner speech. *Memory & Cognition*, 38(8), 1147–1160.
- Ossandon, T., Jerbi, K., Vidal, J. R., Bayle, D. J., Henaff, M.-A., Jung, J., Minotti, L., Bertrand, O., Kahane, P., & Lachaux, J.-P. (2011). Transient suppression of broadband gamma power in the default-mode network is correlated with task complexity and subject performance. *The Journal of Neuroscience*, 31(41), 14521–14530.
- Pacherie, E. (2008). The phenomenology of action: A conceptual framework. *Cognition*, 107(1), 179-217.
- Palmer, E. D., Rosen, H. J., Ojemann, J. G., Buckner, R. L., Kelley, W. M., & Petersen, S. E. (2001). An event-related fMRI study of overt and covert word stem completion. *Neuroimage*, 14(1), 182–193.
- Paulesu, E., Frith, C. D., & Frackowiak, R. S. (1993). The neural correlates of the verbal component of working memory. *Nature*, 362, 342-345.
- Paulhan, F. (1886). Le langage intérieur et la pensée. *Revue Philosophique de la France et de l'Étranger*, PUF, 21, 26-58.
- Perrone-Bertolotti, M., Grandchamp, R., Rapin, L., Baciú, M., Lachaux, J. P., & Løevenbruck, H. (2016). Langage intérieur. In Pinto, S. & Sato, M. (eds.). *Traité de Neurolinguistique : Du cerveau au langage*. Louvain-la-Neuve : De Boeck Supérieur, Collection Neuropsychologie, 109–24.
- Perrone-Bertolotti, M., Kujala, J., Vidal, J. R., Hamame, C. M., Ossandon, T., Bertrand, O., ... & Lachaux, J.-P. (2012). How silent is silent reading? Intracerebral evidence for top-down activation of temporal voice areas during reading. *J Neurosci*, 32(49), 17554–17562.
- Perrone-Bertolotti, M., Rapin, L., Lachaux, J. P., Baciú, M., & Løevenbruck, H. (2014). What is that little voice inside my head? Inner speech phenomenology, its role in cognitive performance, and its relation to self-monitoring. *Behavioural Brain Research*, 261, 220–239.
- Postma, A. (2000). Detection of errors during speech production: A review of speech monitoring models. *Cognition*, 77, 97-132.
- Postma, A., & Noordanus, C. (1996). Production and detection of speech errors in silent, mouthed, noise-masked, and normal auditory feedback speech. *Language and Speech*, 39, 375-392.
- Pulvermüller, F., & Fadiga, L. (2010). Active perception: sensorimotor circuits as a cortical basis for language. *Nature Reviews Neuroscience*, 11, 351-360.
- Raichle, M. E. (2010). Two views of brain function. *Trends in Cognitive Sciences*, 14(4), 180–190.
- Rapin, L., Dohen, M., & Løevenbruck, H. (2016). Les hallucinations auditives verbales. In Pinto, S. & Sato, M. (eds.). *Traité de Neurolinguistique : Du cerveau au langage*. Louvain-la-Neuve : De Boeck Supérieur, Collection Neuropsychologie, 347-370.
- Rapin, L., Dohen, M., Polosan, M., Perrier, P., & Løevenbruck, H. (2013). An EMG study of the lip muscles during covert auditory verbal hallucinations in schizophrenia. *Journal of Speech, Language, and Hearing Research*, 56(6), S1882-S1893.
- Reisberg, D., Smith, J., Baxter, D., & Sonenshine, M. (1989). "Enacted" auditory images are ambiguous; "pure" auditory images are not. *The Quarterly Journal of Experimental Psychology*, 41, 619-641.

