Passerelles entre geste et voix chez différentes espèces de primates: Vers une hypothèse multimodale des origines du langage

Marie Bourjade

▶ **To cite this version:**

Marie Bourjade. Passerelles entre geste et voix chez différentes espèces de primates: Vers une hypothèse multimodale des origines du langage. Sciences cognitives. Université Toulouse Jean Jaurès, 2022. tel-03669070

HAL Id: tel-03669070
https://hal.science/tel-03669070

Submitted on 16 May 2022

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.

Passerelles entre geste et voix chez différentes espèces de primates : Vers une hypothèse multimodale des origines du langage

HABILITATION A DIRIGER DES RECHERCHES

Présentée par Dr Marie BOURJADE

Maître de conférences en psychologie

Laboratoire CLLE UMR CNRS 5263

Membre de l'IUF

**

25 mars 2022

Membres du jury

Dr. Corine Astésano, MCF-HDR	Université Toulouse Jean Jaurès
Dr. Alejandrina Cristia, DR (rapporteure)	Ecole Normale Supérieure (Paris)
Dr. Jean-Louis Dessalles, MCF-HDR	Telecom Paris Tech
Pr. Alban Lemasson, PR (rapporteur)	Université de Rennes 1
Pr. Maya Gratier, PR (rapporteure)	Université Paris Nanterre
Pr. Michèle Guidetti, PR émérite (garante)	Université Toulouse Jean Jaurès

Tout un tas de primates

CC BY-NC, Lucie Rétive, https://danslafusee.noblogs.org

Remerciements

Je souhaite remercier Michèle Guidetti d'avoir accepté de superviser ce travail avec enthousiasme. Je lui suis reconnaissante également de co-diriger la thèse de Mawa Dafreville avec moi et d'assurer pour nous deux un soutien institutionnel sans faille. Je remercie Mawa pour sa sympathie et la qualité de son travail qui a fait significativement avancer les propos rapportés ci-après.

Je remercie également les membres du jury d'avoir accepté d'évaluer ce travail et de venir en discuter avec moi lors de la soutenance orale le 25 mars 2022. Je suis reconnaissante envers Alban Lemasson pour le soutien apporté à ce projet de recherche et pour le contact de Shirley Strum qui s'est avéré fructueux, comme en témoignent les pages qui suivent.

J'ai obtenu mon poste de maître de conférences à Toulouse en septembre 2014 et je suis reconnaissante de l'accueil chaleureux et du soutien que m'ont réservé les collègues d'ex-ECCD et sympathisantes. Merci à Michèle de nouveau, Valérie Tartas, Olga Volckaert-Legrier, Hélène Cochet, Sören Frappart, et Lucie Bouvet. Les échanges scientifiques que nous avons menés au cours de mes premières années à Toulouse ont largement contribué à mon acculturation à la psychologie et aux sciences humaines. Merci également à Fabio Montermini de m'avoir accordé sa confiance pour explorer, au-delà du connu, des gestes de babouins 'comme' les mots d'une langue.

Enfin, je souhaite remercier chaleureusement Luke Glowacki, Sören Frappart et Stéphane Vautier qui ont, chacun à leur manière, contribué à infléchir ma ligne scientifique, et m'ont aidé à retrouver la biologiste qui est en moi. Je les remercie pour leur amitié et pour le plaisir partagé de nos longues conversations scientifiques.

Cette note de synthèse a été rédigée au hameau de Jaqués en Ariège, avec la présence intermittente de Lucie qui dessine dans mon dos un après-midi par semaine et sous le regard circonspect de Ben. Merci à Lucie d'avoir dessiné tout un tas de primates ! Et un grand merci à Ben pour sa patience et sa bienveillance à l'égard de mes accaparements académiques.

Sommaire

Chapitre 1. Présentation d'une approche interdisciplinaire 11
1.1. Introduction et présentation de la note de synthèse 11
1.2. Positionnement scientifique 13
1.3. Itinéraire scientifique interdisciplinaire et spécificités de l'approche 16
1.3.1. L'approche éco-éthologique du développement social 16
1.3.2. La cognition intentionnelle dans une perspective d'anthropologie évolutive 18
1.3.3. La psychologie comparée dans une perspective développementale 19
1.3.4. Répondre aux enjeux de l'interdisciplinarité 21
1.4. Conceptions du langage, de la communication et des signaux communicatifs 25
1.4.1. Le langage 25
1.4.2. La communication 28
1.4.3. Les signaux de communication 30
Chapitre 2. Origines évolutives et développementales du langage 33
2.1. Etat de l'art sur les origines du langage 33
2.1.1. Modèles théoriques 33
2.1.1.1. Les modèles lexicaux 34
2.1.1.2. Les modèles gestuels 36
2.1.1.3. Les modèles musicaux 39
2.1.1.4. Les modèles multimodaux 41
2.1.1.5. Les apports de la paléontologie 44
2.1.1.6. Synthèse 45
2.1.2. Sources d'évidences vocales, gestuelles, multimodales chez les PNH 46
2.1.2.1. La communication vocale des primates non humains 47
2.1.2.2. La communication gestuelle des primates non humains 51
2.1.2.3. La communication multimodale des primates non humains 57
2.2. Les apports d'une psychologie évolutionnaire, développementale et comparée 60
2.2.1. La fonction adaptative du développement 60
2.2.2. Les conceptions *evo-devo* appliquées au langage 62
2.2.2.1. Canalisation épigénétique de l'acquisition du langage 62
2.2.2.2. Linguistique développementale évolutionnaire 64
2.2.2.3. Les apports du cas particulier de l'extinction des langues 67
2.2.3. La pragmatique développementale dans une perspective comparée 68

2.2.4. Origines développementales du langage humain 71
2.2.4.1. La communication dyadique du nourrisson 72
2.2.4.2. Emergence des interactions triadiques et de l'intentionnalité 74
2.2.4.3. Multimodalité et sensibilité à l'attention 78
2.3. Origines du couplage intention – modalité (vocale) de communication 81
2.3.1. Le problème de la transition des modalités et le contrôle vocal 81
2.3.2. Etudier la flexibilité du couplage intention – modalité de communication 84
2.3.3. Opérationnalisation des construits 86
2.3.3.1. La communication intentionnelle 86
2.3.3.2. L'attention sociale 87
2.3.3.3. La sensibilité à l'attention d'autrui 89
2.4. Les apports d'une approche comparative 91
2.4.1. Des inférences évolutives à partir des comparaisons entre espèces ? 91
2.4.2. La nécessaire prise en compte de l'évolution culturelle chez l'humain 95
Chapitre 3. La communication gestuelle multimodale 98
3.1. Approche expérimentale de la communication intentionnelle 98
3.1.1. Le paradigme de requête alimentaire 98
3.1.2. Sensibilité à l'attention d'autrui 100
3.2. Effet de l'expérience individuelle sur l'acquisition de l'intentionnalité 101
3.2.1. Coordination des gestes et des regards 102
3.2.2. Acquisition des comportements d'attention conjointe 103
3.2.3. Développement de la reconnaissance de l'attention visuelle 104
3.3. Intentionnalité de la communication intraspécifique 106
3.3.1. Répertoire de gestes communicatifs 106
3.4. Importance des ruptures attentionnelles et de l'accès au visage 107
3.4.1.1. Synthèse des travaux sur la communication gestuelle multimodale 107
3.4.1.2. La communication sensible à l'attention dans la littérature 109
Chapitre 4. Les passerelles entre geste et voix 112
4.1. Axe 1 : Rupture attentionnelle et communication multimodale 113
4.1.1. Les modèles théoriques de conflit attentionnel 113
4.1.1.1. L'hypothèse de la communication sensible à l'attention 115
4.1.1.2. Prédictions testables 117
4.1.2. Projet 1.1. Développement d'une communication sensible à l'attention 118

4.1.2.1. La comparaison entre espèces 119
4.1.2.2. La comparaison entre populations humaines 120
4.1.3. Projet 1.2. Approche sémiotique du couplage Signe - Référent 122
4.1.3.1. L'attention conjointe multimodale en situation écologique 123
4.1.3.2. Corrélats neuronaux de l'attention conjointe multimodale 124
4.1.3.3. Sémiotique expérimentale des jeux de coordination 125
4.1.3.4. Grammaticalisation des signes visuels et auditifs d'un langage émergent 126
4.2. Axe 2 : Linguistique animale à l'interface geste-voix 127
4.2.1. Projet 2.1. Modélisation phonologique de la communication multimodale 128
4.2.1.1. Unités motrices des structures linguistiques 129
4.2.1.2. Modélisation phonotactique des corpus de menace et de salutation 131
4.2.2. Projet 2.2. Sémantique et pragmatique des conflits sociaux 132
4.2.2.1. Le dictionnaire du geste et de la voix chez le babouin 133
4.2.2.2. La pragmatique du conflit social ou comment négocier de l'aide 134
4.3. De la mathématisation au translationnel : perspectives à long terme 135
4.3.1. Mathématisation de l'évolution du langage 136
4.3.2. L'économie de l'attention dans le domaine de la santé 137
4.3.2.1. La dépression post-partum sous l'angle évolutionnaire 138
4.3.3. Conclusion 139
Chapitre 5. Bibliographie 140

Avant-propos à l'attention du lecteur

Usage des sigles dans ce document :

Afin de faciliter la lecture, les sigles les plus couramment utilisés sont répertoriés ci-dessous avec leur déclinaison.

ACM : Attention conjointe multimodale

ASO : *Apparently satisfying outcome*

ASSH: *Attention Sensitive Signaling Hypothesis*

DAC : Dernier ancêtre commun

DPP: Dépression post-partum

FLB : *Faculty of language in the broad sense*

FLN : *Faculty of language in the narrow sense*

MA : Millions d'années

PNH : Primates non humains

Usage de l'italique dans ce document :

La mise en italique est utilisée comme emphase dans le discours et pour l'usage des termes anglais. Ces derniers sont alternativement mis entre parenthèses quand un terme existe en français mais qu'il semble sous-spécifié, ou conservé directement dans le texte quand il n'existe pas de traduction satisfaisante.

Chapitre 1. Présentation d'une approche interdisciplinaire

1.1. Introduction et présentation de la note de synthèse

Cette note de synthèse en vue de l'obtention de l'Habilitation à Diriger des Recherches (HDR) porte sur mes recherches passées, réalisées en postdoctorat à partir de 2011, ainsi que sur mes recherches plus récentes réalisées depuis que je suis maître de conférences à Toulouse, mais aussi sur celles complètement actuelles, et enfin sur celles que je souhaite développer dans le futur. Je désigne cet ensemble de travaux comme « mes travaux de recherche » ou « cette recherche » car il forme de mon point de vue un tout cohérent, un questionnement général. Je parlerai de « cette note de synthèse » pour délimiter l'application des décisions au contenu de ce document (par exemple concernant l'usage de la définition d'un terme). Enfin, je distinguerai la temporalité de ces travaux en désignant les recherches passées par « mes travaux antérieurs », mes recherches actuelles et futures par « mon programme de recherche » qui se trouve être structuré en deux « axes » avec des « projets », et des « lignes de recherche ».

Le propos rapporté dans ce document concerne mon questionnement autour des origines du langage et de l'approche comparative. Il est centré sur l'usage des modalités gestuelles/visuelles et vocales/audibles en fonction des dynamiques attentionnelles et d'éléments contextuels plus larges. Dans les quatre chapitres qui constituent cette note, j'aborderai la présentation de mon positionnement scientifique et de mon parcours interdisciplinaires, avec les enjeux et problèmes de l'interdisciplinarité (Chapitre 1), puis une présentation théorique des modèles d'évolution du langage en articulation avec les apports de la psychologie du développement et la problématisation de *cette recherche* (Chapitre 2), une synthèse de *mes travaux antérieurs* réalisés sur la communication gestuelle du babouin olive en captivité (Chapitre 3), et une présentation de *mon programme de recherche* sur l'exploration des passerelles entre le geste et la voix avec des projets en cours et à venir (Chapitre 4).

Le Chapitre 1 s'ouvre sur une présentation succincte de mon positionnement scientifique permettant de délimiter le sujet et sa problématique centrale. Je présente ensuite rapidement mon parcours interdisciplinaire en soulignant les apports que j'ai glanés au cours de mes expériences scientifiques assez diverses. J'introduis ce qui selon moi, représente des obstacles concrets à l'interdisciplinarité et je propose de menues solutions permettant de les dépasser à l'échelle de mes travaux de recherche. Je termine ce chapitre introductif par des définitions

théoriques des concepts de langage, communication et signaux de communication, qui se trouvent au cœur des investigations rapportées ici.

Dans le (long !) chapitre 2, je présente un état de l'art des théories et évidences dont nous disposons concernant les ressemblances entre le langage humain et la communication animale. Le point de départ est de balayer les 'familles' de modèles théoriques telles que proposées par Fitch (2010), puis d'en souligner les manques et les complémentarités. Je dresse ensuite un état des connaissances des études sur la communication des primates qui peuvent enrichir les modèles d'évolution du langage sur les aspects gestuels, vocaux et orofaciaux. J'explique en suivant les apports d'une approche développementale, comparée et évolutionnaire[1] (*developmental and comparative evolutionary psychology*) à l'étude de l'évolution du langage, en rappelant les fonctions adaptatives du développement, et en les illustrant avec des exemples sur le langage. J'introduis par la suite la problématisation de mes travaux de recherche par la présentation du problème dit 'de la transition des modalités' soulignant les lacunes des modèles existants pour rendre compte de l'intégration des modalités gestuelles et vocales à l'usage commun du langage dans la vie de tous les jours (Bottineau, 2018; Vigliocco et al., 2014). Je proposerai une manière de dépasser ce problème en présentant mon programme de recherche. Je termine ce chapitre par la présentation de l'approche comparative et par une réflexion sur la nécessité de prendre en compte l'évolution culturelle dès lors que l'on travaille sur l'humain.

Le chapitre 3 est une synthèse de mes recherches postdoctorales sur la communication gestuelle intentionnelle des babouins olive de la Station de Primatologie de Rousset (UPS 846, CNRS). Je présente l'approche expérimentale sur laquelle ces travaux reposent, en introduisant les études antérieures ayant utilisé le paradigme de requête alimentaire. Je rapporte ensuite mes résultats portant sur la communication sensible à l'attention en situation expérimentale. Je continue cet exposé par la série d'expériences que j'ai menée pour examiner l'effet de l'expérience individuelle dans l'acquisition de la communication intentionnelle et des comportements d'attention conjointe chez le babouin olive. Ces travaux mettent en lumière que la communication gestuelle est résolument multimodale en particulier dans les situations de

[1] Le terme *évolutionnaire* est employé de manière synonyme à *évolutive*, en particulier quand le terme « évolutif » ou « évolutive » pourrait être compris dans une acception large, i.e., pour faire mention de changements au cours du temps. Le terme *évolutionniste* n'est employé qu'une fois à la fin du manuscrit pour désigner les chercheurs intéressés par l'évolution. Enfin, j'emploie de terme « développement » pour faire référence aux changements qui surviennent au cours de la vie des individus et « évolution » pour faire référence aux changements qui surviennent au cours des temps géologiques, i.e., l'évolution biologique.

rupture attentionnelle, c'est-à-dire lors des interruptions d'attention visuelle de la part du receveur.

Le chapitre 4 présente mon programme de recherche comportant des travaux en cours et de nombreux projets centrés sur deux passerelles entre la communication gestuelle-visuelle et orale/vocale-auditive. Le premier axe vise à étudier les effets des ruptures attentionnelles sur les modalités de communication de l'émetteur. J'énonce les théories du conflit attentionnel qui peuvent être utilisées pour rendre compte de l'émergence d'un protolangage multimodal. Je présente ensuite les deux projets principaux de cet axe centrés respectivement sur les modalités de communication de l'enfant selon l'attention de sa mère, et sur les modalités de l'attention conjointe en lien avec le développement langagier ultérieur. Le deuxième axe de recherche propose d'examiner l'émergence de structures linguistiques avec l'approche originale d'un modèle animal et d'un cadre sémiotique multimodal. Je présente dans ce cadre, les deux projets principaux centrés sur la phonologie des séquences de menace du babouin olive et sur la pragmatique des conflits sociaux dans la société très complexe de cette espèce. Ce chapitre se termine sur des perspectives à long terme, plus fondamentales ou plus appliquées, qui s'inscrivent dans la continuité des travaux précédents.

1.2. Positionnement scientifique

Bien que la très grande majorité des langues humaines soient parlées oralement, et ce à l'exception des langues signées, les travaux sur les origines évolutives du langage n'ont pas encore élucidé *l'histoire évolutive du contrôle intentionnel de la voix* (Fitch, 2010; Locke, 2006, 2017). De nombreux modèles théoriques proposent que le langage ait d'abord évolué sous forme de systèmes gestuels auxquels se seraient progressivement ajoutées des vocalisations (Levinson & Holler, 2014; Meguerditchian & Vauclair, 2014). Etonnamment, tout autant de modèles théoriques présupposent le contraire, arguant que le langage a nécessairement évolué à partir des compétences vocales de nos ancêtres communs avec les autres espèces de primates (Fitch, 2010; Zuberbühler et al., 2011).

Le langage est une manifestation comportementale, cognitive et sociale structurellement complexe et mobilisant tous les niveaux d'organisation du vivant : de la respiration cellulaire à des processus cognitifs complexes comme la planification d'une phrase en double tâche de l'analyse du discours, en passant par des activations neuronales multiples qui sous-tendent ces processus (Fitch, 2010; Levinson, 2016). Si dans les années 1980, les scientifiques tentaient

d'enseigner des langues naturelles à des animaux (e.g., perroquet gris du Gabon, bonobo, chimpanzé: Pepperberg, 2002; Savage-Rumbaugh et al., 1986; Terrace et al., 1979), la science contemporaine a quelque peu délaissé cette approche holistique pour une approche du langage 'multi-composant' (*multicomponent*) ciblant séparément chacune des adaptations et des compétences mobilisées dans l'activité langagière afin d'en retracer les origines et l'histoire évolutives. Cette approche a permis des avancées « buissonnantes » permettant d'identifier différents précurseurs langagiers chez de nombreuses espèces animales, parfois au détriment d'une approche plus intégrative (pour une vision intégrative, voir néanmoins : Prieur et al., 2020). Ainsi, nous avons appris que l'humain possédait un « gène de l'articulation », FOXP2 également présent chez le chimpanzé avec deux mutations en moins (Enard et al., 2002; Mozzi et al., 2016), que le larynx descendu n'était ni une spécificité humaine (Fitch & Reby, 2001), ni une condition nécessaire à la production de voyelles ou de consonnes (Boë et al., 2017; Lameira, 2018)[2], que les chimpanzés possèdent une aire cérébrale analogue à celle de Broca dans le cerveau humain qui est activée par la communication gestuelle et vocale (Taglialatela et al., 2008), que certains oiseaux ont une syntaxe phonologique caractéristique de la double articulation du langage (Berwick et al., 2011), que les chiens font du « *fast-mapping* » en associant de nouveaux mots à de nouveaux objets (Kaminski, Call, & Fischer, 2004), que la communication gestuelle des singes anthropoïdes et non-anthropoïdes est intentionnelle (Bourjade, 2016; Graham et al., 2020; Townsend et al., 2017) et qu'ils possèdent des rudiments de théorie de l'esprit (Lewis & Krupenye, 2021), même si cela reste source de controverses (par exemple: Leavens et al., 2017; Penn & Povinelli, 2013; Tomasello et al., 2005). Ces découvertes, dont la liste est loin d'être exhaustive, illustrent à la fois la vitalité des recherches sur l'évolution du langage et l'urgente nécessité d'en intégrer les différentes composantes, les synergies, les impasses et les manques. A cet égard, l'émergence du contrôle volontaire et de l'usage intentionnel de la voix manque à la connaissance actuelle de l'histoire évolutive du langage (Locke, 2006, 2009) et la controverse entre les origines gestuelles et vocales apparaît être une impasse empêchant d'en élucider les mécanismes (Slocombe et al., 2011).