- Rochet-Capellan, A., Richer, L., & Ostry, D. J. (2012). Nonhomogeneous transfer reveals specificity in speech motor learning. *Journal of Neurophysiology*, 107(6), 1711-1717.
- Rochet-Capellan, A., & Schwartz, J.-L. (2007). An articulatory basis for the labial-to-coronal effect:/pata/seems a more stable articulatory pattern than/tapa. *The Journal of the Acoustical Society of America*, 121, 3740-3754.
- Saffran, E. M. & Marin, O. S. M. (1977). Reading without phonology: Evidence from aphasia. *Quarterly Journal of Experimental Psychology*, 29, 515-525.
- Sato, M., Baciú, M., Lœvenbruck, H., Schwartz, J.-L., Cathiard, M.-A., Segebarth, C., & Abry, C. (2004). Multistable representation of speech forms: a functional MRI study of verbal transformations. *Neuroimage*, 23(3), 1143-51.
- Scott, M. (2013). Corollary Discharge Provides the Sensory Content of Inner Speech. *Psychological Science*, 24, 1824-1830.
- Scott, M., Yeung, H. H., Gick, B., & Werker, J. F. (2013). Inner speech captures the perception of external speech. *The Journal of the Acoustical Society of America*, 133, EL286-EL292.
- Seal, M. L., Aleman, A., & McGuire, P. K. (2004). Compelling imagery, unanticipated speech and deceptive memory: neurocognitive models of auditory verbal hallucinations in schizophrenia. *Cognitive Neuropsychiatry*, 9(1-2), 43-72.
- Shergill, S. S., Brammer, M. J., Fukuda, R., Bullmore, E., Amaro, E., Murray, R. M., & McGuire, P. K. (2002). Modulation of activity in temporal cortex during generation of inner speech. *Human Brain Mapping*, 16(4), 219-227.
- Shimizu, A., & Inoue, T. (1986). Dreamed speech and speech muscle activity. *Psychophysiology*, 23, 210-4.
- Shuster, L. I., & Lemieux, S. K. (2005). An fMRI investigation of covertly and overtly produced mono-and multisyllabic words. *Brain and Language*, 93(1), 20-31.
- Smadja, S. (in press). *La Recherche en littérature : approches stylistiques du monologue intérieur*. Paris: Hermann.
- Smith, B., Hillenbrand, J., Wasowicz, J., & Preston, J. (1986). Durational characteristics of vocal and subvocal speech-implications concerning phonological organization and articulatory difficulty. *Journal of Phonetics*, 14, 265-281.
- Smith, J., Wilson, M., & Reisberg, D. (1995). The role of subvocalization in auditory imagery. *Neuropsychologia*, 33, 1433-1454.
- Smith, S., Brown, H., Toman, J., Googman, L. (1947). The Lack of Cerebral Effects of d-Tubocurarine. *Anesthesiology*, 8(1), 1-14.
- Sokolov, A. (1972). *Inner speech and thought*. Plenum Press New York.
- Stephens, G. L., & Graham, G. (2000). *When Self-Consciousness Breaks: Alien Voices and Inserted Thoughts*. Cambridge, MA: MIT Press.
- Stricker, S. (1885). *Du langage et de la musique*. Traduit de l'allemand par Frédéric Schwiedland, Alcan, Paris.
- Sumbly, W. H., & Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *The Journal of the Acoustical Society of America*, 26, 212-215.
- Swiney, L., & Sousa, P. (2014). A new comparator account of auditory verbal hallucinations: how motor prediction can plausibly contribute to the sense of agency for inner speech. *Frontiers in Human Neuroscience*, 8, 72-86.
- Taine, H. (1870). *De l'intelligence*, 2 vols. Paris: Hachette.
- Tian, X., & Poeppel, D. (2012). Mental imagery of speech: linking motor and perceptual systems through internal simulation and estimation. *Frontiers in Human Neuroscience*, 6, 152-162.
- Tian, X., & Poeppel, D. (2013). The Effect of Imagination on Stimulation: The Functional Specificity of Efference Copies in Speech Processing. *Journal of Cognitive Neuroscience*, 25 (7), 1020-1036.
- Tian, X., Zarate, J. M., & Poeppel, D. (2016). Mental imagery of speech implicates two mechanisms of perceptual reactivation. *Cortex*, 77, 1 – 12.
- Tremblay, S., Houle, G., & Ostry, D. J. (2008). Specificity of speech motor learning. *Journal of Neuroscience*, 28(10), 2426-2434.
- Tsakiris, M., Schütz-Bosbach, S., & Gallagher, S. (2007). On agency and body-ownership: Phenomenological and neurocognitive reflections. *Consciousness and Cognition*, 16(3), 645-660.

- Vallar, G., & Cappa, S. F. (1987). Articulation and verbal short-term memory: Evidence from anarthria. *Cognitive Neuropsychology*, 4, 55-77.
- Vercueil, L., & Perrone-Bertolotti, M. (2013). Ictal inner speech jargon. *Epilepsy & Behavior*, 27(2), 307-309.
- Vicente, A., & Martínez-Manrique, F. (2016). The Nature of Unsymbolized Thinking. *Philosophical Explorations*, 19, 173-187.
- Vygotski, L. S. (1934/1986). *Thought and Language*. English Translation by Alex Kozulin. The MIT Press, Cambridge, MA: MIT Press.
- Warren, R. M. (1961). Illusory changes of distinct speech upon repetition - the verbal transformation effect. *British Journal of Psychology*, 52, 249-258.
- Watson, J. B. (1919). *Psychology from the standpoint of a behaviorist*. (J. B. Lippincott, Ed.). Philadelphia.
- Weber, R. J., & Bach, M. (1969). Visual and speech imagery. *British Journal of Psychology*, 60, 199-202.
- Weber, R. J., & Castleman, J. (1970). The time it takes to imagine. *Perception & Psychophysics*, 8, 165-168.
- Wilding, J., & White, W. (1985). Impairment of rhyme judgments by silent and overt articulatory suppression. *The Quarterly Journal of Experimental Psychology Section A*, 37, 95-107.
- Wilson, M., & Emmorey, K. (1998). A "word length effect" for sign language: Further evidence for the role of language in structuring working memory. *Memory & Cognition*, 26, 584-590.
- Wolpert, D. (1997). Computational approaches to motor control. *Trends in cognitive sciences*, 1(6), 209-216.
- Wolpert, D., & Kawato, M. (1998). Multiple paired forward and inverse models for motor control. *Neural Networks*, 11, 1317-1329.
- Wolpert, D., Miall, R. C., & Kawato, M. (1998). Internal models in the cerebellum. *Trends in Cognitive Sciences*, 2(9), 338-347.
- Wolpert, D. M., Ghahramani, Z., & Jordan, M. I. (1995). An internal model for sensorimotor integration. *Science*, 269, 1880-1882.
- Yao, B., Belin, P., & Scheepers, C. (2011). Silent reading of direct versus indirect speech activates voice-selective areas in the auditory cortex. *Journal of Cognitive Neuroscience*, 23, 3146-3152.
- Yao, B., Belin, P., & Scheepers, C. (2012). Brain 'talks over' boring quotes: Top-down activation of voice-selective areas while listening to monotonous direct speech quotations. *NeuroImage*, 60, 1832-1842.

Figure captions

Figure 1. Our adaptation of the Predictive Control Account of inner speech. During overt speech, given a desired sensory state, an inverse model computes motor commands that are sent to the motor system, which produces speech movements and sounds. These are then processed by the sensory system, producing an actual sensory experience and resulting in an actual sensory end state. This actual sensory state provides a sense of ownership and can be compared with the desired state (C1) to improve the inverse model. In parallel, an efference copy mechanism takes place, depicted in dashed lines. A forward model predicts the sensory consequences of the motor commands. The predicted sensory feedback (or rather its end state) can be compared with the desired sensory state (C2) to adjust the motor commands, even before the action is executed. In addition, when the two states are close enough, a sense of agency is felt. The predicted sensory feedback, to which a delay is applied, is also compared with the actual sensory feedback (C3), to improve the forward model, and to further contribute to agency. During covert speech, the lines and boxes in light grey are irrelevant. In parallel with the motor commands, inhibitory signals (in red) are sent to the motor system, preventing actual articulator movement from occurring. A residual actual sensory feedback may still be experienced, giving rise to the sense of ownership. The predicted sensory signal computed by the efference copy mechanism yields inner language percepts: the inner voice heard and/or the inner articulation felt and/or the inner sign/gesture seen. Its end product is compared with the desired sensory state (C2) to adjust the motor commands while providing a sense of agency if the two states are sufficiently similar. TPJ, temporo-parietal junction; LIFG, left inferior frontal gyrus; M1, primary motor cortex.