Mes travaux de recherche visent ainsi à élargir le champ des connaissances sur l'acquisition et les origines du langage sous l'angle original de l'évolution de la *multimodalité*, avec l'objectif de dépasser cette controverse. L'enjeu principal est d'élucider comment les individus

[2] Contrairement à l'hypothèse avancée par le phonéticien Philip Lieberman selon laquelle la langue humaine a évolué pour produire des voyelles et que le larynx descendu permet l'articulation fine (pour une actualisation du débat voir: Lieberman, 2012)

appartenant à différentes espèces aujourd'hui bien vivantes résolvent le problème d'intégration des modalités communicationnelles, visuelles et auditives, afin de proposer des scénarios évolutifs inférentiels sur nos espèces ancêtres (Harmon, 2018). Mes travaux explorent deux passerelles comportementales entre le « geste » et la « voix », alternativement du côté du receveur et du côté du producteur. Chacune constitue un axe de recherche présenté ci-dessous et détaillé dans le Chapitre 4.

Du point de vue du receveur, la gestuelle peut être multimodale, c'est-à-dire qu'il est possible d'en percevoir les composantes visuelles, tactiles ou auditives, comme pour les gestes produisant des sons, (e.g., « frapper des mains » ; ci-après dénommés gestes *sonores*) ou les gestes induisant un contact physique (ci-après dénommés gestes *tactiles*). De même, la parole et plus généralement la communication orale comporte des aspects voisés (e.g., les voyelles, mobilisant les cordes vocales) et des aspects non-voisés (e.g., certaines consonnes et bruits de bouche ne mobilisant pas les cordes vocales, comme un soupir) ainsi que des aspects visuels (e.g., les mouvements articulatoires des lèvres) utilisés dans la compréhension (Driver, 1996; Katz & Mehta, 2015). Ce premier axe de recherche investigue les situations de rupture attentionnelle (essentiellement sur le plan visuel et plus marginalement tactile) de la part du receveur afin d'en quantifier les effets sur l'usage différentiel et/ou combiné des modalités communicationnelles chez le producteur et sur l'attention conjointe. Il s'appuie sur un travail expérimental réalisé en post-doctorat présenté dans le Chapitre 3. Il est étendu par la suite à des observations naturalistes de dyades mère-enfant chez différentes espèces de primates et différentes populations humaines, et à une approche expérimentale spécifique (Cf. chapitre 4).

Du point de vue du producteur, aussi bien la gestuelle que la communication orale sont souvent *multi-dimensionnelles* (Forrester, 2008) dès lors qu'elles mobilisent différentes parties anatomiques de concert. Par exemple, réaliser un geste silencieux avec la main associé à une expression faciale silencieuse est multi-dimensionnel, unimodal (visuel). Elles peuvent également être *multimodal*es, lorsqu'un signal est naturellement encodé dans au moins deux modalités sensorielles (multimodalité physique, ou *signal multimodal,* e.g., parler), ou lorsque plusieurs signaux émis dans des modalités sensorielles différentes sont combinés (multimodalité comportementale ou *combinaison multimodale,* e.g., parler en fronçant les sourcils) (Fröhlich & van Schaik, 2018). Ce deuxième axe de recherche explore les contraintes motrices et exécutives (inhibition, flexibilité) de la gestualité corporelle, manuelle et orofaciale du babouin olive afin d'en découvrir les principes organisateurs et structurants. Pour ce faire, il

combine une approche expérimentale à l'application de formalismes linguistiques qui relèvent de la phonotaxe[3], de la sémantique et de la pragmatique.

Un des objectifs ultimes de cette recherche est de caractériser des trajectoires développementales aux interfaces comportement / cognition / langage qui permettent : (i) d'intégrer les différentes modalités de communication dans un modèle développemental cohérent, (ii) de rendre compte de la dimension *audiovisuelle* du langage humain *ET* de la communication gestuelle et orale des primates non humains (ci-après, PNH) et, (iii) de mieux renseigner les modèles théoriques retraçant l'histoire évolutive des facettes multimodales du langage (Vigliocco et al., 2014; Wilcox & Xavier, 2013).

1.3. Itinéraire scientifique interdisciplinaire et spécificités de l'approche

« On reconnaît sa voie en découvrant les chemins qui s'en éloignent. »

Albert Camus (1986, p.153)

Mon travail de recherche s'appuie sur un parcours interdisciplinaire associant la biologie du comportement en milieu naturel à l'étude des sciences cognitives animales et humaines. A cette base, viennent s'ajouter les influences de la psychologie du développement, de l'anthropologie évolutive et de la bio-psycho-linguistique appelant à une approche résolument *Evo-Devo*[4] des origines évolutives du langage (Locke, 2009).

1.3.1. L'approche éco-éthologique du développement social

J'ai mené mon travail de doctorat sur une population de chevaux de Przewalski (*Equus ferus przewalskii*) qui faisait l'objet d'un programme de conservation et de réintroduction en

[3] Etude de l'enchainement des phonèmes dans une langue. Les transitions courantes, rares ou inexistantes ainsi que les contraintes générales qui s'y rapportent constituent les règles phonotactiques de la langue.

[4] Le champ de *l'Evo-Devo* (i.e., *evolutionary developmental biology*) en biologie réfère à la comparaison des gènes de développement de plusieurs espèces afin d'en inférer les origines évolutives et la parenté. Etendu à une hérédité extra-génétique (par exemple culturelle), l'approche *Evo-Devo* s'intéresse à la plasticité des trajectoires développementales et à l'origine héritée et développée de la variation entre individus (la variation étant le matériau sur lequel opère toute sélection).

Mongolie[5], l'aire historique de l'espèce. Ma formation initiale en biologie, écologie et éthologie m'a permis de réaliser une collecte de données de terrain, basée sur l'observation scientifique du comportement en conditions naturalistes (Altmann, 1974; Martin & Bateson, 2007). L'objectif de ma thèse était de décrire, quantifier et prédire les influences sociales intergénérationnelles (verticales) et intragénérationnelles (horizontales) qui opèrent au sein des familles de chevaux[6]. Je me suis particulièrement intéressée au développement du réseau social du jeune de 1 et 2 ans, en relation avec l'évolution de la proximité et de la relation affiliative à sa mère. J'ai montré que les jeunes chevaux conservent une relation spatiale privilégiée avec leur mère qui reste stable au cours du développement tandis que le profil interactionnel évolue et se diversifie (Bourjade, 2007). J'ai également mis en évidence un effet régulateur des adultes sur les comportements d'agression des jeunes et sur leurs relations spatiales et affiliatives (Bourjade, de Boyer des Roches, et al., 2009). Cet effet régulateur des adultes a été répliqué sur différentes populations (Fureix et al., 2012). Dans les familles naturelles de chevaux de Przewalski où la proportion d'adultes est forte, les jeunes sont quatre fois moins agressifs que dans les familles où elle est la plus faible et ils maintiennent des distances interindividuelles similaires à celles des adultes (Bourjade, de Boyer des Roches, et al., 2009). Dans des conditions expérimentales, l'ajout de chevaux adultes dans des groupes de jeunes induit une baisse du taux d'agression et un développement du réseau social qui se traduit par des préférences plus marquées pour les plus proches voisins (Bourjade et al., 2008).

En résumé, la diversité de partenaires présents autour du jeune cheval représente autant d'opportunités d'apprendre à se comporter dans un environnement socialement complexe. La manière dont le jeune fait progressivement face à cette complexité et les mécanismes susceptibles de la réduire, e.g., garder toujours le même individu comme voisin (Strum & Latour, 1987), constituent des pistes intéressantes pour investiguer l'évolution de la complexité sociale et les structures cognitives sous-jacentes.

En plus de m'avoir offert une excellente formation de terrain, ce travail doctoral m'a permis d'accéder à une première approche du développement social, dont je réutilise bon nombre d'éléments dans mes projets de recherche actuels. En particulier, l'étude des influences sociales

[5] Informations disponibles sur le site de l'association Takh : https://www.takh.org/fr/

[6] Noter que lorsque j'ai démarré ma thèse en 2004, la notion de culture animale ou non-humaine n'était ni populaire ni acceptée comme elle l'est aujourd'hui ; de même la notion d'intentionnalité recevait une adhésion très marginale et ce n'est qu'à partir de 2009 que les premières publications sur la cognition du cheval ont commencé à paraitre.

dans le développement, abordée avec un modèle animal pour lequel l'héritabilité était plus communément pensée en termes génétiques par les modèles sociobiologiques, participe de l'approche '*Evo-Devo*' mise en œuvre dans mes travaux actuels. Il s'agit de reconnaitre des mécanismes de transmission des caractères qui ne soient pas seulement génétiques, mais qui puissent être épigénétiques, comportementaux, ou culturels (Jablonka & Lamb, 2014; Stamps, 2003; West et al., 2003) – et rejoignent ainsi certains modèles théoriques du développement humain (Bard & Leavens, 2014; Bruner, 1975; Vygotsky, 1978). Il s'agit également de reconnaître le caractère créatif du développement, en ce sens qu'il produit une variabilité sur laquelle la sélection naturelle peut ensuite opérer (Ho, 1998; Lerner, 1998). La fonction adaptative du développement sera développée au Chapitre 2.

1.3.2. La cognition intentionnelle dans une perspective d'anthropologie évolutive

Le deuxième axe de ma thèse qui était de caractériser les influences sociales horizontales m'a donné l'opportunité de m'intéresser aux processus de décision de groupe. J'ai examiné les modèles théoriques de leadership et de décisions par consensus dans les déplacements des familles de chevaux (Conradt & Roper, 2005; Couzin et al., 2005). J'ai montré qu'aucun des déplacements collectifs n'était réalisé sous contrainte coercitive, indiquant que chaque membre du groupe prenait, à minima, la décision individuelle de se déplacer (Bourjade & Sueur, 2010). Dans cette perspective, j'ai pu mettre en évidence que les décisions de groupe étaient partagées entre plusieurs individus qui préparaient le départ en indiquant une direction, plus ou moins soutenue par d'autres (Bourjade, Thierry, et al., 2009). Le résultat crucial de ce travail était que la dynamique décisionnelle (nombre de chevaux qui préparent, nombre de directions indiquées, durée de la préparation et durée de l'adhésion au déplacement) était significativement associée à des contextes écologiques précis, dépendants du site d'*arrivée* (Bourjade, Thierry, et al., 2009). Ainsi, les chevaux prennent plus de temps pour décider d'aller plus loin, dans des lieux très spécifiques où ils ont des activités tout aussi spécifiques, e.g., visiter le point d'eau, lécher certaines pierres riches en minéraux, retourner la terre pour manger les racines des plantes grasses situées dans un lieu précis, etc. En résumé, les processus décisionnels entre les membres d'un même groupe sont influencés par la finalité du déplacement ; finalité qui n'est pas encore manifeste sur le plan spatio-temporel lors du processus décisionnel. Ce résultat fascinant a motivé mon étude pour la cognition intentionnelle et c'est la raison pour laquelle j'ai choisi d'étudier par la suite la capacité des animaux à anticiper un évènement futur, à différentes temporalités.

La capacité à simuler une action dans le futur qui soit déconnectée des éléments perceptibles du présent revêt un intérêt particulier dans l'organisation du comportement humain (Schacter et al., 2008). Cette habilité émerge aux alentours de 5 ans chez l'enfant et peut être ensuite mobilisée dans de nombreuses situations sociales, scolaires et de résolution de tâche (Atance & O'Neill, 2001; Barragan-Jason et al., 2018). Ajuster son comportement présent à cette représentation du futur revient à mettre en place un comportement orienté vers un but, c'est-à-dire un comportement intentionnel. Lors de mon post-doctorat à l'IPHC (UMR CNRS 7178) et à l'institut Max Planck d'anthropologie évolutive, j'ai étudié expérimentalement la capacité de cinq espèces de PNH à adopter dans le présent un comportement qui serait bénéfique dans le futur. J'ai utilisé une tâche d'échange d'objets qui comporte une dimension sociale venant ajouter de l'incertitude à l'évènement futur. Ce travail a mis en lumière les difficultés et habiletés des PNH à anticiper une situation future qui ne présente pas d'éléments saillants dans le présent (Bourjade et al., 2012; Bourjade, Call, et al., 2014) : excepté quelques orangs-outangs et bonobos, l'écrasante majorité des chimpanzés, macaques et capucins a échoué à la tâche et à tous ses contrôles qui impliquaient un délai temporel entre les deux actions clefs. Il a été marquant de constater que les chimpanzés échouaient massivement alors qu'ils réussissaient une tâche équivalente dont la dimension sociale (échange d'objet) était remplacée par une tâche d'utilisation d'outils. La dimension sociale/coopérative de la tâche a donc affecté les performances des chimpanzés (Bourjade, Call, et al., 2014; Dufour & Sterck, 2008), corroborant les résultats obtenus par d'autres chercheurs (e.g., Hare & Tomasello, 2004).

Ces premiers travaux sur la cognition des primates m'ont permis de développer de bonnes compétences expérimentales et d'apprendre à obtenir l'adhésion des sujets à une tâche cognitive. J'ai également bénéficié d'un environnement scientifique à l'institut Max Planck qui mêlait anthropologie, psychologie et biologie autour de la question de l'évolution cognitive et cela a par la suite orienté mes recherches vers une approche résolument comparative et interdisciplinaire. Ces premiers résultats sur la cognition des primates ont également contribué à nourrir ma pensée sur l'intentionnalité et ses corrélats comportementaux sur lesquels je continue de travailler aujourd'hui.

1.3.3. La psychologie comparée dans une perspective développementale

Récemment, le babouin olive (*Papio anubis*) est devenu un modèle d'étude intéressant pour ses comportements latéralisés au même titre que le chimpanzé (Hopkins & Vauclair, 2012),

notamment du fait qu'une majorité de droitiers a été recensée dans la population captive élevée à la Station de Primatologie de Rousset (UPS CNRS 846). C'est dans ce contexte que j'ai réalisé une série d'études postdoctorales sur la communication gestuelle du babouin olive. Naturellement, je me suis intéressée à sa nature intentionnelle tout en interrogeant ses bases neurobiologiques sous-jacentes.

Cette préférence pour la main droite concerne en particulier un geste communicatif de menace (i.e., « *hand slap* » : Meguerditchian, Molesti, & Vauclair, 2011 ; Meguerditchian & Vauclair, 2006), de même qu'un geste de requête alimentaire enseigné à certains sujets (Bourjade, Meunier, Blois-Heulin, & Vauclair, 2013 ; Meguerditchian & Vauclair, 2009). Par ailleurs, cette préférence manuelle n'est pas retrouvée pour les tâches de manipulation d'objets suggérant que le caractère communicatif de certains gestes, plus que la motricité de la main en elle-même, pourrait être associé à une dominance hémisphérique gauche, analogue à la dominance hémisphérique observée pour le langage humain (Bourjade et al., 2013). S'il est admis que l'émergence des préférences manuelles a précédé l'apparition du langage (Corballis et al., 2012), le fait que la dominance hémisphérique pour le langage provienne d'une spécialisation hémisphérique pour ses aspects moteurs (i.e., dimension motrice du geste : Forrester, Quaresmini, Leavens, Mareschal, & Thomas, 2013 ; Greenfield, 1991) ou ses aspects communicatifs (i.e., dimension communicative du geste : Vauclair, 2004) fait cependant toujours débat.

Ce travail initial visait à étayer l'hypothèse des origines gestuelles du langage (Corballis, 2003; Hewes, 1973; Hopkins & Vauclair, 2012; Meguerditchian, Cochet, et al., 2011; Vauclair, 2004) qui a ainsi constitué ma porte d'entrée dans ce champ d'étude. Le chapitre 3 propose une synthèse détaillée des travaux portant sur l'usage intentionnel de gestes appris et spontanés, et sur l'effet de l'expérience individuelle dans l'acquisition de cette intentionnalité. Ce travail postdoctoral a été pour moi l'occasion de réaliser que les individus adultes impliqués dans des expériences de cognition constituaient un modèle développemental « idéal » puisqu'il devenait possible de contrôler leur expérience individuelle et par conséquent d'en mesurer les effets à l'écart des synergies développementales traditionnelles avec la maturation, les périodes sensibles et les influences culturelles et écologiques naturellement environnantes. Cette « période » expérimentale m'a également été très utile pour constituer un répertoire de référence des comportements de communication intentionnelle chez le babouin captif. Il est advenu naturellement que je poursuive ces études sur la communication intentionnelle dans un contexte

naturaliste, en y réalisant une micro-analyse du comportement communicatif tenant compte du contexte social et environnemental.

1.3.4. Répondre aux enjeux de l'interdisciplinarité

Aujourd'hui, l'essentiel de mon travail de recherche s'appuie sur les méthodes de l'éthologie et des sciences cognitives que j'utilise pour répondre à des questions relevant de la psychologie du développement, de la psycholinguistique, de la primatologie et/ou de l'anthropologie évolutive. L'enjeu principal est d'élucider comment les individus résolvent le problème d'intégration des modalités communicationnelles, i.e., modalité visuelle, auditive, mais aussi tactile, dans diverses situations sociales. La méthodologie générale consiste ainsi à générer des descriptions quantitatives sur *l'acquisition* des aspects multimodaux du langage et de la communication chez différentes espèces de primates, qui permettent par la suite de mieux informer les modèles théoriques *d'évolution* du langage. L'approche est indiscutablement comparative, entre espèces, entre cultures, mais aussi entre disciplines.