Figure 2. A cerebral landscape of wilful covert word production with one own's voice. Lemma retrieval is handled by the left middle temporal gyrus. The lemma is converted to a lexeme, in a multisensory format, through two pathways, one for auditory representation (a) and one for somatosensory (b) representations. The auditory specification of the desired auditory state activates the left pSTG and STS, arrow 1a. The parallel somatosensory pathway activates the aSMG and S1, arrow 1b. An inverse model transformation then takes place, involving two pathways. The auditory specification is fed to the TPJ, arrow 2a. The somatosensory specification is sent to the cerebellum (arrow 2b). Motor programs are then specified: the transformed auditory goals are sent from the TPJ to the LIFG and to the left ventral premotor cortex, arrow 3a; the transformed somatosensory goals are sent from the cerebellum to the lower M1, arrow 3b. The motor programs issued by LIFG are themselves sent to M1 (arrow 4) integrating the two motor programs computed in the auditory and the somatosensory pathways. Articulation is inhibited, via a signal issued in rostral prefrontal cortex (BA 10) and anterior cingulate gyrus (BA 32) and sent to M1 only, or to both LIFG and M1. A residual somatosensory feedback may be felt (aSMG and S1), resulting from attenuated motor commands being sent to the motor system. The efference copy mediated by LIFG is sent to the TPJ (arrow 4a) and is inverted into a predicted auditory signal, activating pSTG and STS (arrow 5a). The other copy, in M1, is sent to the cerebellum (arrow 4b) and is inverted into a predicted somatosensory signal, activating aSMG and S1 (arrow 5b). C2 (between predicted and original desired states) takes place at two sites, in auditory and somatosensory cortices. pSTG, posterior superior temporal gyrus; STS, superior temporal sulcus; aSMG, anterior supramarginal gyrus; S1, primary somatosensory cortex; TPJ, temporo-parietal junction; LIFG, left inferior frontal gyrus; M1, primary motor cortex.

Figure 1

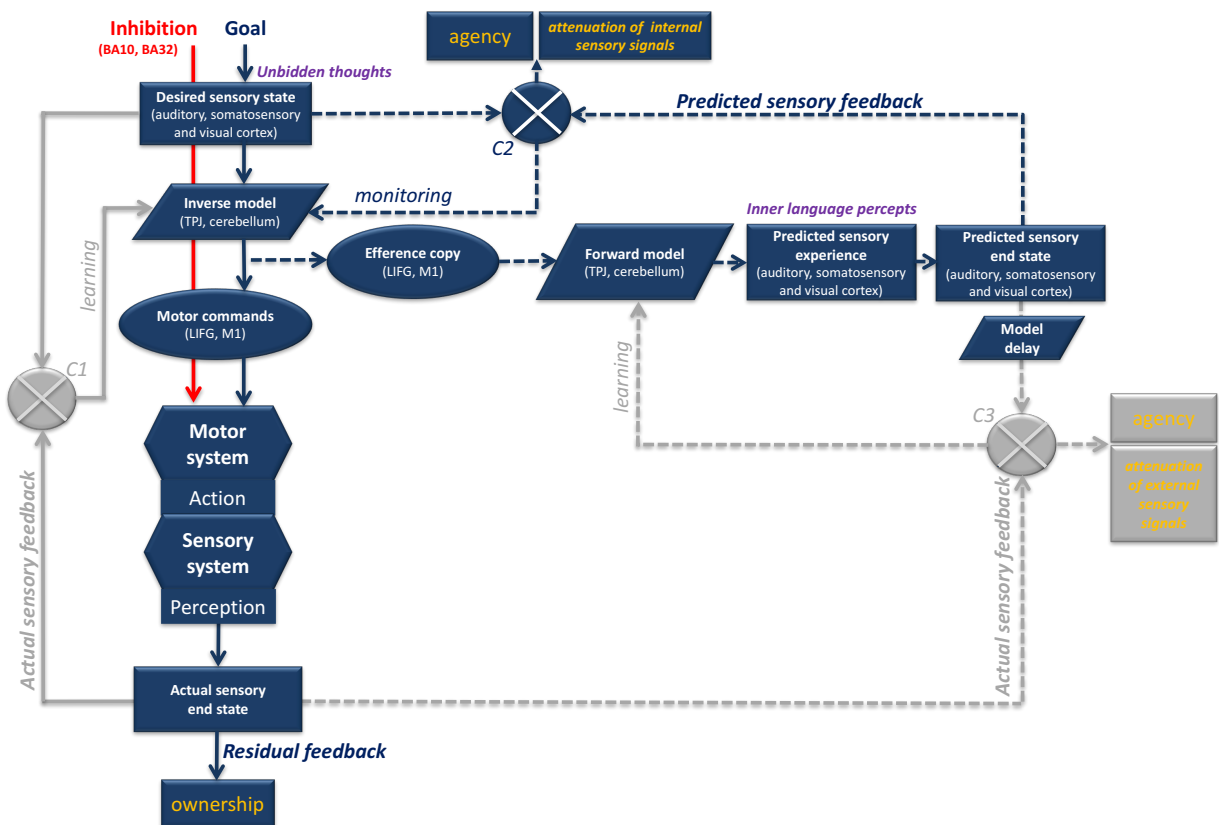
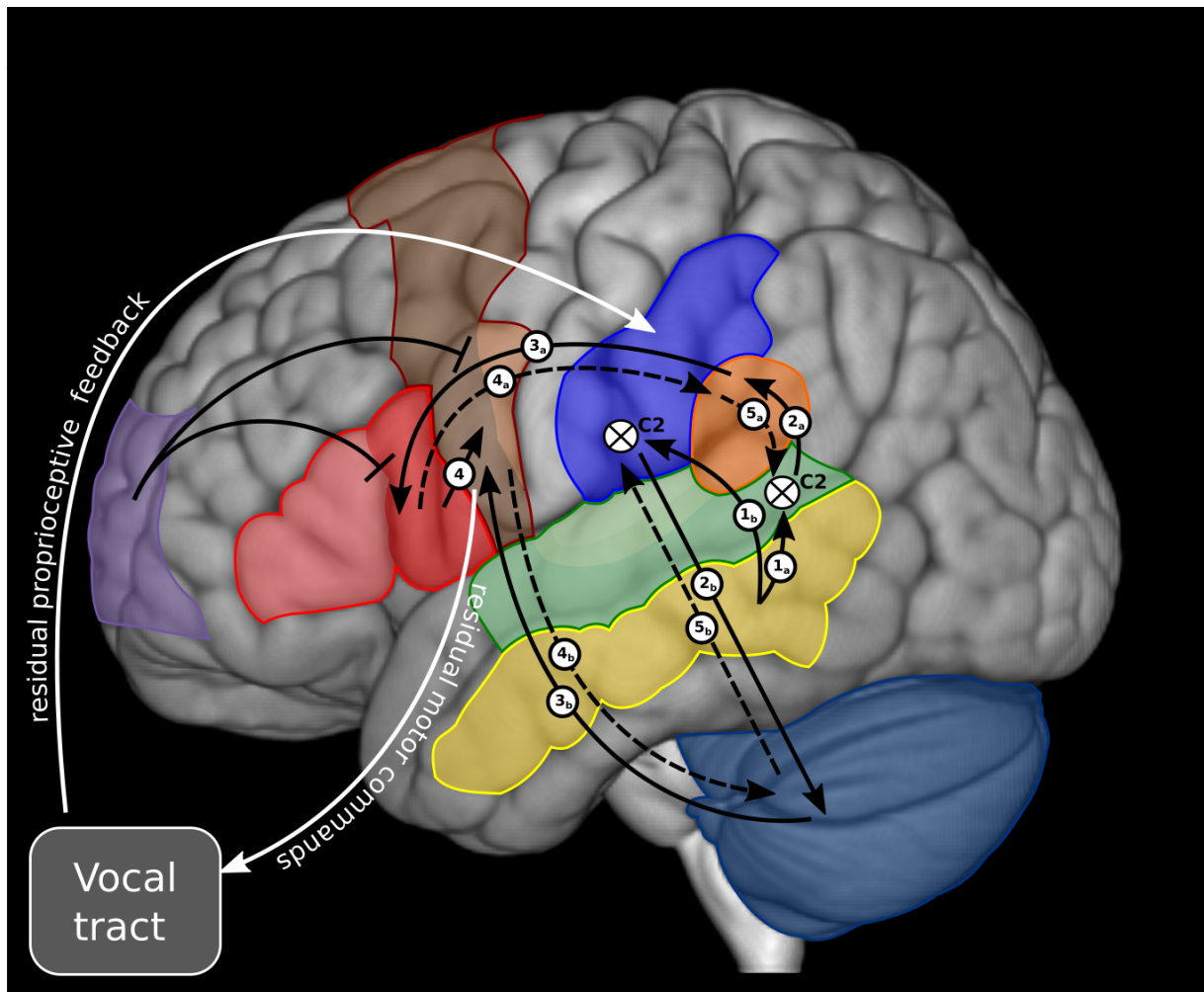