En effet, le champ de l'évolution du langage mobilise de nombreuses disciplines scientifiques poursuivant le même objet d'étude et offrant comme premier défi le soin de disséquer, d'établir, et d'harmoniser les terminologies afin de s'assurer que les mêmes buts scientifiques sont *effectivement* poursuivis. En d'autres termes, les problèmes de terminologie posent aux études interdisciplinaires des défis de validité externe. A cet égard, mon premier geste a été de faire correspondre les littératures de psychologie comparée avec celles de psychologie du développement sur la gestuelle intentionnelle (Bourjade, 2016; Bourjade et al., 2020). La section suivante offre un petit aperçu de la polysémie des termes comme « geste communicatif », « communication » ou « langage ». Les enjeux de terminologie semblent dépassables à condition d'accepter que les processus ou états auxquels ils font référence dans différentes espèces, cultures ou paradigmes puissent ne pas être entièrement recouvrants. Toutefois, cela nécessite de délimiter ce qu'il y a de commun et de différent à l'aide de critères objectifs et utilisables, communiqués dans les études (Bourjade et al., 2020).

Au-delà de la terminologie, les approches comparatives et interdisciplinaires posent des problèmes théoriques. Par exemple, le stade sensori-moteur caractéristique du développement du jeune enfant selon la théorie de Jean Piaget (1936) peut être utilisé pour décrire le développement des schèmes moteurs chez d'autres espèces animales. En revanche, il est plus difficile d'utiliser le formalisme du stade préopératoire et des opérations concrètes pour décrire

les activités égocentriques ou morales chez des espèces animales. De la même manière, s'il est possible de penser les influences intergénérationnelles sur le développement du jeune non-humain (Danchin & Pocheville, 2014; Jablonka & Lamb, 2014; Stamps, 2003; West et al., 2003), il est bien plus délicat d'étudier le rôle médiateur des outils de la culture dans cet achèvement comme le propose la théorie de Lev Vygotski[7] (Vygotsky, 1978). En effet, à la différence de l'espèce humaine, les « protocultures » ou « cultures animales » documentées chez les non-humains reposent essentiellement sur des transmissions immatérielles. La fonction d'outil culturel que pourrait par exemple revêtir une relation sociale nécessite d'interroger au préalable le processus de formation de cette relation, son usage et ses conséquences – un champ d'investigation à développer (Strum, 2012).

Dans ce contexte, il advient qu'une théorie unique ne permet pas d'embrasser le sujet dans son ensemble et que l'articulation des modèles théoriques doit être réalisée de sorte que le cadre s'applique à chaque espèce étudiée, mais également à chaque échelle temporelle : développementale et phylogénétique. En s'appuyant sur la logique du plus petit dénominateur commun, l'objet d'étude de l'éthologie tel que formulé par Nico Tinbergen dans son ouvrage fondateur *Les buts et les méthodes de l'éthologie* (Tinbergen, 1963) offre une première étape satisfaisante. Ainsi, pour comprendre le comportement, le chercheur doit investiguer quatre questions fondamentales que sont : 1) la cause proximale, c'est-à-dire les mécanismes responsables de son occurrence, 2) l'ontogenèse, c'est-à-dire son développement et ses changements au cours de la vie des individus, 3) la cause ultime, c'est-à-dire la fonction adaptative du comportement, ce pour quoi il continue d'être sélectionné et 4) la phylogenèse, c'est-à-dire l'histoire évolutive du comportement au sein d'un phylum ou d'un clade[8], composée des aspects hérités d'espèces ancêtres, des réversions et des caractères dérivés (nouvellement apparus) au sein d'un clade. L'intérêt que revêt indéniablement cette approche est qu'elle se focalise sur les manifestations observables d'individus en interaction avec leur environnement.

[7] Vygotski avec un « i » final en français, avec un « y » final en anglais. J'emploie la version française dans le texte mais mon éditeur de références conserve la version anglaise pour les références bibliographiques.

[8] Un clade est un ensemble d'espèces apparentées issues du même ancêtre commun.

A cette vision centrée sur le comportement[9], s'ajoutent des conceptions contemporaines du développement qui s'inscrivent dans les grands changements conceptuels des théories de développement de la seconde moitié du XXème siècle. A savoir, la reconnaissance et la prise en compte : (i) d'un nombre croissant de niveaux d'organisation du vivant impliqués dans le développement, (ii) du développement « vie entière » (*life-span*), (iii) de l'activité première de l'individu dans son développement, ce qui implique que s'il y a un déterminisme, il n'est absolument pas total, (iv) de l'environnement complexe dans lequel se trouve l'individu et avec lequel il est en interaction, i.e., conceptions *épigénétiques*, et (v) de la nature bidirectionnelle du développement, en ce sens que l'individu et son propre changement participent de la niche développementale (i.e., « l'enfant est le père de l'adulte ») ; l'ensemble impliquant une vision non-réductionniste de la biologie (e.g., Gottlieb, 1987, 2007), une compréhension dynamique des changements culturels et environnementaux et une approche située en contexte (Lerner, 1998). Le chapitre 2 proposera une présentation plus détaillée de cadres théoriques du développement susceptibles de s'appliquer à cette recherche comparative.

Par-delà la nature observable et développementale du comportement, les quatre questions de Tinbergen n'offrent pas la possibilité directe de s'intéresser aux motifs internes sous-jacents de la conduite, ni à l'expérience subjective qui peut en être ou en faire l'objet[10]. En résumé, il s'agit d'un modèle d'étude du comportement et non de la cognition. Sans m'aventurer dans l'histoire des sciences cognitives, il semble nécessaire de préciser que les modèles théoriques qui permettent de relier le comportement à la cognition, ou à l'état mental, ou interne du sujet sont eux aussi variables selon les théories et les comportements considérés. L'exemple des conceptions de la pensée et du langage chez l'enfant par Jean Piaget et Lev Vygotski en donne une illustration classique. Pour le premier le langage est simplement le reflet de la pensée, c'est-à-dire le *reflet* des processus internes, de la construction interne de l'enfant (Piaget, 1936). Tandis que pour Vygotski, le langage est *l'outil* de la pensée, c'est-à-dire qu'il est internalisé progressivement par l'enfant et ce faisant transforme sa pensée (Vygotsky, 1978). Chez Piaget, l'usage du langage est sous-tendu par le processus interne à l'enfant du développement de son

[9] Noter qu'il ne s'agit pas pour autant d'une vision béhavioriste reposant uniquement sur des relations de stimulus-réponses.

[10] C'est précisément *le* défi des sciences cognitives d'identifier si l'expérience subjective résulte du comportement ou si le comportement résulte de l'expérience subjective. A cet égard, la philosophie et les neurosciences convergent vers l'idée que l'expérience subjective résulte du comportement et provient de l'extérieur (James, 1884; Murdoch, 1962; Talmi & Frith, 2011). Selon William James, fuir devant un ours, c'*est* avoir peur. Selon Murdoch, il n'y a pas de concept mental qui soit véritablement *interne*.

intelligence, tandis que chez Vygotski, le même comportement est emprunté à l'extérieur, à l'environnement social et c'est en se l'appropriant que l'enfant développe sa pensée (voir aussi Murdoch, 1962). Le même genre de débat se retrouve en psychologie comparée où l'on considère trois principaux modèles de cognition (Leavens et al., 2017) : un modèle d'équivalence qui stipule que si le comportement *p* est observé, cela implique que la compétence cognitive *q* est présente et inversement (p ⇔ q) ; un premier modèle conditionnel stipulant que si l'organisme possède la capacité *q*, alors il doit émettre le comportement *p* (q => p) ; un deuxième modèle conditionnel stipulant que si l'organisme est capable du comportement *p*, alors c'est qu'il possède la capacité *q* (p => q). De mon point de vue, le modèle d'équivalence et le premier modèle conditionnel privent tout organisme de libre arbitre et de contrôle exécutif puisque si la capacité cognitive est présente alors le comportement *doit* être exprimé. Au contraire, le deuxième modèle conditionnel, retenu dans mes travaux de recherche, s'appuie sur l'observation du comportement *p* pour inférer la présence de la capacité cognitive *q*. Ce modèle implique que l'expression du comportement *p* indique la présence de la capacité *q* à condition que les observations de *p* soient simplement ordonnées si *q* est une grandeur (Bourjade & Vautier, 2017) ou que *p* soit observé de manière stable dans différents contextes si *q* est un état qualitatif discret (Whiten, 1996). Par la contraposée, l'absence de la capacité *q* implique l'absence du comportement *p*. Cela dit, il est important de rappeler qu'il n'est pas possible de raisonner sur l'absence de *p* pour inférer l'absence de *q*. Premièrement ce serait une affirmation du conséquent et donc un raisonnement fallacieux, et deuxièmement, si les organismes possèdent un libre arbitre, et même si celui-ci se limite à un contrôle exécutif en termes d'inhibition comportementale, alors, ils peuvent posséder la capacité *q* et ne pas produire le comportement *p*.

Pour terminer la présentation des enjeux interdisciplinaires que ce travail de recherche doit relever, je rappelle qu'il s'appuie sur la méthode comparative et plus précisément sur la comparaison d'espèces entre elles pour investiguer la phylogenèse des comportements et des compétences cognitives dont on cherche à caractériser l'évolution (ci-après dénommés *traits* ou *caractères*). Il s'agit d'examiner la quatrième question de Tinbergen en y intégrant une approche développementale et adaptative ; ou en d'autres termes, de mêler la question de l'ontogenèse à celle de la phylogénèse en s'intéressant aux causes ultimes potentielles, i.e., aux fonctions adaptatives du développement. Cette approche sera abordée dans le chapitre 2, de même que les contraintes qu'impose la comparaison. Par ailleurs, sur le plan strictement comparatif, et bien que l'ontogenèse ne récapitule pas la phylogénèse, les études ontogénétiques

offrent des opportunités uniques d'identifier les facteurs qui déterminent l'émergence d'un caractère particulier (Gottlieb, 1987), notamment lorsque celui-ci est appris socialement et lorsque les conditions de son acquisition, à savoir une culture ou une organisation sociale spécifique constituant la niche ontogénétique (Stamps, 2003; Vygotsky, 1978; West et al., 2003) sont, elles aussi, susceptibles d'être héritées (Jablonka & Lamb, 2014).

Dans la section suivante, je m'attache à proposer les définitions des concepts clefs de cette recherche que sont le langage, la communication et les signaux communicatifs. Ces choix s'appuient sur la mise en correspondance des champs lexicaux des différentes sous-disciplines impliquées dans ce programme de recherche et permettent de fournir une assise opérationnelle à la description quantitative de la communication intentionnelle, unimodale et multimodale, et des phénomènes attentionnels visuels, tactiles et auditifs, relevant de l'attention sociale, c'est-à-dire l'attention adressée à un congénère[11] (Bourjade, 2017).

1.4. Conceptions du langage, de la communication et des signaux communicatifs

1.4.1. Le langage

Il existe d'innombrables définitions du langage, associées à des conceptions variées qui tendent à souligner le caractère formel de son organisation ou la diversité de ses manifestations, y compris son aspect générativiste et créatif, et son nécessaire ancrage dans le corps, le contexte, la culture. Parmi les définitions les plus classiques, on trouve celles reposant sur un ensemble précis de caractéristiques nécessaires mais non suffisantes ; parmi celles-ci, les caractéristiques de Hockett (1960) sont les plus souvent citées, bien qu'elles ciblent les origines de la *parole* et non les origines du langage. L'application de chacun des critères à la communication animale a fait l'objet de nombreux débats (e.g., Wacewicz & Żywiczyński, 2015) mais une implication générale de cette approche est que la plupart des caractéristiques peuvent être observées chez d'autres espèces animales ; cependant, elles ne le sont jamais toutes ensemble, excepté chez l'humain. Parmi les caractéristiques que l'on peut retrouver dans la communication animale, citons de manière non-exhaustive : le canal oral-vocal et les spécificités acoustiques qui y sont

[11] Individu de même espèce, et par extension dans le contexte interspécifique des travaux présentés dans le chapitre 3, un partenaire de communication.

associées[12], la sémanticité, la discrétion (discontinuité), la double articulation, la prévarication[13] ou encore la transmission culturelle et par apprentissage – bien que chacune fasse l'objet de controverses. En revanche, la réflexivité du langage (utiliser le langage pour parler du langage), la récursivité ou le « déplacement » (*displacement*) qui consiste à utiliser le langage pour échanger des contenus abstraits ou qui n'émanent pas du présent et de la perception directe sont des compétences qui, à quelques exceptions près[14] et à l'heure d'aujourd'hui, sont absentes chez les animaux. Par ailleurs, la nature *arbitraire* qui relie les formes à leurs référents, comme le mot 'lune' à l'objet 'lune', est sujette à débat dans la communication animale puisque les signaux de communication ont évolué sous des pressions de sélection exercées par les receveurs, ce qu'une partie des biologistes considèrent comme une relation « non arbitraire » (Maynard-Smith & Harper, 1995). Enfin, il est nécessaire de mentionner que les critères de Hockett (1960) font également l'objet de critiques en linguistique et en sémiotique. D'une part, l'universalité de ses critères dans toutes les langues naturelles a été remise en question (voir par exemple Everett, 2005) et d'autre part, l'usage d'un critère séparé des autres, par exemple la nature arbitraire d'un signal sans sémanticité compositionnelle (au sens strict) n'est pas acceptée comme une homologie par de nombreux linguistes (e.g., Del Prete, communication personnelle, juillet 2021; Wacewicz & Żywiczyński, 2015). En résumé, pour le biologiste, il n'y a pas d'arbitraire car les signaux résultent des processus évolutifs qui les ont façonnés pour être efficaces sur les receveurs (i.e., Maynard-Smith & Harper, 1995) et pour le linguiste, il n'y a pas d'arbitraire s'il n'y a pas de sémanticité au sens strict puisque la relation forme-référent n'est pas pleine ou totale (i.e., Wacewicz & Żywiczyński, 2015).

A cette approche *bottom-up*, il est possible d'associer une définition théorique (*top-down*) telle que celle proposée par Hauser, Fitch & Chomsky (2002), s'inscrivant dans une perspective évolutionnaire. Ces auteurs envisagent le langage comme une exaptation, c'est-à-dire un trait qui a évolué secondairement, sur la base d'adaptations primaires ayant été sélectionnées pour

[12] Spécificités initialement centrées sur l'oralisation du langage qui est longtemps resté le modèle dominant mais qui inclut maintenant les langues signées, dans une perspective multimodale (Vigliocco et al., 2014; Wacewicz & Żywiczyński, 2015; Wilcox & Xavier, 2013).

[13] Capacité à mentir ou à tromper.

[14] La danse des abeilles est citée comme un exemple de « *displacement* » (Hockett, 1960) qui fait débat (Wacewicz & Żywiczyński, 2015).

les avantages qu'elles fournissaient en remplissant d'autres fonctions[15]. Plus précisément, ils considèrent le langage comme une capacité *incarnée*, reposant sur plusieurs systèmes biologiques (e.g., sensori-moteur) et cognitifs (e.g., conceptuel-intentionnel), dont la majorité est présente chez les autres espèces animales et constitue la faculté de langage au sens large (*Faculty of Language in the Broad sense – FLB*), tandis qu'une minorité de compétences restreintes à la dimension récursive du langage serait, elle, une propriété de la faculté de langage au sens strict (*Faculty of Language in the Narrow sense – FLN*); à ce jour observé uniquement dans l'espèce humaine.

Sur ces bases, je situerai mon propos dans une conception biologique du langage, au travers de laquelle le langage est d'abord un comportement *situé*, réalisé dans un contexte social et mobilisant de concert des adaptations biologiques, cognitives et sociales dont certaines sont héritées quand d'autres sont développées (Bottineau, 2018; Fitch, 2010; Locke, 2009). La nature essentiellement incarnée des manifestations langagières s'accompagne d'une vision énactive/émergente du comportement comme résultant d'un ensemble de conditions internes et externes à l'individu, nécessaires à son expression. Cette vision s'accompagne d'une prise de distance avec la proposition de Hauser et al. (2002) promulguant une acception cognitiviste du langage sous forme de 'systèmes' internes, et préférant une acception comportementale, incarnée et située en contexte – et donc éminemment sociale. La conséquence directe de cette perspective est que mon propos portera sur des *individus capables de communication ou de langage* et non sur des systèmes de communication ou des langues à la manière de la linguistique formelle (Del Prete, communication personnelle, juillet 2021). De plus, les caractéristiques des communications étudiées seront décrites et répertoriées à la manière des caractéristiques de Hockett (1960), sans nécessité d'y assigner la preuve d'une appartenance au langage, dans une opposition qui serait manichéenne, à tous les autres systèmes de communication.

En résumé, mon approche est centrée sur les caractéristiques observables des individus *capables de communication ou de langage* et traite la question de savoir si chaque caractéristique est homologue à celle observée dans le langage humain avec circonspection. L'approche que je défends est celle de se doter d'une définition théorique du langage, associée

[15] L'exemple de l'oralité du langage est souvent utilisé pour illustrer cette hypothèse : le tractus vocal aurait évolué pour des fonctions végétatives et aurait, par la suite, permis de produire la parole, à son tour soumise à des pressions de sélection pour cette nouvelle fonction (secondaire).

à des critères observables et qui soit *indépendante* de la spécialisation humaine pour le langage, sans quoi, nous nous piégeons dans un raisonnement circulaire (e.g., le langage est défini comme étant humain, et nous l'étudions chez d'autres espèces pour finalement trouver qu'elles ne le possèdent pas). Les critères sur lesquels portent essentiellement ce travail sont l'intentionnalité, la volition (la capacité de produire un comportement de manière volontaire), la sensibilité à l'attention du receveur, la double articulation avec une centration sur la combinaison d'unités non signifiantes en un ensemble signifiant (Cf. chapitres 3 & 4).

1.4.2. La communication

A l'instar du langage, il existe de nombreux paradigmes permettant de formaliser la communication, comme une activité sociale, un système, une compétence etc. Le modèle le plus connu est sans conteste celui de la *Théorie de l'information*, proposée par Shannon et Weaver (1949). Ce formalisme permet de relier un *émetteur* à un *récepteur* par l'intermédiaire d'un *canal* permettant l'échange de signaux de communication encodés dans un matériel physique (e.g., un mouvement perceptible, une onde sonore), de nature plus ou moins complexe, mais possédant comme caractéristique la *transmission* d'un *message* constitué d'*information*. Ce paradigme (*code model*) a façonné l'étude de la communication animale et a largement été adopté par les sciences humaines avant de recevoir diverses critiques. En effet, il repose sur une vision simplificatrice des notions de signal communicatif et d'information, les deux locuteurs étant supposés utiliser un répertoire commun et fini, et offre ainsi une vision réductionniste des dimensions sociale et pragmatique de la communication (Font & Carazo, 2010; Scott-Phillips, 2015b). Il repose également sur une notion mathématique de l'information, définie en termes probabilistes, et non logiques, comme étant la réduction de l'incertitude face à *n* alternatives toutes équiprobables. L'information est donc ce qui réduit l'incertitude chez le receveur.

Plusieurs auteurs ont souligné l'usage superflu, voire trompeur de la « métaphore linguistique » pour étudier la communication animale (Font & Carazo, 2010; Rendall et al., 2009). Leur critique porte essentiellement sur la notion de *message* et de *signification* (*meaning*) qui seraient véhiculés dans les échanges communicatifs chez les espèces animales. Selon ces auteurs, il n'y a pas de message équivalent à celui transmis par le langage humain tant qu'il n'y a pas de conscience de la part de l'émetteur que le message véhicule de l'information (Rendall et al., 2009). Ce faisant, les auteurs pointent la nature *ostensive* de la communication humaine,

caractéristique initialement introduite par Grice (1957), et sur laquelle repose la dimension *pragmatique* du langage, visant à réduire le fossé entre la signification intentionnellement désirée par le locuteur et la signification extraite par l'interlocuteur (Scott-Phillips, 2015b; Sperber & Wilson, 2002; van Duijn, 2016). Cette vision 'gricéenne' de la communication repose sur la lecture des intentions mutuelles des interlocuteurs et la question de savoir si les animaux usent de leurs signaux à ce degré d'intentionnalité de deuxième ordre (Dennett, 1971) continue de faire débat dans la littérature (Graham et al., 2020; Moore, 2016; Scott-Phillips, 2015a). Communiquer de manière ostensive implique que (i) l'émetteur ait l'intention d'induire une réponse chez le receveur, (ii) qu'il considère son action comme étant la transmission d'un message ou d'un signal communicatif et (iii) qu'il considère que le receveur partage la même vision, c'est-à-dire que ce dernier 'comprend' que l'on s'adresse à lui pour lui transmettre un message ou tout au moins, il 'comprend' l'intention de communication (Graham et al., 2020; Moore, 2016). Ces acceptions ne sont pas sans poser problème dans la recherche sur la communication animale, dans la mesure où il n'est pas possible d'accéder aux représentations des animaux (Bourjade et al., 2020; Leavens et al., 2017; Penn & Povinelli, 2013). Si la recherche sur l'animal peut se satisfaire de la métaphore tant qu'elle reste indépendante des études sur le langage (Shettleworth, 2010), ce n'est plus le cas dès lors que les disciplines se rencontrent et qu'il s'agit de mettre en correspondance les phénomènes. Le Chapitre 2 s'attachera à proposer des définitions opérationnelles de la communication intentionnelle et ostensive qui soient applicables à tous les modèles d'étude mentionnés dans cette note de synthèse.

Finalement, il est possible d'envisager la communication dans une acception comportementale et fonctionnelle qui s'émancipe quelque peu des notions de message et d'information, en revenant au comportement communicatif qui repose sur une *interaction sociale*. Cette perspective sociale de la communication, qui sera adoptée dans le présent écrit, repose essentiellement sur la modification (observable) du comportement ou de l'état de connaissance sur le monde chez le receveur d'un signal communicatif ; en d'autres termes sur l'*influence* que la communication opère sur le receveur (Bourjade et al., 2020; Font & Carazo, 2010; Maynard-Smith & Harper, 1995; Rendall et al., 2009). Plus précisément, la notion d'information comme réduction de l'incertitude peut être remplacée par la notion *d'information fonctionnelle* qui constitue toute information à laquelle il est bénéfique de réagir pour le receveur (Font & Carazo, 2010), ou plus généralement tout contenu influent sur le comportement du receveur (Rendall et al., 2009).

Ici, je conserverai la notion d'*influence*, comprenant à la fois les signaux honnêtes et trompeurs (voir ci-dessous), produits au bénéfice de l'émetteur, et sur lesquels le receveur peut agir *différentiellement* ; en répondant ou ne répondant pas, et en interprétant le signal selon les bénéfices qu'il en retire. En effet, j'ai proposé avec mes co-auteurs (2020) que cette influence ne soit pas entièrement opérée par l'émetteur des signaux communicatifs ; au contraire, que la *communication* se distingue précisément de l'*interaction* par la marge d'interprétation des signaux qu'elle offre au receveur ; les motifs et intérêts de l'émetteur et du receveur pouvant diverger (Maynard-Smith & Harper, 1995).

1.4.3. Les signaux de communication

Définir la communication revient essentiellement à caractériser la nature de l'échange communicatif. Sur le plan comportemental, la communication est un construit abstrait tandis que les signaux de communication s'observent. Mes travaux de recherche reposent sur une conception des signaux de communication comme étant mécaniquement ineffectifs, c'est-à-dire que l'émission du signal ne suffit pas à remplir sa fonction communicative, qui reste à la charge du receveur (Bourjade et al., 2020; Scott & Pika, 2012). Autrement dit, c'est la réaction du receveur qui remplit la fonction communicative. Cette définition peut être illustrée avec une expression faciale de menace. La menace est émise par l'émetteur du signal, mais pour que la menace soit effective, le receveur doit en tenir compte et y adapter son comportement. A l'inverse, une agression physique est mécaniquement effective et ne laisse pas au receveur le soin d'interpréter et d'ajuster son comportement. Selon la définition utilisée dans ce travail, la menace est un signal communicatif tandis que l'agression physique ne l'est pas. Cette définition peut paraître réductionniste aux psychologues et aux linguistes qui travaillent sur des nuances interprétatives subtiles. Par exemple, nous avons proposé qu'une gifle puisse être un signal communicatif chez l'humain, mais pas chez le PNH, car même si les effets mécaniques du signal ne peuvent pas être ignorés par le receveur humain, il est toujours possible pour l'humain d'interpréter la signification de la gifle selon le contexte, l'histoire interactionnelle et d'autres sources symboliques et culturelles (Bourjade et al., 2020). Cet espace d'interprétation est précisément ce que j'appelle *communication*, et qui présente des obstacles épistémiques différents chez l'homme et l'animal – avec notamment l'impossibilité d'accéder à la subjectivité de l'animal. Pour répondre aux enjeux de l'interdisciplinarité, la définition retenue est celle applicable à toutes les espèces, sans pour autant s'interdire d'étudier des signaux supplémentaires chez l'humain.

Par ailleurs, une des grandes questions biologiques de la communication concerne l'évolution de son honnêteté, ou en d'autres termes, l'obtention d'un système évolutivement stable concernant les échanges communicatifs (Maynard-Smith & Harper, 1995; Scott-Phillips, 2008). Maynard-Smith & Harper (1995) définissent un *signal* communicatif comme une action ou une structure qui accroit la valeur adaptative de l'émetteur en altérant (positivement ou négativement) le comportement des organismes qui le détectent, et qui possède des caractéristiques (physiques) ayant évolué du fait de ces effets. Selon leur typologie, un signal peut indiquer des propriétés propres à l'émetteur, ou à l'environnement, mais il se distingue des *indices* qui sont des éléments externes que les individus peuvent utiliser pour orienter leur comportement, plus généralement prendre des décisions (Maynard-Smith & Harper, 1995). Par exemple, observer un individu manger une nourriture attractive peut être un indice que cette nourriture a été trouvée à proximité dans l'environnement. L'individu qui mange n'émet pas de signal mais son comportement peut être utilisé par d'autres comme un indice ; il est source d'information publique, disponible pour quiconque (Dall et al., 2005; Danchin et al., 2004; Wagner & Danchin, 2010). La question de la fiabilité des signaux repose sur différentes stratégies évolutives : (i) soit l'émetteur et le récepteur ont les mêmes intérêts et dans ce cas il n'y a aucune raison que le signal ne soit pas fiable, (ii) certains signaux sont 'handicapants' pour l'émetteur de sorte que le coût pour le produire assure qu'il soit honnête, (iii) les indices sont physiquement reliés à une qualité intéressante pour le receveur et, (iv) l'usage non fiable de la communication peut être régulé par des mécanismes comme la réputation dès lors que les interactions sont répétées entre les partenaires (réputation directe) (Maynard-Smith & Harper, 1995; Scott-Phillips, 2008).

Ici, je retiendrai une définition active (et non structurelle) du signal communicatif telle qu'il s'agit d'une *action transitoire* dirigée vers un ou plusieurs destinataires, mécaniquement ineffective et reposant sur la réponse du receveur pour satisfaire sa fonction communicative. Cette définition exclut les indices et exclut également les signalements inhérents au phénotype qui seraient indépendants du contexte (Wagner & Danchin, 2010). Ainsi, un nourrisson de 3 mois ne *signale pas* sa dépendance à ses parents par son manque de tonus musculaire, comme on pourrait l'attendre sous l'hypothèse du handicap en considérant le phénotype comme un signalement (Seabright, communication personnelle, octobre 2020). En revanche, il *signale* sa dépendance par ses pleurs et ses cris, selon la définition de signal communicatif transitoire utilisée dans mes travaux de recherche.

Un dernier mot doit être ajouté sur la genèse des signaux de communication. Comme introduit plus haut, la perspective évolutionnaire soutient l'idée que les signaux de communication ont évolué pour leurs effets sur les receveurs et que certaines propriétés physiques des signaux (e.g., les cris d'alarme) sont particulièrement efficaces pour remplir cette fonction. A l'inverse, le langage humain est entièrement appris et les enfants qui ne l'acquièrent pas durant la période infantile et juvénile ne l'acquièrent jamais complètement (Cf. chapitre 2). Pour cette raison, de nombreuses hypothèses ont été avancées concernant la genèse des gestes communicatifs des primates qui répondent à une ontogenèse longue, semblent appris par les individus, présentent une forte variabilité inter-individuelle et inter-groupe, et sont affectés par l'expérience individuelle (Call & Tomasello, 2007; Hobaiter & Byrne, 2011a; Pika & Fröhlich, 2019; Prieur et al., 2020). De mon point de vue, la canalisation phylogénétique des signaux n'empêche pas les individus d'apprendre socialement la manière appropriée de les produire (Hobaiter & Byrne, 2011a, 2011b; Seyfarth & Cheney, 1986; Snowdon & Hausberger, 1997; West et al., 2003) et cette recherche ne fait pas de distinction entre les signaux communicatifs sur la base de leur origine.

Chapitre 2. Origines évolutives et développementales du langage

2.1. Etat de l'art sur les origines du langage

2.1.1. Modèles théoriques

> *« … L'onomatopée du chacal, de l'écureuil et du sanglier sont utilisés dans le huchement pour conduire le bétail […]. L'expression yaw yaw renvoie à la figure du sanglier ; elle traduit une émission sonore produite exclusivement par le sanglier, en l'occurrence le grognement. En cela, cette onomatopée relève de l'ordre de l'équivalence phonique ; la relation qu'elle sous-tend entre le son et le sens est « naturelle » et figurative. »*
>
> *Romain Simenel (2018, pp.487-488)*

Nim, Washoe, Viki … nombreux sont les chimpanzés auxquels les chercheurs ont tenté d'apprendre à parler en les « immergeant » dans la vie humaine : la maison, le repas, le jeu, les câlins. Aucun n'a su prononcer un mot mais ils ont été capables d'apprendre la langue des signes américaine avec un succès bien supérieur. De ce constat a résulté l'abandon de l'approche holistique de l'évolution du langage (Fitch, 2010). Peut-être l'anglais américain manquait-il d'iconicité sonore, d'idéophones et de gestes vocaux que les chimpanzés auraient pu imiter plus facilement ? Peut-être cet anglais-là n'était-il pas enseigné en passant par le « parler bébé » ou manquait-il de lien avec l'environnement « naturel » du chimpanzé ? Peut-être qu'une conception processuelle et intégrée du langage (Wilcox & Xavier, 2013), en lieu et place d'une conception normative centrée sur la performance aurait-elle était plus féconde ? Le langage comme une conception universelle, unifiée et normative a trouvé ses limites avec les grands singes entrainés au langage. Aujourd'hui, le champ d'étude de l'évolution du langage adopte une vision multi-composante de ce dernier émanant des ensembles de traits nécessaires à sa production et sa compréhension (Hauser et al., 2002). En guise de rappel, certains de ces traits appartenant au FLB (*Faculty of Language in the Broad sense*) sont impliqués dans d'autres fonctions que le langage, tandis que les traits du FLN (*Faculty of Language in the Narrow sense*) caractérisent plus précisément les structures linguistiques *per se*. Les modèles théoriques d'évolution du langage portent donc sur l'évolution de certains de ces traits – généralement quatre ou cinq – pouvant appartenir à l'un des deux systèmes, FLB ou FLN, ou les deux. Fitch (2010) propose une admirable synthèse des travaux théoriques et empiriques réalisés dans le domaine de l'évolution du langage jusqu'aux années 2010. Il dépeint le champ théorique à l'aide de trois familles de modèles ; à savoir les modèles *lexicaux*, *gestuels* et

musicaux. Chaque famille de modèles tente d'expliquer des aspects distincts des origines et de l'évolution du langage et par conséquent et en très grande majorité, les modèles ne sont pas mutuellement exclusifs.

2.1.1.1. Les modèles lexicaux

Les modèles lexicaux sont ceux qui tentent d'expliquer l'évolution des caractéristiques structurelles et organisationnelles du langage. Ce sont des modèles d'évolution des traits du FLN, majoritairement portés par des linguistes et anthropo-linguistes. Ils s'appuient sur l'hypothèse d'un protolangage, comme étape intermédiaire entre un système de communication ancestral, présents chez le dernier ancêtre commun (DAC ci-après) à l'humain et aux chimpanzés et bonobos, qui serait de nature *lexicale*. Le modèle de Derek Bickerton (2007) est l'un des plus représentatifs. Il stipule un protolangage lexical sans syntaxe qui aurait préludé à l'apparition 'catastrophique' (comprendre soudaine) d'une syntaxe élaborée caractéristique du langage « moderne ». Pour argumenter cette apparition soudaine, le modèle s'appuie sur l'explosion des outils culturels retrouvés autour des fossiles humains après *Homo erectus*. Bien que son hypothèse concorde avec d'autres sur la temporalité de l'apparition d'un langage « moderne », il n'explique pas *comment* la syntaxe aurait soudainement évolué. Par ailleurs, il tend à omettre l'importance de ce qu'il considère comme des non-questions et qui représentent l'évolution des traits du FLB ; à savoir si le langage a évolué d'un système de communication ancestral présent chez le DAC, si le langage a évolué à partir des gestes ou des vocalisations, ou encore s'il a évolué dans une seule niche écologique ou plusieurs simultanément – des questions hautement importantes pour qui prend la question de l'évolution au sérieux. Toutefois, sa critique de l'évolution du langage à partir des vocalisations des PNH n'est pas sans intérêt pour le présent propos puisqu'elle pointe des questions cruciales concernant l'évolution du contrôle volontaire de la voix. Bickerton (2007) critique la vision continuiste de l'évolution en argumentant ceci au sujet des cris des primates non humains :

a. Les cris sont génétiquement déterminés ; les mots sont culturellement déterminés,
b. Les cris sont équivalents à des propositions ; les mots ne le sont pas,
c. Les cris ne comportent pas de prédicats[16] ; toute forme de langage en comporte,

[16] En linguistique, le prédicat est ce qu'on dit du sujet ; dans une phrase simple, il y a le sujet et le prédicat.

d. Les cris sont indiciels (fonction expressive) et non symboliques ; les mots sont de réels symboles qui peuvent référer à des entités non présentes ici et maintenant et même non existantes.

Cette critique programmatique, bien que fondée sur une analogie entre cri et mot qu'il serait intéressant de questionner, pointe tout à la fois quatre dimensions centrales du langage et le lien présumé absent avec les vocalisations des primates – ce que diverses études récentes tendent à infirmer comme nous le verrons dans la section 2.1.2..

Un des succès des modèles lexicaux est par ailleurs, d'avoir pris au sérieux la question de la communication *coopérative*. En effet, comme je l'ai introduit dans la section 1.4.3., les biologistes ont toujours porté un intérêt marqué à la résolution du problème coopératif. Il s'agit de comprendre comment l'évolution a pu sélectionner des comportements qui peuvent être manipulés par des « tricheurs ». Dans le cas spécifique de la communication, il s'agit de comprendre comment l'échange d'information 'vraie' ou 'fidèle' sur l'environnement ou le signaleur peut être une *stratégie évolutivement stable* sans être envahie par de faux signalements qui seraient au bénéfice du seul signaleur. Cette question centrale de la biologie de l'évolution revêt une couleur particulière dans le cas du langage dont une des caractéristiques du point de vue de la linguistique, est précisément la *prévarication*, c'est-à-dire son usage à des fins mensongères ou trompeuses, ou plus généralement fictives.

Différents modèles théoriques rendent compte de stratégies évolutives permettant d'expliquer l'évolution du langage comme système coopératif de communication comprenant la possibilité de tromper sans qu'elle soit généralisée. Le modèle *Grooming-to-Gossip* de Robin Dunbar (1998) propose de résoudre ce problème. Dunbar s'appuie sur la théorie de l'évolution et sur des données comparatives entre les babouins et différentes populations humaines pour créer un parallèle intéressant entre l'épouillage des singes et le commérage des humains. Le commérage est répandu dans de très nombreuses cultures et permet de construire une *réputation* ; élément régulateur des pratiques individuelles et permettant d'identifier rapidement les tricheurs dans un groupe. Le modèle propose que l'humain résolve le problème de la communication coopérative de cette manière ; les menteurs (excessifs) sont exclus des communautés par le mécanisme de la réputation. L'argument évolutionnaire chez les primates non humains tient au temps extraordinaire que les singes passent à s'épouiller, en tenant compte de qui épouille qui, en réciproquant leurs services et en utilisant l'épouillage comme « monnaie » d'échange pour d'autres services (Dufour et al., 2009; Molesti & Majolo, 2017; Schino, 2007; Schino et al., 2007; Schino & Aureli, 2008). L'épouillage est une activité sociale par excellence, impliquée

dans la création, le maintien et la négociation des liens sociaux. Mais le temps requis pour l'épouillage d'un individu empêche de multiplier les partenaires. De fait, chaque individu possède généralement deux à trois partenaires d'épouillage préférentiels. En s'appuyant sur l'hypothèse de « l'intelligence sociale » selon laquelle la taille du cerveau d'une espèce reflète la complexité sociale et la taille des groupes, Dunbar propose que les premiers hominines[17] aient vécu dans des groupes sociaux de plus en plus larges ; la taille du groupe exerçant une pression de sélection sur l'activité d'épouillage. En effet, la taille des groupes augmentant, l'épouillage *manuel* (chronophage) n'était plus suffisamment efficace pour établir des liens avec les multiples partenaires environnants. L'épouillage serait devenu *vocal* sous la pression de l'augmentation de la taille du groupe, conduisant les individus à « épouiller » vocalement de nombreux partenaires à la fois. Dunbar considère que l'échange d'information spécifique et propositionnel ait été un prérequis pour que l'épouillage *vocal* deviennent un *commérage*, permettant de se défendre contre les *tricheurs.* La force de ce modèle est de répondre à la fois au problème de la communication coopérative et au problème de la dominance de la modalité vocale dans le langage moderne. Toutefois, les forces sélectives responsables de la sémantique propositionnelle ne sont pas spécifiées, de même que la question du contrôle volontaire de la voix, ou encore du fait que l'épouillage délivre des endorphines, tout comme la musique tandis que la parole ne le fait pas (Fitch, 2010).