Figure 2



Loquor, ergo comunico-cogito-sum

Résumé – Dans ce mémoire, présenté en vue de l’obtention de l’habilitation à diriger des recherches, je défends l’hypothèse que le langage est le fondement de l’humanité car il remplit trois fonctions majeures : la fonction communicative, pour échanger avec nos congénères, la fonction cognitive, pour élaborer et structurer notre pensée, et la fonction métacognitive d’autoconscience, pour l’élaboration du tandem soi/autrui et la conscience de soi dans le temps. Ces trois fonctions sont intriquées, chacune s’appuyant sur les deux autres. Communication, pensée et autoconscience forment une triade systémique, avec des interactions complexes, qui requiert des formes langagières (phonologiques, lexicales, morphosyntaxiques et prosodiques) multiples et sophistiquées. Ainsi, ce sont les pressions conjuguées de la complexification de la communication, de l’élaboration de la pensée et du renforcement de la conscience de soi, qui ont conduit au langage humain actuel, dans toute sa diversité. Chacune de ces fonctions est toutefois susceptible de perturbation, mettant en danger les deux autres. Il faut donc que les erreurs commises par les systèmes de production et perception du langage soient détectées et éventuellement corrigées, d’une part, et que ces systèmes distinguent les événements auto-générés de ceux produits par autrui, d’autre part. Pour se déployer normalement, la triade communication-pensée-autoconscience requiert ainsi une cybernétique, un mécanisme général de contrôle, qui se décline en plusieurs dispositifs, *monitorings* verbaux et attribution de l’agentivité. Cette cybernétique verbale permet la régulation de notre flux verbal et de ceux d’autrui, pour la communication, le contrôle de nos verbalisations intérieures, pour la pensée, le suivi de soi et d’autrui dans nos constructions narratives autobiographiques, pour l’autoconscience.