2.1.1.2. Les modèles gestuels

Les modèles gestuels présupposent généralement l'existence d'un protolangage gestuel ayant précédé le langage moderne. Les humains de toutes les cultures produisent des gestes co-verbaux accompagnant le discours, et ce, même dans des situations où ces gestes ne peuvent pas être perçus par l'interlocuteur (comme au téléphone par exemple). Les modèles gestuels considèrent ces gestes comme des « fossiles » d'un protolangage gestuel ancestral et explorent si les espèces phylogénétiquement proches en partagent les caractéristiques. L'hypothèse gestuelle du langage telle qu'étudiée aujourd'hui remonte à la proposition de Gordon Hewes (1973) qui reposait sur trois arguments principaux : (i) un argument comparatif s'appuyant sur les compétences des primates non humains en matière de communication gestuelle, (ii) un argument sur la sémanticité des gestes (Fay et al., 2013; e.g., iconicity and pantomime: Fay,

[17] Groupe d'espèces comprenant les australopithèques et les représentants du genre *Homo*.

Ellison, et al., 2014a; Fay, Lister, et al., 2014b), et (iii) un argument sur le contrôle neural cortical des gestes communicatifs. A ces trois arguments originaux, s'ajoutent la découverte du rôle des gestes intentionnels dans le développement du langage chez l'enfant (e.g., Iverson & Goldin-Meadow, 2005; Iverson & Thelen, 1999), la formation spontanée de *home-sign languages* par les enfants sourds (e.g., Goldin-Meadow, 2012; Rissman et al., 2020), et la découverte des neurones miroirs et de leur rôle putatif dans l'émergence du langage (Gallese et al., 1996; Keysers et al., 2003; Rizzolatti & Craighero, 2004). Ces arguments sont discutés en détails par Fitch (2010) et je ne reprendrai ici que les éléments utilisés plus explicitement dans mes travaux de recherche.

L'argument *comparatif* repose sur l'usage flexible et intentionnel des gestes communicatifs par les grands singes (Tomasello, 2008; Tomasello & Call, 2019) et plus généralement sur toutes les caractéristiques socio-cognitives qui sous-tendent leur communication gestuelle et se trouvent être des prérequis importants du langage (Liebal & Call, 2012; Liebal & Oña, 2018; Tomasello, 2006, 2008). Parmi ces compétences cruciales pour le langage, on retrouve la flexibilité de l'usage des gestes qui peuvent être produits ou inhibés volontairement, la dissociation entre la forme et la fonction du geste, mise en évidence par le fait qu'un geste peut être produit dans différents contextes fonctionnels et qu'une fonction communicative peut être satisfaite avec différents gestes (e.g., Hobaiter & Byrne, 2014). Cette disposition à un couplage multiple des formes et des fonctions du geste en fait un précurseur idéal de la pragmatique et de la sémantique du langage moderne (Tomasello, 2008). De plus, les grands singes adressent leurs gestes communicatifs à des partenaires précis, en les regardant et en tenant compte de leur état attentionnel, et en répétant ou modifiant leur geste initial si le but communicatif n'est pas atteint ; trois critères majeurs de la *sémanticité intentionnelle* (Bates et al., 1975; Leavens, 2004; Townsend et al., 2017). Ce champ comparatif s'applique également aux populations humaines étudiées pour leurs capacités différentielles à s'exprimer gestuellement ou vocalement (en dehors de l'usage de la langue maternelle) lors d'expériences en laboratoire, interculturelles, ou de sémiotique expérimentale (e.g., Fay et al., 2018). Les humains ont une tendance naturelle à utiliser les gestes plus facilement que la voix, et ce, dès l'enfance comme le reprend l'argument ontogénétique des origines gestuelles du langage (Tomasello, 2008).

L'argument *ontogénétique* repose sur l'observation des comportements communicatifs de l'enfant préverbal durant les trois premières années de vie. Les premiers signes de communication intentionnelle apparaissent lorsque l'enfant commence à trianguler et coordonner son attention visuelle avec celle de l'adulte vers une entité extérieure – processus

dénommé *attention conjointe* (Bourjade, 2017; Carpenter et al., 1998; Scaife & Bruner, 1975). Ces séquences de triangulation s'accompagnent progressivement de gestes de pointage produits par l'enfant en direction de l'objet et à l'attention de l'adulte par le biais d'alternance de regards entre les deux cibles (Beuker et al., 2013; Camaioni et al., 2004; Carpenter et al., 1998; Liszkowski et al., 2012). Cette triangulation visuelle et gestuelle marque l'entrée dans la communication intentionnelle et référentielle desquelles découle l'entrée dans le langage (Cochet & Byrne, 2016; Colonnesi et al., 2010; Kita, 2003; Liszkowski et al., 2012). En effet, les gestes de pointage sont peu à peu couplés à des vocalisations, des babillages et des proto-mots (Aureli et al., 2017; Igualada et al., 2015; McCune & Zlatev, 2015) jusqu'à former les premières phrases bimodales (Guidetti et al., 2014; Iverson & Goldin-Meadow, 2005). Ainsi, cette phase de communication gestuelle intentionnelle ouvre à l'enfant une voie royale à l'entrée dans le langage oral-verbal et constitue donc un précurseur ontogénétique du langage (Colonnesi et al., 2010; Iverson & Goldin-Meadow, 2005; Tomasello, 2008).

L'argument *neural* repose sur l'hypothèse d'un système gestuel latéralisé dans l'hémisphère gauche déjà présent chez les grands singes et qui constituerait un précurseur neuronal à la dominance hémisphérique gauche du langage (Corballis, 2003; Hopkins & Vauclair, 2012). Le point central repose sur la spécialisation hémisphérique d'une fonction à l'échelle de la population entière ; dans le cas de l'humain, 90% de la population présente une préférence pour la main droite dans les activités manuelles (Knecht et al., 2000), impliquant un contrôle de l'hémisphère gauche. Cette manifestation transcende les cultures (e.g., Raymond & Pontier, 2004) et l'évolution humaine (e.g., Faurie & Raymond, 2004; Uomini, 2009). L'autre spécialisation hémisphérique gauche présente chez la majorité des humains est celle pour le langage. Cette double spécialisation hémisphérique a probablement évolué de manière contingente, la préférence manuelle ayant précédé la spécialisation hémisphérique pour le langage (Corballis et al., 2012; Forrester et al., 2013).

Deux catégories de modèles théoriques se disputent la cause de cette continuité. La première établit un lien *moteur* entre la préférence manuelle pour les activités manuelles complexes (e.g., l'utilisation d'outils) et la motricité nécessaire au langage (Corballis et al., 2012; Forrester et al., 2013; Greenfield, 1991). Il est possible d'associer à cette catégorie, les modèles sur les neurones miroirs, qui proposent que l'aire de Broca prenne en charge la détection et la représentation d'actions manuelles et hiérarchiques complexes indépendamment de la modalité (Arbib, 2005). Les neurones miroirs seraient impliqués dans l'imitation des actions simples,

puis complexes, puis communicatives, favorisant la parité entre les rôles de locuteurs et d'interlocuteurs.

L'autre catégorie de modèles théoriques établit un lien *communicatif* entre la préférence manuelle pour les gestes communicatifs, qui revêtent à la fois une dimension motrice et communicative, voire linguistique, et la fonction communicative du langage (Hopkins & Vauclair, 2012; Meguerditchian & Vauclair, 2014). Le contrôle hémisphérique gauche pour les gestes communicatifs représenterait ainsi un précurseur de la dominance hémisphérique pour le langage moderne. Cette hypothèse est bien soutenue par les résultats expérimentaux chez les enfants humains qui acquièrent le langage et montrent une préférence pour la main droite dans les gestes de pointage plus que dans les activités non communicatives (e.g., Cochet & Vauclair, 2010a). Les travaux comparatifs menés chez les chimpanzés et les babouins ont permis de mettre en évidence un biais populationnel droitier plus important pour les gestes communicatifs que pour les gestes de manipulation (Meguerditchian & Vauclair, 2009, 2014; Meunier et al., 2012), et plus robuste aux contingences locales (Bourjade et al., 2013; Chapelain et al., 2012). Cet argument neural bénéficie également du soutien des études menées sur les signeurs ; une dominance hémisphérique gauche est également trouvée et le même réseau neuronal est activé par les langues de signes et la parole (e.g., Courtin et al., 2011; Emmorey et al., 2007), renforçant l'hypothèse que la spécialisation hémisphérique pour le langage n'est pas spécifique d'une modalité, et que geste et parole sont les deux faces de la même pièce.

2.1.1.3. Les modèles musicaux

Les modèles musicaux (i.e., vocaux) sont certainement les plus intuitifs sur le plan évolutionnaire ; ils tentent d'expliquer l'évolution du langage oral à partir des vocalisations des espèces parentes. Cela dit, ils font face à des difficultés importantes du point de vue comparatif puisque les vocalisations des primates non humains (i) ne sont pas sous contrôle cortical (Jürgens, 2002), (ii) constituent un répertoire de signaux phylogénétiquement hérités de petite taille et sans modifications ultérieures, et (iii) ne relèvent pas d'un apprentissage par imitation comme c'est le cas du langage humain ou des oiseaux chanteurs (Cheney & Seyfarth, 2018; Fitch, 2010). Toutefois, ces modèles animaux non primates apportent des éléments cruciaux sur les conditions environnementales pouvant favoriser le développement de la modalité vocale. De plus, les études récentes sur les vocalisations des PNH ont permis d'identifier des caractéristiques partagées avec le langage humain, notamment sur les plans sémantique et

pragmatique (Schlenker et al., 2017). Enfin, les modèles musicaux recoupent largement les modèles de communication envers les apparentés (*Kin communication models*) permettant d'offrir alternativement (i) une solution au problème de la communication coopérative, (ii) des forces sélectives à la prédominance de la modalité vocale chez l'humain, ou encore (iii) un contexte adaptatif à l'évolution de la sémanticité.

La majorité des modèles musicaux tendent à expliquer l'évolution de la phonologie, soit pour son organisation hiérarchique et ses contraintes de production, i.e., phonotaxe, soit pour sa capacité de générativité (création d'ensembles signifiants), i.e., la double articulation du langage (Martinet, 1984). Les scénarios sont généralement basés sur l'acquisition d'une phonologie primitive (*bare phonology*) sans sémanticité. De cette apparition découle l'apparition de l'imitation vocale pour les aspects syntaxiques simples de cette phonologie (e.g., structure de la phrase, segment). L'étape suivante est l'apparition d'une signification holistique des chants, avec une correspondance globale entre une phrase ou un chant constitué d'éléments phonologiques complexes, et un ensemble sémantique complexe. Cette connexion entre des signaux arbitraires et des entités concrètes génère naturellement l'arbitraire saussurien du lien entre signifiant et signifié. A ce stade, les modèles tentent d'expliquer l'apparition de la sémantique propositionnelle ou analytique. Le modèle de Kirby (1999) basé sur des simulations informatiques montre par exemple qu'il n'y a pas besoin de changements évolutifs supplémentaires pour que les agents d'une population virtuelle « découpent » accidentellement les significations complexes en morceaux de taille inférieure ; et ce découpage se fait naturellement le long de régularités phonologiques (modèle d'émergence de la morphologie).

Les modèles de communication envers les apparentés apportent des informations complémentaires, notamment sur les pressions de sélection permettant de passer d'une étape à l'autre. Le modèle de Dean Falk (2004) reprend l'idée initiale d'Ellen Dissanayake (1992) sur l'évolution de la musique : les capacités musicales (e.g., prosodie, chant) ont une fonction centrale dans la vie quotidienne, notamment entre les mères et les enfants. Falk (2004) y ajoute les éléments d'anthropologie évolutive permettant de dessiner le contexte adaptatif dans lequel les berceuses, les chants, ou encore le langage adressé aux jeunes enfants auraient pu évoluer. Sa proposition s'appuie sur les changements posturaux progressifs auxquels les hominines ont fait face après la généralisation de la bipédie (soit à partir des australopithèques, il y a 4 millions d'années (MA)). Le bassin des mères bipèdes est devenu plus étroit, menant à des naissances plus précoces – phénomène connu sous le nom de *conflit obstétrical* (Rosenberg & Trevathan, 1995). De plus, la modification anatomique du pied a progressivement supprimé la capacité de

s'accrocher à la mère comme le font les bébés singes sur les mères quadrupèdes. L'effet combiné de l'immaturité des nourrissons et de leur incapacité locomotrice et anatomique à s'agripper à la mère a (i) conduit les mères à poser le bébé au sol pour vaquer à leurs occupations – l'hypothèse porte le nom de *putting-the-baby-down*, (ii) modifié le rapport du bébé à la proximité maternelle : à l'initiative du bébé chez les quadrupèdes mais à l'initiative de la mère chez les bipèdes, (iii) accru la nécessité de porter attention à la progéniture plus dépendante, (iv) accru le besoin du nourrisson d'attirer l'attention sur lui et ses besoins, (v) favorisé une forme vocale d'échanges attentionnels : du côté du parent le chant ou le *motherese* pour fournir de l'attention et sécuriser le bébé ; du côté du bébé, les cris, les pleurs, puis les babillages pour attirer l'attention du parent. D'autres modèles proches soulignent davantage le rôle donné à l'enfant dans l'émergence du babillage (Locke, 2006, 2017) ou la coévolution des échanges attentionnels par les moyens du chant entre les parents qui doivent fournir de plus en plus d'attention et les enfants qui développent des stratégies de plus en plus efficaces pour en demander (Mehr & Krasnow, 2017). Ces modèles seront repris dans les théories du conflit attentionnel au Chapitre 4.

2.1.1.4. Les modèles multimodaux

Si Falk (2004) s'est concentrée sur la communication adressée à l'enfant en soulignant les aspects vocaux et chantés, plusieurs passages de son écrit soulignent que les gestes, les expressions faciales et la coordination de l'attention visuelle de la mère avec son enfant sont parties prenantes du propos. Il est donc relativement aisé d'interpréter la théorie de Falk comme une hypothèse multimodale des origines du langage. Ainsi, la communication *vocale et visuelle* adressée *à distance* par la mère bipède à son enfant dépendant aurait constitué le substrat d'un protolangage *multimodal*. La principale force de cette proposition est de fournir un contexte adaptatif à l'usage de la modalité *vocale* chez les hominines, qui soit absent chez les autres primates quadrupèdes.

Une autre catégorie de modèles multimodaux s'intéresse à l'oralité des signaux de communication. La proposition initiale de Peter MacNeilage (1998), intitulée *Frame/Content Theory* est à l'origine une théorie de l'évolution de la *parole*, et non du langage. Toutefois, elle permet d'éclairer des aspects cruciaux concernant l'évolution de la modalité vocale et de son contrôle volontaire. En particulier, MacNeilage présuppose que la parole aurait évolué à partir des activités orales cycliques de nos ancêtres communs avec les primates non humains. La

cyclicité de la mastication et de l'ingestion de nourriture aurait formé une matrice (*frame*) dans laquelle peut s'insérer un contenu (*content*) vocalisé, reproduisant la double articulation du langage que l'on peut trouver entre une syllabe et ses phonèmes. Les activités orales cycliques auraient donc peu à peu joué un rôle dans la communication des primates ; rôle dont on retrouve la trace dans certaines expressions faciales rythmiques comme le *lip-smacking*, i.e., un mouvement rythmique d'ouverture et de fermeture des lèvres et de la langue, destiné à des partenaires affiliés. Cette hypothèse a été examinée et étendue par Asif Ghazanfar (2013) qui a notamment étudié le caractère audiovisuel (i.e., bimodal), de la perception et de la production de la parole chez l'humain. Il note que la perception du langage est bimodale dès la naissance et ce, dans toutes les cultures, pointant le caractère universel et fondamental de la nature *audiovisuelle* du langage. Il souligne par ailleurs que la majorité des vocalisations des primates ne fonctionnent pas sur la base de ce couplage caractéristique entre le mouvement des lèvres et le son produit – les deux modalités sont largement découplées chez les PNH. Il existe cependant des exceptions fort intéressantes sur le plan évolutionnaire. Le *lip-smacking* est par exemple produit dans le respect de ce couplage audiovisuel, et ce, à la même fréquence que la parole humaine, soit entre 3 et 8 Hz. Cette hypothèse du rythme a depuis été étendue à d'autres vocalisations (Bergman, 2013; Lameira et al., 2015).

Plus récemment, deux propositions théoriques émanant de la psycholinguistique et de la psychologie comparée ont proposé des scénarios évolutifs intégrant les gestes et les vocalisations. Levinson & Holler (2014) proposent un scénario originellement gestuel-visuel permettant d'expliquer l'anatomie de l'œil humain, i.e., une scléra blanche très visible (Kobayashi & Kohshima, 2001; mais voir Perea-García et al., 2019 pour un contre-argument) et l'inhabituelle tolérance au regard mutuel. Ils appuient leur scénario sur ce que Levinson (2006) a dénommé le moteur de l'interaction (*interaction engine*) encapsulant les « mœurs communicationnelles » de l'humain, à savoir le face à face, les regards mutuels, l'attention conjointe, la communication coopérative, la connaissance partagée (*common ground*), auxquels ils raccrochent la très particulière capacité de tour de parole qui impose une pression énorme sur la production et la perception du langage – les deux processus étant menés par le cerveau en double tâche[18]. Selon ce scénario, le contrôle volontaire des vocalisations apparaît chez les ancêtres immédiats d'*Homo heidlebergensis* présentant un canal vertébral thoracique élargi

[18] En effet, il faut entre 600 et 1500 ms pour planifier une réponse tandis que le délai moyen de tour de parole est de 200 ms.

permettant le contrôle fin de la respiration. En comparaison, le propos de Frölich et al. (2019a) insiste sur l'équivalence des modalités gestuelles et vocales des primates non-humains en termes de compétences partagées avec le langage. Ils proposent que la capacité de contrôle vocal soit présente chez les grands singes et le DAC, de même qu'un usage des gestes iconiques. Leur point central est de fournir une solution à la prévalence de la modalité vocale chez l'humain (le dénommé *problème de la transition*, repris ci-dessous). Alors que les grands singes utilisent des gestes pour la communication proximale et les vocalisations pour la communication distale, l'humain fait exactement l'inverse. Selon les auteurs, cet état de fait dénote une *inversion modale* qui aurait émergé chez les hominines avec la généralisation de la communication face à face. Les auteurs proposent que l'usage de la modalité vocale autorise plus de complémentarité et se prête mieux à la structure compositionnelle du langage. Ils ne discutent pas le rôle du canal vertébral thoracique dans le contrôle de l'articulation fine et leurs arguments en faveur du contrôle vocal chez les PNH sont minces en regard des évidences inverses. Certes, plusieurs études en captivité ont montré un apprentissage possible de vocalisations, e.g., la production de *coo-calls* dans des programmes de renforcement chez le macaque (Coudé et al., 2011; Hage & Nieder, 2013); l'apparition de vocalisations « apprises » chez les chimpanzés captifs (Hopkins et al., 2007). Ces résultats apportent sans conteste des informations cruciales dans la compréhension des bases neurales de la communication intentionnelle et du langage ; mais ils ne sont pas suffisants pour affirmer une équivalence de contrôle cortical avec le contrôle gestuel des PNH, ni avec celui des vocalisations humaines. L'enjeu d'accroitre notre compréhension concernant l'apparition du contrôle vocal au cours de l'évolution des primates réside dans l'identification de *contextes* favorisant la combinaison des signaux visuels et auditifs, a fortiori vocaux, ayant pu faire l'objet de pressions sélectives. Bien que les modèles évolutionnaires parlent plus volontiers de *transition*, les modèles développementaux préfèrent la notion de *réorganisation* des modalités de communication au cours de l'acquisition du langage, rendant mieux compte de la conception multimodale actuelle du langage (Guidetti et al., 2014).

Un dernier mot doit être porté au qualificatif « multimodal » de ces modèles. Certains modèles présupposent un *résultat* multimodal, qui peut être atteint par des évolutions unimodales distinctes (e.g., Levinson & Holler, 2014), tandis que d'autres présupposent que le *processus* évolutif a impliqué les deux modalités depuis le départ (e.g., Fröhlich, Sievers, et al., 2019b; MacNeilage, 1998), quand finalement, certains modèles présupposent que les *origines* du

langage sont multimodales (McNeill, 2012; Wacewicz & Zywiczynski, 2017; Zywiczynski et al., 2017).

2.1.1.5. Les apports de la paléontologie

L'approche comparative peut également concerner des espèces éteintes sur la base des restes fossiles. Cette approche paléontologique amène des éléments rares mais cruciaux permettant d'éliminer les scénarios évolutifs non plausibles au regard des découvertes paléontologiques. Deux sources d'évidence sont pertinentes en regard de la question du contrôle vocal : la découverte du gène FOXP2 impliqué dans l'articulation orale, et la largeur du canal vertébral de la cage thoracique permettant un contrôle fin de la respiration (et donc de l'articulation).

Le gène FOXP2 a été identifié suite à une série de cas cliniques de dysphasies développementales au sein d'une même famille. Les fonctions touchées par les patients atteints relevaient de problèmes articulatoires, d'erreur de généralisation de règles de grammaire, de difficulté d'imitation vocale (Vargha-Khadem et al., 1995). Le gène identifié et séquencé, FOXP2, présentait des mutations chez les patients atteints de dysphasie comparé à la population générale (Lai et al., 2001). Par la suite, ce gène a été identifié et séquencé chez de nombreux mammifères, montrant un très fort conservatisme au cours de l'évolution : seuls deux acides aminés diffèrent entre la protéine sécrétée par l'humain et celle du chimpanzé (Enard et al., 2002). Depuis, 10 gènes impliqués dans des troubles du langage ou des dysphasies développementales ont été identifiés et séquencés (Mozzi et al., 2016). Ces gènes ne présentent pas de différences chez les espèces de mammifères et d'oiseaux qui apprennent par imitation vocale et les autres. Le séquençage de FOXP2 chez l'homme de Neandertal indique une variante similaire à celle d'*Homo sapiens,* suggérant que Neandertal avait des capacités articulatoires similaires aux nôtres.

Dans le même registre, la découverte d'un élargissement du canal vertébral au niveau des vertèbres thoraciques chez l'homme moderne est interprétée comme un élément anatomique saillant dans le contrôle des mouvements fins liés à la respiration et l'articulation. L'étude des restes fossiles des premières espèces d'*Homo* indique que cet élargissement n'était pas présent chez *Homo erectus*, sur un fossile âgé de 1.4 Ma alors qu'il l'est chez *Homo heidelbergensis* (0.5 Ma) qui est le DAC de Neandertal et Sapiens (Dediu & Levinson, 2013). Cet argument anatomique est donc utile pour éliminer des scénarios proposant un contrôle articulatoire fin trop précoce au cours de l'évolution. Toutefois, cela ne dit rien des capacités de contrôle vocal

en général ; il est en effet tout à fait possible de produire des vocalisations complexes sur l'expiration et rien n'interdit à un hypothétique protolangage d'être produit au rythme de la respiration.

2.1.1.6. Synthèse

En résumé, les modèles lexicaux réussissent le passage du lexique à la syntaxe en résolvant le problème de la communication coopérative. Ils nécessitent toutefois d'être articulés à d'autres modèles permettant d'expliquer les traits cognitifs et comportementaux sous-tendant le contrôle *volontaire* et *intentionnel* de l'usage des signaux[19]. A cet égard, les modèles gestuels sont sans doute les plus convaincants. Ils ont pour caractéristique d'être particulièrement ancrés dans les résultats d'expérience et d'observation et bénéficient ainsi d'un important soutien empirique. Ils sont également les moins spéculatifs, mais aussi les plus incomplets dans le sens où ils ne permettent pas d'expliquer comment la modalité vocale a prévalu au cours de l'évolution humaine, ni quelles pressions de sélection ont été en jeu sur le contrôle vocal depuis le DAC (Fitch, 2010). Les modèles musicaux tendent à expliquer l'évolution de la phonologie et de la double articulation du langage, en proposant des mécanismes pour expliquer l'apparition d'une sémantique holistique, puis propositionnelle. Ils sont intéressants à recouper avec les modèles de communication vers les apparentés qui offrent les meilleures perspectives de résolution du problème de la coopération, tout en proposant des contextes adaptatifs expliquant le déploiement de la modalité vocale chez l'humain.

Dans cette note de synthèse, j'articule les arguments principalement comparatifs et ontogénétiques de l'hypothèse *gestuelle* des origines du langage, avec des modèles à la fois *musicaux* et d'apparentement (*kin communication models*), suivant ainsi la proposition de Fitch (2010) selon laquelle le problème de la communication coopérative se résout entre parents et enfants. Les principaux modèles musicaux auxquels je fais appel sont ceux de Falk (2004), Locke (2006), et Mehr &Krasnow (2017). Ce dernier n'est pas à proprement parler un modèle d'évolution du langage mais un modèle d'évolution de la musique (tel qu'affirmé par les auteurs). Toutefois, la « musique » pointée par l'article est en fait un chant adressé aux enfants,

[19] La distinction entre contrôle volontaire et intentionnel est la suivante : un signal est sous contrôle volontaire s'il est possible de le produire ou de l'inhiber sur commande ; il est intentionnel s'il est émis dans l'attente d'un effet sur le destinataire. Plus largement, la communication intentionnelle peut être définie comme l'usage de signaux de communication ayant pour visée la modification du comportement ou de l'état mental d'autrui. Ces distinctions seront discutées dans la section 2.3.3.1..

soit une *berceuse* – ce qui représente (i) le chant le mieux reconnu à travers les cultures (Mehr et al., 2019), (ii) une étape très probable d'un protolangage, et (iii) un comportement vocal (*frame*) qui peut contenir ou non des contenus sémantiques (content). Dans le Chapitre 4, je présenterai la manière dont ces modèles peuvent être utilisés dans la théorie plus générale du conflit parent-enfant de Trivers (1974) ; ce qui rassemble tous ces modèles sous une unité que j'ai dénommée *le conflit attentionnel*.

Enfin, il est notoire de constater que les modèles théoriques de l'évolution du langage proposent des scénarios complexes, souvent spéculatifs et parfois non opérationnalisés. Prendre au sérieux l'évolution du langage implique pourtant de développer des modèles théoriques qui soient testables d'une part, et dont l'opérationnalisation est compatible avec l'approche comparative d'autre part. Par exemple, les modèles computationnels d'évolution distinguent les caractères quantitatifs des caractères qualitatifs et leur assignent des probabilités d'apparition et de réversion différentes (Harmon, 2018). Or, simplement avec la question de l'intentionnalité, ce point n'est pas consensuel. A cet égard, la transdisciplinarité de l'objet d'étude appelle à créer des passerelles entre les éléments théoriques des modèles et leur opérationnalisation dans les études. Pour cela, des théories auxiliaires sont nécessaires sans quoi il est impossible d'ancrer l'étude dans le réel. Les théories auxiliaires que je mobilise pour étudier les phénomènes de multimodalité et de contrôle exécutif de la communication sont pour cette raison présentées dans la section suivante (2.2.). Ces enjeux constituent également un développement de mon programme de recherche, évoqué à la fin du Chapitre 4.

2.1.2. Sources d'évidences vocales, gestuelles, multimodales chez les PNH

> *"If baboons wore hats, they could simplify their social negotiations saving a lot of time and effort. These material symbols contain information about roles and expectations of behavior. If only baboons had "hats", they could begin to shift from complex society where almost everything impinges simultaneously and needs to be continually negotiated towards a complicated society where negotiations are simplified, more focused, and produce more stable outcomes that can travel over space and time."*
>
> *Shirley Strum (2012, p.11)*

Les différents modèles théoriques présentés ci-dessus sont supportés par des évidences empiriques dans le domaine vocal, gestuel ou multimodal. Plusieurs revues de questions publiées récemment permettent d'en avoir une vision relativement exhaustive (Bergman et al., 2019; Cheney & Seyfarth, 2018; Fröhlich, Sievers, et al., 2019b; Prieur et al., 2020; Rodrigues

et al., 2021). Pour cet écrit, il s'agira plutôt de présenter quelques-unes des études relevant de la communication vocale, gestuelle ou multimodale des primates afin de souligner les contributions de chacun des systèmes de communication existant chez les espèces actuelles aux origines et à l'évolution du langage moderne.

2.1.2.1. La communication vocale des primates non humains

Les études sur la communication vocale des primates recèlent d'informations importantes concernant (i) l'espace vocal et la modulation acoustique des vocalisations, (ii) la double articulation des unités vocales en ensembles signifiants plus larges, (iii) la référence et la sémanticité des vocalisations, y compris dans une approche pragmatique, (iv) la flexibilité de production des vocalisations.

L'étude du tractus vocal des babouins de Guinée montre par exemple que cette espèce est capable de prononcer cinq types de voyelles – et la même voyelle peut être présente dans différentes vocalisations, à la manière des phonèmes dans les langues (Boë et al., 2017). Chez les orangs-outans, Adriano Lameira (2018) rapporte l'existence d'un répertoire riche de sons comme des clics, *smacks* et *raspberries* (air soufflé entre les lèvres pincées) qui diffèrent d'une population à l'autre, formant des dialectes. La structure acoustique de ces sons en fait de bons candidats pour des proto-consonnes desquelles les consonnes dites non voisées comme [s] ou [k][20] auraient pu évoluer. Par ailleurs, l'étude systématique de la composition des vocalisations des primates non humains a permis de mettre en évidence l'existence de régularités syntaxiques et statistiques (Prieur et al., 2020; Schlenker, Chemla, Schel, Fuller, Gautier, Kuhn, Veselinović, Arnold, Cäsar, Keenan, et al., 2016). Par exemple, des unités vocales (signifiantes ou non) peuvent être combinées entre elles pour former des ensembles signifiants plus larges (Schlenker et al., 2017). Chez les mones de Campbell, le même suffixe *-oo* peut être ajouté à deux racines, *krak* ou *hok* et modifier leur signification dans le même sens (Ouattara et al., 2009; Schlenker et al., 2017), constituant un exemple de morphologie. Chez le colobe guéréza, la combinaison de *snort* immédiatement suivie de *roar* semble indiquer un cas d'alerte peu spécifié ; moins spécifié que chaque vocalisation prise séparément. Ainsi, il ne s'agirait pas d'une combinaison sémantique, ni morphologique, mais pourrait être un exemple de complexité

[20] A la différence de [z] ou [g] voisées

phonologique (Schlenker et al., 2017). A l'inverse, la combinaison par le cercopithèque hocheur de *pyow* et de *hack* qui ont chacun leur signification indépendante, en séquence *pyow-hack* qui permet d'induire un mouvement de groupe, représente un cas de combinaison d'unités sémantiques en un ensemble sémantique supérieur (Arnold & Zuberbühler, 2008). Toutefois, la signification de l'ensemble supérieur ne semble pas dériver de la signification de chacun des éléments, limitant les possibilités qu'il s'agisse d'un réel cas de compositionnalité. Schlenker et collaborateurs (2017) proposent que le *pyow-hack* fonctionne comme un idiome ; syntaxiquement combinatoire mais pas sémantiquement compositionnel. Une étude récente sur les séquences vocales des chimpanzés de Taï offre un portrait similaire : les chimpanzés combinent des unités sémantiques entre elles mais à ce jour, il n'est pas évident de savoir si cela représente un cas de compositionnalité *stricto sensu* (Girard-Buttoz et al., 2021). De nombreux exemples sont développés dans Bergman et al. (2019) sans que l'on sache exactement à quel formalisme linguistique les combinaisons en séquence réfèrent. L'analyse des séquences vocales des PNH a également permis de mettre en évidence des régularités statistiques d'ordre syntaxique (e.g., le *boom* du mone de Campbell toujours en début de séquence sous forme de paires, *boom boom* (Schlenker, Chemla, Schel, Fuller, Gautier, Kuhn, Veselinović, Arnold, Cäsar, Keenan, et al., 2016)) et morphologique. En effet, les vocalisations du macaque de Formose se conforment à une loi de brévité dénommée loi de Zipf, selon laquelle les unités vocales les plus couramment utilisées sont également les plus courtes (Semple et al., 2010). De même, les séquences vocales des geladas (Gustison et al., 2016) et des chimpanzés (Fedurek et al., 2017) répondent à la loi de Menzerath selon laquelle la taille totale de la séquence est négativement corrélée à la taille de ses constituants.

Les vocalisations des primates ont longtemps été étudiées pour leur caractère référentiel envers une entité extérieure (Cheney & Seyfarth, 2018). L'étude pionnière de Seyfarth et al. (1980) sur les vocalisations des singes vervets avait permis de mettre en évidence des vocalisations référant à la nature du prédateur : léopard, aigle ou serpent étaient indiqués par des vocalisations différentes. Depuis, d'autres études ont rapporté des résultats équivalents (e.g., Ouattara et al., 2009 chez le monde de Campbell) et la nature référentielle des vocalisations a fait l'objet de controverses et de redéfinitions (Fischer, 2017; Scarantino & Clay, 2015; Wheeler & Fischer, 2015) sans empêcher l'application d'une approche sémantique telle que définie par le formalisme linguistique (Schlenker et al., 2017). Au contraire, les approches sémantique et pragmatique de la communication vocale des primates non humains ont offert une vision

différente. Les singes vervets par exemple utilisent les *grunts*[21], *wrrs* et *chutters* dans les rencontres inter-groupes alors que ces mêmes vocalisations sont utilisées dans d'autres contextes (Cheney & Seyfarth, 2018). De même, les enfants chimpanzés produisent des *grunts* et des *whimpers* dans des contextes fonctionnels émotionnellement différents (Dezecache et al., 2021). Ces deux études remettent en question le postulat de singularité du développement vocal humain, selon lequel l'enfant humain serait le seul à explorer vocalement des contextes émotionnels variés avec les mêmes vocalises ; i.e., la *flexibilité fonctionnelle* (Oller et al., 2013).

La prise en compte du contexte dans lequel est produite chacune des vocalisations permet de faire une analyse de la pragmatique de la communication vocale des PNH. Dans sa conception formelle, la pragmatique consiste à étudier les inférences interprétatives qui peuvent être faites d'un énoncé donné en fonction des éléments contextuels environnants – un exercice coopératif à la charge du receveur visant à raisonner sur les motivations et les intentions du locuteur (Grice, 1957; Moore, 2017; Schlenker et al., 2017). Lorsque plusieurs inférences possibles se trouvent en compétition les unes avec les autres, la sélection d'une interprétation repose généralement sur le Principe d'Informativité[22], selon lequel le receveur priorise la sélection par le locuteur de la plus grande quantité d'information à propos du sujet. Appliquée à la sémanticité des vocalisations des mones de Campbell tel que proposé par Schlenker (Schlenker et al., 2017) cette approche pragmatique indique que si *hok* (alerte d'un danger non terrestre) est produit, et non *hok-oo* qui est plus informatif (*hok-oo* indique une alerte non terrestre (*hok*) de faible intensité (*-oo*)), l'interprétation de *hok* est 'enrichie' par la reconnaissance que *hok-oo* n'est pas produit : ainsi cela désigne une menace non terrestre et dont l'intensité n'est pas faible, par déduction la présence d'une menace importante dans le ciel (présence d'un rapace).

Bien que les vocalisations des PNH soient majoritairement innées et que le répertoire montre peu de modifications au cours de la vie des individus, les PNH ont une flexibilité considérable à propos de *l'usage* des vocalisations ; c'est-à-dire qu'ils apprennent le contexte dans lequel produire une vocalisation donnée (Bouchet et al., 2017; Lemasson, 2011; Snowdon &

[21] Le terme *grunt* pourrait se traduire par grognement mais le terme anglais désigne en réalité une vocalisation qui résulte de la fermeture de la glotte par l'activation des muscles du larynx suivie de l'émission soudaine d'une voyelle – avec la bouche fermée ou ouverte, mais sans intervention d'articulateur supra-glottal. Le terme anglais est donc conservé dans l'ensemble du document.

[22] Principe selon lequel lorsqu'un locuteur énonce une phrase S qui évoque (entre en compétition avec) une phrase S', et si S' est plus informative que S, alors S' est fausse (si S' était vraie, le locuteur l'aurait énoncée).