Le premier chapitre est consacré des essais de formalisation théorique décrivant certaines des composantes de la cybernétique verbale. J’argumente, d’abord, comment le langage sous-tend les fonctions de communication, pensée et autoconscience. M’appuyant sur de nombreux travaux en psycholinguistique et neurosciences, je décris ensuite, pour chacune de ces trois fonctions, les dispositifs de *monitoring* verbaux et d’attribution de l’agentivité ainsi que leur implémentation neurocognitive hiérarchique et prédictive. Enfin, je propose les premières briques d’un formalisme neuro-computationnel décrivant la cybernétique verbale intégrée qui gouverne la triade communication-pensée-autoconscience. Le deuxième chapitre présente des travaux expérimentaux dont l’objectif est d’observer et tester l’implémentation neuro-computationnelle des composantes de la cybernétique verbale. Je fournis des données comportementales et neurophysiologiques sur la production ou la perception du langage (extériorisé et/ou intérieur), chez l’adulte, au cours du développement chez l’enfant et chez des personnes souffrant de troubles du langage, de l’audition, de pathologies neurologiques ou psychiatriques. Le troisième chapitre propose des applications thérapeutiques, éducatives et technologiques pour la remédiation des troubles du langage. Le dernier chapitre présente le programme de recherche envisagé. Sur le plan théorique, je propose de mieux caractériser les substrats neurophysiologiques de la cybernétique verbale sous-tendant la triade communication-cognition-autoconscience, dans le but de construire un formalisme neuro-computationnel intégré. Sur le plan empirique, je compte poursuivre le recueil de données comportementales et neurophysiologiques permettant d’étayer ces propositions théoriques. Sur le plan appliqué, j’envisage des retombées pour le développement, la remédiation ou la rééducation du langage, gardant à l’esprit que c’est par le langage que nous communiquons, pensons et existons.

Loquor, ergo comunico-cogito-sum

Abstract – In this manuscript, I argue for the hypothesis that language is at the heart of humanity because it serves three major functions: a communicative function, to exchange with our fellow creatures, a cognitive function, to elaborate and structure our thoughts, and a metacognitive function of auto-noesis. These three functions are interlinked, each relying on the other two. Communication, thinking and auto-noesis form a systemic triad, involving complex interactions, and requiring multiple and sophisticated linguistic forms (phonological, lexical, morphosyntactic and prosodic). Human language in its present diverse shape, results from the combined pressures of communication complexification, thinking elaboration and self-awareness enhancement. Each of these functions can be perturbed, however, endangering the other two. The errors committed by the language the production and perception systems need to be detected and potentially corrected. In addition, the language systems need to distinguish between self-generated and other-generated events. To unfold adequately, the communication-thinking-auto-noesis triad requires a cybernetics, i.e. a general control mechanisms, which includes several devices: verbal monitoring and agency attribution. This verbal cybernetics allows for the control of our own verbal flows as well as those of others (during communication), for the monitoring of our inner verbalising (for thinking), and for self- and other-monitoring in our autobiographical narrative constructions (for auto-noesis). The first chapter is dedicated to theoretical formalising, and attempts at describing some of the components of this verbal cybernetics. I first provide arguments for the claim that language underlies communication, thinking and auto-noesis functions. Then, building on many works in psycholinguistics and neuroscience, I describe the verbal monitoring and agency attribution devices, for each of the three functions, as well as their hierarchical neurocognitive and predictive implementation. Finally, I describe the first building blocks for a neurocomputational formalism aimed at describing the integrated verbal cybernetics that govern the communication-thinking-auto-noesis triad. The second chapter provides a description of experimental works carried out with the aim of observing and testing the neurocomputational implementation of some components of the verbal cybernetics. I describe behavioural and neurophysiological data on language production or perception (overt or covert), in adults, in infant or children, and in people with language, hearing, neurological or psychiatric disorders. The third chapter lists a few therapeutic, educative and technological applications of this work, in the domain of general language remediation. The final chapter presents the research program I consider for the future. In terms of theoretical research, I plan to better characterize the neurophysiological substrates of the verbal cybernetics that underlies the communication-thinking-auto-noesis triad, in order to build an integrated neurocomputational formalism. In terms of empirical research, I intend to carry on collecting behavioural and neurophysiological data, to support these theoretical propositions. In terms of applications, I consider potential outcomes in the field of language development, remediation or re-education, keeping in mind that it is thanks to language that we can communicate, think and exist.