Hausberger, 1997), et les décisions de vocaliser et de répondre à une vocalisation sont elles aussi flexibles (Bergman et al., 2019; Cheney & Seyfarth, 2018; Lemasson, 2011). Un nombre croissant d'études montre par exemple que les échanges vocaux des PNH suivent des règles conversationnelles similaires à celles des humains en respectant des 'tours de parole' manifestés par des regards orientés, le respect d'un délai de réponse, l'absence de vocalisation simultanée (Bouchet et al., 2017; Lemasson, 2011; Pougnault et al., 2020) ; et que cette structure conversationnelle se développe au cours de la vie des individus (e.g., Bouchet et al., 2017), à la manière dont elle se développe chez le nourrisson humain (e.g., Gratier et al., 2015; Northrup & Iverson, 2020). Quelques études récentes montrent également que la communication vocale des PNH répond à certains critères de la communication intentionnelle comme le fait d'obtenir un effet désiré sur le receveur (le but communicatif), y compris en y associant des alternances de regards (e.g., chimpanzé: Schel et al., 2013) ou en modifiant la structure acoustique des cris dans une séquence répétée (e.g., macaque japonais: Koda, 2004) ou selon la composition de l'audience (e.g., chimpanzé: Slocombe & Zuberbühler, 2007). Ainsi, la flexibilité de la communication vocale est plus importante que ce qui était initialement envisagé et plusieurs espèces montrent notamment des vocalisations « gradées » dont la structure acoustique s'étend sur un continuum dépendent d'éléments contextuels et/ou ontogénétiques (Lemasson, 2011; Taylor et al., 2021). Chez le chimpanzé, une étude de repasse (*playback experiment*) a montré que les différents types de cris (*scream*) émis lors des conflits étaient discriminés sur la base de leurs caractéristiques acoustiques ; et que les chimpanzés pouvaient également discriminer un cri émis en réponse à une agression, d'un cri émis en réponse à de la frustration sociale (Slocombe et al., 2009). Une compétence similaire semble exister chez le babouin olive, chez lequel les cris des femelles adultes ne reçoivent pas la même réponse selon la sévérité de l'agression et/ou le contexte de frustration sociale (observation personnelle). Finalement, plusieurs études rapportent des cas de convergence sociale[23] dans la structure des vocalisations, menant à des dialectes vocaux qui suggèrent des influences sociales au sein des groupes (Fischer et al., 2020; Lemasson, 2011; Lemasson et al., 2011). Ces éléments forment les bases d'une communication *coopérative* fondée sur des *inférences interprétatives concurrentes* du

[23] La convergence vocale est un cas particulier de *convergence sociale* caractérisée par une variabilité intra-groupe inférieure à la variabilité inter-groupe pour un comportement donné. Les cas de convergence vocale peuvent générer des dialectes, tandis que la convergence sociale de comportements visuels tend à être interprété en termes de traditions ou protocultures.

point de vue des receveurs ; en d'autres termes, ils invitent à en étudier la pragmatique, ce qui sera repris à la fin du Chapitre 4 (section 4.2.2.).

Un autre exemple de modification du répertoire vocal provient des cas atypiques d'innovation vocale chez les chimpanzés (e.g., Hopkins et al., 2007) et orangs-outans (Lameira et al., 2015, 2016) captifs. Ces études indiquent que dans des environnements nouveaux, et au contact des humains, les singes anthropoïdes peuvent enrichir leur répertoire vocal, produire de nouvelles vocalisations (Hopkins et al., 2007; Lameira et al., 2016), babiller (Lameira et al., 2015) ou encore combiner des gestes avec des sons non voisés comme le *raspberry*[24] pour obtenir l'attention de l'humain (Hopkins et al., 2007; Taglialatela et al., 2011). Plusieurs études ont examiné le contrôle neural sous-jacent à la production de ces sons non voisés et de ces nouvelles vocalisations chez les chimpanzés et montrent que (i) l'asymétrie faciale est opposée à celle du répertoire de cris naturels et indique un contrôle de l'hémisphère gauche (Reynold Losin et al., 2008; Wallez et al., 2012), (ii) la combinaison du *raspberry* avec un geste de requête/pointage accroit la préférence manuelle pour la main droite (Hopkins & Cantero, 2003), (iii) cette même combinaison est associée à une augmentation de l'activation de l'homologue de l'aire de Broca dans l'hémisphère gauche (Taglialatela et al., 2008, 2011), (iv) les individus produisant ces sons atypiques ont une plus grande quantité de matière grise dans l'homologue de l'aire de Broca que ceux qui ne les produisent pas (Bianchi et al., 2016), et (v) ces sons peuvent être appris par renforcement positif (Russell et al., 2013). Ces études montrent qu'un contrôle volontaire sur les signaux oraux est possible et suggèrent que le contrôle peut porter sur les cordes vocales (Lameira et al., 2016) et sur les articulateurs (Taglialatela et al., 2008).

2.1.2.2. La communication gestuelle des primates non humains

La communication gestuelle des primates non humains a reçu un intérêt grandissant ces dernières décennies pour (i) sa flexibilité et son ontogénèse (Liebal et al., 2019; Liebal & Call, 2012), (ii) ses propriétés intentionnelles (Meguerditchian, Cochet, et al., 2011; Tomasello, 2008; Tomasello & Call, 2019), (iii) ses propriétés sémantiques et pragmatiques (Byrne et al., 2017; Hobaiter & Byrne, 2014), ou encore (iv) son contrôle hémisphérique (Arbib, 2005; Hopkins & Vauclair, 2012; Meguerditchian & Vauclair, 2014).

[24] Il s'agit d'une fricative bilabiale produite par l'expiration d'air vers des lèvres fermées, voire pincées.

Initialement, les gestes manuels, puis corporels et orofaciaux des singes anthropoïdes ont été étudiés pour leur flexibilité apparente, leur apprentissage individuel et leurs caractéristiques apparemment intentionnelles (Tomasello, 2008; Tomasello et al., 1985, 1994). Cette flexibilité concerne l'usage, mais aussi la production et l'apprentissage de gestes ; et peut être interprétée de différentes manières selon les hypothèses ontogénétiques favorisées (Byrne et al., 2017; Leavens et al., 2005; Liebal et al., 2019). Sur le plan factuel, les répertoires de gestes diffèrent entre les individus d'une même espèce, de nombreux gestes idiosyncratiques ont été décrits (principalement en captivité), des différences de répertoires entre groupes captifs ont été rapportés comme des cas de convergence sociale, et les individus au contact des humains développent des gestes qui leur sont adressés sans difficulté (Liebal et al., 2019; Liebal & Call, 2012; Tomasello, 2008). De plus, l'ontogenèse des gestes est longue et leur acquisition, leur fréquence d'expression et les contextes dans lesquels ils sont utilisés varient au cours de la vie des individus (Genty et al., 2009; Hobaiter & Byrne, 2011a; Liebal et al., 2019 for a review). Les gestes sont utilisés dans des contextes fonctionnels différents, et ne sont donc pas attachés à une seule et unique fonction. A cet égard, différents gestes peuvent être utilisés à la même fin communicative. Cette dissociation des moyens et des fins représente (i) une caractéristique importante de la communication intentionnelle (Call & Tomasello, 2007; Tomasello, 2008; Townsend et al., 2017) (ii) une relation arbitraire entre la forme du signe et la signification associée, qui varie selon l'intention de l'émetteur, telle que la pragmatique le présuppose (Moore, 2017), (iii) une flexibilité fonctionnelle, entendue comme une capacité à dissocier le signe de son contexte émotionnel, affectif et environnant, à la manière des vocalises humaines (Oller et al., 2013).

Les observations d'individus tantôt captifs, tantôt sauvages ont peu à peu conduit les principales équipes étudiant les gestes communicatifs des singes anthropoïdes à diverger sur l'origine des gestes et par conséquent, les manières d'interpréter cette flexibilité. A ce jour, plusieurs hypothèses concurrentes et mutuellement exclusives inondent la littérature (voir une synthèse dans: Liebal et al., 2019; Prieur et al., 2020). Les gestes peuvent ainsi être considérés comme (i) hérités phylogénétiquement (Byrne et al., 2017), (ii) ritualisés ontogénétiquement (Call & Tomasello, 2007; Tomasello, 2008), ou (iii) négociés socialement, sur la base d'interactions répétées (Pika & Fröhlich, 2019). Notons toutefois qu'initialement, Tomasello et ses collaborateurs avaient mentionné que la plupart des gestes étaient hérités et peu flexibles comme la majorité des signaux de communication animaux, mais que *certains* gestes, dans des contextes évolutionnaires 'neutres' comme le jeu ou le *grooming* semblaient plus flexibles et

répondaient aux critères de l'intentionnalité (Tomasello, 2008; Tomasello & Call, 2019). Une des conséquences visibles de cette approche est que la taille des répertoires de *gestes* rapportées par cette équipe est généralement inférieure à la taille des répertoires de gestes échantillonnés dans la nature et pour lequel une conception plus large de 'geste' est utilisée (Bourjade et al., 2020; Byrne et al., 2017; Liebal & Call, 2012). Une autre conséquence probable de l'approche de Tomasello et collaborateurs est que les contextes plus restreints dans lesquels sont observés ces gestes se prêtent à la répétition. En effet, le partenaire de jeu, la mère ou l'individu avec lequel on s'épouille, sont des partenaires avec lesquels il est possible d'avoir de nombreuses interactions répétées au cours du temps, et qui rendent possible la ritualisation ontogénétique[25] d'un geste ou sa négociation sociale. Inversement, les équipes ayant travaillé majoritairement sur des populations sauvages ont rapporté des répertoires de grandes tailles, largement partagés entre les différentes espèces de singes anthropoïdes et inter-compréhensibles d'une espèce à l'autre (voir une synthèse dans: Byrne et al., 2017; gorille: Genty et al., 2009; bonobo: Graham et al., 2017; chimpanzé: Hobaiter & Byrne, 2011a). Ces évidences soutiennent bien l'hypothèse d'une origine phylogénétique des gestes. Notons également que les premiers répertoires de gestes réalisés sur des singes non-anthropoïdes rapportent une flexibilité similaire à celle observée chez les grands singes en termes de dissociation geste/contexte, de variabilité de la taille des répertoires individuels et de variabilité selon les classes d'âge et de sexe (Gupta & Sinha, 2019; Molesti et al., 2020).

A la suite des études pionnières menées par Tomasello et collaborateurs (e.g., 1985, 1994), la très grande majorité des études a porté sur l'usage intentionnel des gestes communicatifs. A ce stade, il est important de souligner qu'une partie de ces études souffre d'un raisonnement circulaire qui consiste à proposer que les gestes communicatifs soient intentionnels, avant d'examiner s'ils le sont avec une série de critères, qui mène indéniablement à montrer qu'ils le sont bel et bien. Cet écueil a été discuté à de nombreuses reprises (Bourjade et al., 2020; Hobaiter & Byrne, 2017; Liebal et al., 2019; Scott & Pika, 2012), semble en partie dépassé dans les études récentes (Aychet, Blois-Heulin, Palagi, et al., 2021; Fröhlich et al., 2017), et est absent de la majorité des études expérimentales qui utilisent généralement des définitions plus restreintes. Dans un souci de brièveté, j'illustre ici les principales caractéristiques

[25] Transformation d'une *action* mécaniquement effective en *signal* communicatif (action ritualisée) par le biais d'interactions répétées au cours du développement (et non au cours de l'évolution, auquel cas il s'agirait de ritualisation phylogénétique). Une définition complète de ce terme est présentée dans de nombreux ouvrages, avec la description du mécanisme de modelage (*shaping*) (e.g., Call & Tomasello, 2007; Pika, 2008).

intentionnelles généralement incombées aux gestes (Cf. Townsend et al., 2017) avec des études menées sur la communication intraspécifique - tandis que les apports des études expérimentales réalisées avec un partenaire humain seront discutés dans le Chapitre 3. La communication gestuelle des primates non humains imprègne tous les contextes de la vie sociale. A l'inverse de la majorité des vocalisations qui ne sont pas orientées vers un partenaire spécifique, les gestes sont *adressés* à un receveur particulier et attentif, par les moyens d'une *orientation visuelle* vers cet individu ; e.g., chimpanzés (Hobaiter & Byrne, 2011a), gorilles (Genty et al., 2009), orangs-outans (Liebal et al., 2006), babouins olive (Molesti et al., 2020), mangabeys à front rouge (Aychet, Blois-Heulin, Palagi, et al., 2021). De plus, tous les gestes ont une composante visuelle, mais certains ont également une composante tactile ou acoustique supplémentaire. Toutes les études menées à ce jour rapportent que les gestes sont principalement adressés à des partenaires visuellement attentifs, et que lorsque l'attention visuelle n'est pas disponible, des gestes de contact ou auditifs sont favorisés (Genty et al., 2009; Hobaiter & Byrne, 2011a; Liebal et al., 2006; Molesti et al., 2020). Cette *sensibilité à l'attention* du receveur sera discutée plus amplement dans les chapitres suivants (Chapitres 3 et 4).

Par ailleurs, les gestes sont émis avec un but communicatif qui est satisfait au moyen de la *réponse* du receveur, laquelle si elle n'est pas satisfaisante ou absente, est associée à une persévérance de la part du producteur pour obtenir l'effet escompté. La *persévérance vers le but* peut prendre la forme de la répétition du même geste ou d'une élaboration de la communication à l'aide d'une diversification des signaux, e.g., orangs-outans (Russon & Andrews, 2011), chimpanzés (Fröhlich, Wittig, et al., 2019c; Roberts et al., 2013), macaque à bonnet (Gupta & Sinha, 2019). Plusieurs études récentes attestent d'un développement similaire de ces compétences intentionnelles chez les individus immatures ; la proportion de gestes sensibles à l'état attentionnel du partenaire augmente avec l'âge chez le chimpanzé (Dafreville et al., 2021; Fröhlich, Wittig, et al., 2019c) et le babouin olive (Molesti et al., 2020) ; de même que la proportion de gestes associés à des regards (e.g., Molesti et al., 2020) et à une persévérance vers le but communicatif (e.g., Fröhlich, Wittig, et al., 2019c). Notons également que ces marqueurs d'intentionnalité se retrouvent adéquatement utilisés dans les *règles conversationnelles* qui structurent les négociations sociales lors, par exemple, de l'initiation d'un déplacement (e.g., chimpanzé: Fröhlich et al., 2016) ou d'une demande de portage (e.g., bonobo: Rossano, 2013) entre mères et enfants.

Bien qu'ils surviennent dans des contextes fonctionnels différents, les gestes des singes non-anthropoïdes ont été étudiés pour leurs significations précises (*meaning*) déterminées à l'aide

de la réponse du receveur, i.e., l'ASO (*Apparently Satisfying Outcome*), représentant une issue apparemment satisfaisante du point de vue du signaleur. L'effet perlocutoire d'une réponse satisfaisante est la cessation de la communication par le signaleur (Cartmill & Byrne, 2010; Hobaiter & Byrne, 2014). Avec cet outil, il devient possible de construire un lexique des significations exprimées avec les différents gestes du répertoire. Comme le montrent Hobaiter & Byrne (2014), la polysémie existe, mais la plus grande flexibilité vient de l'usage de gestes différents pour 'dire la même chose'. Par la suite, les auteurs ont catégorisé une deuxième fois leurs répertoires de gestes selon la signification, déterminée du point de vue du receveur chimpanzé, i.e., avec l'ASO (Hobaiter & Byrne, 2017)[26]. Certains gestes pourraient également avoir une fonction *déictique* dans la communication, comme le *directed-scratch* des chimpanzés (Pika & Mitani, 2006), les postures tenues par les macaques à bonnet lors des séances d'épouillage (Gupta & Sinha, 2016) ou le *beckoning gesture* des bonobos, également iconique[27] (Genty & Zuberbühler, 2014) ; trois instances de communication où le geste désigne une entité extérieure, en l'occurrence spatiale, à la sémanticité de l'énoncé.

Les singes anthropoïdes sont également singuliers dans leur manière de gesticuler en *séquence*. Gorilles, bonobos, chimpanzés et orangs-outans ont été observés en train de gesticuler en longues séquences, notamment en absence d'une réponse immédiate du receveur (Cartmill & Byrne, 2007; Genty & Byrne, 2010; Graham et al., 2017; Hobaiter & Byrne, 2011b; Liebal et al., 2006; Roberts et al., 2013). Deux cas de figure émergent des études : (i) après avoir attendu une réponse qui n'est pas survenue, le signaleur répète le même geste ou change de geste (Cartmill & Byrne, 2007; Hobaiter & Byrne, 2011b; Liebal et al., 2006), ou (ii) le signaleur réalise une série rapide de gestes souvent variés (Genty & Byrne, 2010; Hobaiter & Byrne, 2011b). Alors que le premier cas représente un moyen de persévérance ou d'élaboration en cas d'échec de la communication, le deuxième cas, qui concerne principalement des individus immatures, est plus couramment interprété comme une emphase sur les gestes du signaleur qui augmenterait par ce moyen les chances d'obtenir l'effet escompté. En effet, la fréquence des séquences 'rapides' diminue avec l'âge chez les chimpanzés, tandis que la probabilité de sélectionner un geste efficace augmente (Hobaiter & Byrne, 2011b). Ces résultats ont été interprétés comme un 'élagage' du répertoire de gestes phylogénétiquement hérités que les immatures explorent puis sélectionnent (Byrne et al., 2017). Les séquences de gestes des

[26] Ces études ont débouché sur une large expérimentation de sciences participatives où des participants naïfs étaient amenés à donner leur avis sur la signification des gestes des chimpanzés (https://greatapedictionary.ac.uk).

[27] C'est-à-dire qui reproduit la forme d'un mouvement ou d'un objet auquel il fait référence.

chimpanzés répondent également à la loi de Menzerath selon laquelle plus longue est la séquence, plus courts sont ses constituants (Heesen et al., 2019) ; et celles des gorilles satisfont la loi de Zipf selon laquelle la taille des items est négativement corrélée à leur fréquence d'usage (Genty & Byrne, 2010).

La préférence manuelle pour la communication gestuelle fournit d'autres sources d'évidence en soutien à l'hypothèse gestuelle des origines du langage. Les arguments sont synthétisés notamment par Hopkins & Vauclair (2012) et Meguerditchian et al. (2011). La force de l'argument repose sur la découverte de biais populationnels robustes pour la préférence de la main droite chez certaines espèces (e.g., chimpanzé: Hopkins et al., 2005; e.g., babouin olive: Meguerditchian, Molesti, et al., 2011; e.g., macaque de Tonkéan: Meunier et al., 2013) – bien que d'autres espèces montrent des préférences individuelles sans biais populationnels (e.g., bonobo: Chapelain et al., 2011; e.g., capucin moine: Meunier & Vauclair, 2007). D'autant plus que le biais est plus marqué pour les gestes communicatifs que pour les tâches manuelles simples ou complexes (Meguerditchian & Vauclair, 2014) et plus robuste aux facteurs positionnels et environnants (Bourjade et al., 2013; Chapelain et al., 2012).

Chez les chimpanzés captifs, cette préférence manuelle se traduit par une asymétrie *anatomique* dans le gyrus frontal inférieur et *fonctionnelle* avec l'activation prédominante de l'homologue de l'aire de Broca dans l'hémisphère gauche (Taglialatela et al., 2006, 2008). Une asymétrie anatomique contra-latérale a récemment été découverte dans la partie médiane du sillon central des cerveaux de babouins olive droitiers ou gauchers pour une tâche manuelle complexe (Margiotoudi et al., 2019). La spécialisation hémisphérique pour les gestes, et a fortiori les gestes communicatifs, pourrait être un caractère ancestral présent chez les singes de l'ancien monde. Par ailleurs, la découverte de neurones miroirs dans le cortex prémoteur ventral du macaque rhésus (Gallese et al., 1996) a procuré un regain d'intérêt pour l'hypothèse gestuelle. En effet, ces neurones s'activent à la réalisation d'un geste manuel, en particulier alimentaire (mettre quelque chose à la bouche), et à la vue de cette même action perpétrée par un autre individu (Ferrari et al., 2003). Le pouvoir 'magique' des neurones miroirs repose sur la suggestivité de leur action pour une série d'activités motrices et cognitives centrales dans la communication coopérative et le langage (Uddin et al., 2007), à savoir : (i) le lien moteur et cognitif/représentationnel entre la main et la bouche, (ii) les capacités d'imitation motrice manuelle et orofaciale, (iii) le lien entre la conscience de soi et d'autrui et les possibles dérivations, e.g., empathie, théorie de l'esprit (Kampe et al., 2003; Uddin et al., 2007), (iv) la multi-sensorialité de ces neurones et leur capacité à mettre la voix de l'humain en miroir, i.e.,

ils déchargent à la production vocale et à la perception de l'écho de voix (Keysers et al., 2003; Rizzolatti & Craighero, 2004).

2.1.2.3. La communication multimodale des primates non humains

La communication multimodale, définie comme l'usage de signaux multimodaux ou de combinaisons[28] de signaux émis dans différentes modalités sensorielles (Fröhlich & van Schaik, 2018) reste marginalement étudiée à ce jour. L'émergence de cet intérêt émane sans doute de la prise de conscience qu'étudier la communication unimodale ne permettait pas d'étudier adéquatement l'évolution du langage moderne, conçu comme un phénomène multimodal (Levinson & Holler, 2014; Slocombe et al., 2011; Vigliocco et al., 2014; Wilcox & Xavier, 2013). Les études qu'il est possible de regrouper sous ce terme sont disparates et peu nombreuses, mais citons les intérêts de ces approches pionnières en regard de (i) l'intégration audiovisuelle de l'expression orale, (ii) l'étude des gestes sonores, (iii) l'étude des combinaisons de gestes et de vocalisations, ou plus généralement de différentes modalités perceptibles.

L'intégration des modalités visuelles et auditives de la parole est le mode par défaut de la compréhension de la parole humaine dès le plus jeune âge (Driver, 1996; Patterson & Werker, 1999, 2003). Les primates non humains produisent également des expressions faciales associées à la communication vocale, de même que silencieuses (Ghazanfar, 2013; Micheletta et al., 2013; Waller et al., 2017). A l'instar des humains, les macaques, capucins et chimpanzés reconnaissent les correspondances visuelles et auditives des vocalisations de leur espèce (Ghazanfar, 2013). Toutefois, une différence fondamentale émerge quant au rythme auquel sont produites ces vocalisations ; la parole humaine est produite à un rythme moyen de 5 Hz (intervalle de 3-8 Hz) créé par les oscillations mandibulaires correspondant à l'alternance des consonnes et des voyelles (Bergman et al., 2019; Ghazanfar, 2013). A cet égard, les vocalisations des PNH sont plus 'figées' dans la mesure où la bouche ne bouge pas de manière rythmique (Bergman et al., 2019). Il y a cependant des exceptions intéressantes sur le plan de l'évolution de la parole. Plusieurs études sur le *lip-smacking*, une expression orofaciale affiliative correspondant à une oscillation rythmique de la bouche, avec les dents visibles ou

[28] Il s'agit de considérer les combinaisons de signaux au sens large, c'est-à-dire la production de plusieurs signaux de manière synchrone ou asynchrone, avec un chevauchement ou alignés.

non, avec ou sans rétractation du scalp (Micheletta et al., 2013) ont mis en évidence un rythme similaire à la parole humaine (e.g., gelada: Bergman, 2013; macaque: Ghazanfar et al., 2012). Chez le macaque à crête, cette expression peut être produite de manière silencieuse ou accompagnée d'un léger *grunt* qui accroit la probabilité d'obtenir une réponse affiliative (Micheletta et al., 2013). Chez le gelada, les vocalisations accompagnant les *lip-smacks* sont dénommées *wobbles* et possèdent des caractéristiques rythmiques similaires à celles de la parole humaine (Bergman, 2013a; Gustison et al., 2012; Gustison & Bergman, 2017). Chez l'orang-outan captif *Rocky,* Lameira et collaborateurs (2016) ont montré que les babils de cet individu étaient produits au rythme caractéristique du langage. Dans l'ensemble, la pertinence scientifique des expressions orofaciales pour l'évolution du langage a été minimisée du fait qu'elles ont longtemps été considérées comme innées, émotionnellement déclenchées, et inflexibles, y compris chez l'humain (Fitch, 2010; Johnson, 2005; Matsumoto & Willingham, 2006, 2009; Sherwood et al., 2003). Toutefois, plusieurs études récentes montrent qu'elles satisfont les critères de la communication intentionnelle, suggérant qu'elles sont plus flexibles qu'il n'y parait (e.g., mangabey à front rouge: Aychet, Blois-Heulin, Palagi, et al., 2021; chimpanzé: Florkiewicz & Campbell, 2021).

D'autres sources d'évidence ont mis en avant l'usage de gestes manuels possédant une composante audible ; ci-après *gestes sonores*. Ces gestes sonores sont utilisés par les mères orangs-outans pour coordonner leurs déplacements avec leur progéniture (Fröhlich, Lee, et al., 2019a), ou par les chimpanzés adolescents pour communiquer avec leur mère lorsqu'elle est visuellement inattentive (Dafreville et al., 2021). En captivité, les gestes sonores consistant à frapper la cage ou à frapper dans ses mains ont été documentés chez plusieurs espèces (e.g., babouin: Bourjade, Meguerditchian, et al., 2014; chimpanzé: Meguerditchian et al., 2012) et sont utilisés envers un partenaire humain visuellement ou socialement inattentif (Bourjade, Meguerditchian, et al., 2014; Leavens et al., 2004, 2010). Ces situations de ruptures attentionnelles ont également permis de révéler l'usage combiné de gestes et de vocalisations apprises, émises pour regagner l'attention d'un expérimentateur humain par les chimpanzés (Hopkins et al., 2007) et activant l'homologue de l'aire de Broca dans le cerveau du chimpanzé (Taglialatela et al., 2008, 2011). L'intérêt évolutionnaire des gestes sonores pour l'évolution du langage repose sur le recours à la modalité auditive de la part des *receveurs*. En effet, une critique générale qui peut être faite à l'étude de la communication chez les primates non humains et son biais sur la *production*, au détriment de la *compréhension* – et ce, à l'exception des expériences de repasse vocale (*playback experiment*) (Fröhlich & van Schaik, 2018). Or,

indépendamment des causes multiples et des origines diverses qui ont permis au langage d'évoluer en tant que système d'émission de signaux, il est bien sûr nécessaire de s'intéresser à la manière dont la réception de ces signaux a évolué, au moins pour les raisons suivantes : (i) les receveurs sont les plus à même d'exercer une pression de sélection sur les signaux de communication (Higham & Hebets, 2013; Maynard-Smith & Harper, 1995; Partan & Marler, 1999; Rendall et al., 2009; Rowe, 1999), (ii) la redondance, la désambiguïsation, ou la complémentarité des signaux s'observe dans les réponses des receveurs en fonction du contexte et de l'expérience des individus (Fröhlich & van Schaik, 2018; Higham & Hebets, 2013), (iii) en étudiant la perception des signaux par les receveurs, et notamment leur contexte d'usage, il est possible d'identifier des forces sélectives ayant favorisé la prédominance de la modalité vocale chez l'humain (Fitch, 2010). L'intérêt des ruptures attentionnelles sera développé plus en avant dans les deux chapitres suivants (Chapitres 3 & 4).

L'étude des combinaisons multimodales de signaux, bien que balbutiante, fournit des informations cruciales et complémentaires sur les fonctions proximales et ultimes de la multimodalité et les possibilités de complémentarité de signification (*meaning*) qu'offre le découplage modal (Fröhlich & van Schaik, 2018). A cet égard, la documentation d'un geste instrumental permettant aux orangs-outans qui le produisent de modifier la structure acoustique d'un cri offre la première indication de combinaison sémantique multimodale, i.e., le *Kiss squeak* (Lameira et al., 2012). En effet, le *Kiss squeak* permet aux mâles qui le produisent de baisser la fréquence fondamentale du cri sans altérer les autres composants acoustiques, et de paraitre ainsi plus gros, ce qui présente un intérêt face aux mâles rivaux. Chez les chimpanzés, deux études ont rapporté l'usage de combinaisons de gestes et de vocalisations. Chez les chimpanzés de Kanyawara, les combinaisons multimodales permettent d'augmenter le taux de réponse en comparaison à une vocalisation produite seule, mais pas à un geste seul (Wilke et al., 2017). Chez les chimpanzés de Budongo, les combinaisons multimodales sont utilisées après un échec de communication vocale, mais pas après un échec de communication gestuelle (Hobaiter et al., 2017). Ces résultats sont interprétés par les auteurs en termes d'information publique : les vocalisations sont perçues par une large audience tandis que les gestes silencieux ne sont perçus que par les receveurs immédiats. Ajouter une vocalisation dans ces contextes 'intimes' ne serait pas toujours au bénéfice des interactants. Une interprétation alternative est que le geste vient spécifier l'*adressage* de la vocalisation à un receveur spécifique, ajoutant une dimension ostensive à la communication (Moore, 2016). Dans un registre similaire, les femelles macaques qui approchent une autre femelle ont plus de chances de recevoir un comportement

affiliatif qu'agonistique si elles couplent leur approche d'un *grunt* adressé à la femelle approchée (Silk et al., 2000). Une récente étude chez les mangabeys à front rouge a examiné les séquences de signaux en détaillant les modalités et les composantes (geste manuel ou corporel, vocalisation, expression faciale) des séquences. Les auteurs rapportent huit clusters de séquences de signaux souvent observés en association. La proportion de séquences multimodales est supérieure dans les contextes de jeu social et d'agression, tandis que les séquences unimodales sont majoritaires dans les contextes sexuels, affiliatifs, de soumission et d'épouillage (Aychet, Blois-Heulin, & Lemasson, 2021). Ces approches intégrées qui permettent de dépasser la dichotomie visuel/auditif ou geste/vocalisation sont à privilégier dans la recherche de passerelles entre geste et voix (Cf. Chapitre 4).

2.2. Les apports d'une psychologie évolutionnaire, développementale et comparée

2.2.1. La fonction adaptative du développement

> *« Nous vivons sur l'avenir : « demain », « plus tard » […]. Ces inconséquences sont admirables, car enfin il s'agit de mourir. Un jour pourtant l'homme constate qu'il a trente ans. Il affirme ainsi sa jeunesse. Mais du même coup, il se situe par rapport au temps. Il y prend sa place. […] Demain, il souhaitait demain, quand tout lui-même aurait dû s'y refuser. Cette révolte de la chair, c'est l'absurde. »*
>
> *Albert Camus (1986, p.30)*

Les théories gestuelles et musicales de l'évolution du langage qui sont articulées dans cette recherche reposent en grande partie sur l'ontogénèse des facultés de communication et sur la relation parent-enfant. A cet égard, il est nécessaire de se munir d'une théorie du développement qui soit compatible avec cette approche. La psychologie développementale évolutionnaire s'intéresse aux *avantages adaptatifs*[29] des caractéristiques développementales des espèces (Geary & Bjorklund, 2000). En s'appuyant sur la théorie des traits d'histoire de vie, elle s'accompagne d'une conception épigénétique du déterminisme développemental permettant d'articuler l'héritage propre à l'espèce avec une plasticité développementale propre à l'individu, qui prend lieu et place dans son environnement écologique, social et historico-culturel (Bogin, 1997; Bronfenbrenner, 1977; Geary & Bjorklund, 2000). Plusieurs théories allient les notions

[29] C'est-à-dire qui confère un avantage en termes de survie ou de reproduction à son porteur. Dans une vision étendue de l'hérédité, il est possible de considérer la transmission de variants culturels équivalents à la transmission de gènes à la génération suivante (Jablonka & Lamb, 2014; Micheletti, 2020).

de préformisme, de biologie et d'évolution avec une approche épigénétique du développement (Jouen & Molina, 2007). La plupart reposent sur la notion de *paysage épigénétique* introduite par le généticien Conrad Waddington (2014). Jusqu'alors, les biologistes pensaient que les innovations observées chez les organismes étaient le produit de la sélection naturelle. Waddington opéra un tournant majeur en proposant que le rôle *créatif* appartienne au processus développemental : une grande partie de la variation entre individus, qui constitue le matériau sur lequel opère ensuite la sélection naturelle, provient du développement lui-même (Gottlieb, 1987; Ho, 1998). Si une *invention développementale* est adaptée à l'environnement, elle est répétée par l'individu et par tous ceux chez qui elle émerge ; ce faisant, elle devient canalisée[30]. La notion de *canalisation épigénétique* conjugue l'héritage biologique à un ensemble de comportements probables dont seuls une petite portion émerge. Ainsi, le déterminisme est seulement partiel et il est compatible avec le rôle actif de l'individu dans son développement (Jouen & Molina, 2007; Lerner, 1998) et avec une vision située, en contexte écologique (Bronfenbrenner, 1977). Ces changements développementaux forment ainsi une base de changement évolutionnaire (Gottlieb, 1987).

Les études comparatives suggèrent qu'une des fonctions d'une maturité différée dans le temps est la possibilité d'ajuster et d'entrainer les compétences physiques et sociales de l'individu nécessaires à la vie adulte dans un environnement donné (Geary, 1995; Geary & Bjorklund, 2000; Joffe, 1997). Chez les primates, la période d'immaturité allongée semble être une réponse adaptative à la complexité sociale des espèces : les espèces qui ont le développement le plus long, sont celles qui ont le néocortex le plus développé et la société la plus complexe (Joffe, 1997). L'espèce humaine possède cinq *stades d'histoire de vie*, conçus comme des étapes incontournables de l'ontogenèse composée des phases de croissance, maturation et vieillissement, propres à l'espèce et séquentiellement ordonnées, i.e., c'est-à-dire relevant de l'orthogénèse. Le premier est le stade *nourrisson* durant lequel le bébé est nourrit au lait[31]. Dans les sociétés traditionnelles, le sevrage[32] se déroule entre 2 et 3 ans, âge auquel le jeune entre dans le stade de l'*enfance* qui n'existe pas chez les autres primates. L'enfance prend fin vers

[30] Dans une variante proche, la *théorie de la dynamique des systèmes* s'appuie sur cette idée de canalisation épigénétique ; les *attracteurs développementaux*, c'est-à-dire des patrons de comportements qui émergent comme étant les plus appropriés dans un environnement donné, s'avèrent être plus ou moins canalisés (Jouen & Molina, 2007).

[31] Sein maternel ou biberon

[32] Entendu comme l'arrêt du lait maternel (ou de son substitut alimentaire)

sept ans quand le jeune humain possède sa dentition définitive lui permettant de s'alimenter comme un adulte ; c'est l'entrée dans le stade *juvénile*, qui s'étend jusqu'à la puberté. La puberté signe le début de *l'adolescence* qui s'achève à l'âge adulte lorsque l'individu présente un comportement sexuel adulte.

L'*enfance*, qui est un stade propre à l'espèce humaine, offre à l'individu la possibilité de poursuivre sa maturation cérébrale sous l'influence des stimulations socio-éco-culturelles et ainsi de se développer en s'adaptant à l'environnement local. Par ailleurs, les enfants humains conservent des caractéristiques néoténiques[33] qui continuent de stimuler les soins parentaux et l'attention qui leur est préférentiellement adressée. Du point de vue du parent, l'enfance représente un allongement de la période néonatale à coût évolutif réduit : la mère n'allaitant plus, les soins allo-parentaux[34] deviennent possibles, ce qui favorise la coopération et soutient les apprentissages (Bogin, 1997). La transition du stade nourrisson au stade de l'enfance met donc fin à un coût parental élevé, notamment pour les mères allaitantes, et modifie le conflit parent-enfant d'investissement attentionnel (Bogin, 1997; Trivers, 1974). Ce point crucial pour ce travail de recherche sera développé plus en avant dans le Chapitre 4 (section 4.4.1.).

2.2.2. Les conceptions *evo-devo* appliquées au langage

2.2.2.1. Canalisation épigénétique de l'acquisition du langage

Comme cela a été évoqué plus haut, il existe une dizaine de gènes séquencés qui jouent un rôle dans les compétences linguistiques (Enard et al., 2002; Mozzi et al., 2016). Concernant FOXP2, la souris, le chimpanzé et l'humain partagent la majorité de la séquence d'ADN excepté deux nucléotides[35] qui ont muté depuis la divergence entre l'humain et le chimpanzé et semblent être en partie responsables des capacités articulatoires uniques de notre espèce. A ces bases génétiques, s'ajoutent des compétences linguistiques présentes à la naissance.

Aldrige et collaborateurs (1999) ont par exemple, exposé 16 bébés de moins de deux jours (âgés de 4 à 33 heures) à une expérience présentant un visage de femme en train de prononcer les voyelles /a/, /u/, /y/, /i/. Dans les conditions « congruentes » le visage et le son étaient

[33] Caractéristiques juvéniles ou d'immaturité conservés au cours de la vie de l'individu et notamment à l'âge adulte.

[34] Soins promulgués par des non-parents, comme un grand frère, une grand-mère ou un adulte non apparenté.

[35] Elément de base de la séquence d'ADN

concordants (e.g., visage qui prononce /a/ avec le son /a/). Dans les conditions « incongruentes », le visage et le son étaient discordants (e.g., visage qui prononce /a/ avec le son /u/). Le comportement relevé chez les nouveau-nés était la durée du premier regard, un indicateur d'orientation sociale très utilisé chez les nouveau-nés. Pour les voyelles /i/, /y/ et /a/, les bébés regardaient plus longtemps l'image quand elle était concordante avec le son. Les auteurs concluent que les bébés naissent avec une sensibilité structurale aux mouvements visibles de la parole, strictement analogue à la sensibilité structurale pour ses caractéristiques auditives. L'exposition des bébés aux sons de leur langue maternelle durant la vie fœtale ne permet pas d'expliquer directement ce résultat puisque si le fœtus est exposé à la voix de sa mère, il ne perçoit pas son visage.

Perani et al. (2011) ont étudié l'activation cérébrale[36] de nouveau-nés de deux jours auxquels une histoire était racontée. Trois voix étaient présentées ; la voie naturelle, la voie fredonnée et la voie d'automate, c'est-à-dire débarrassée de ses contours prosodiques, émotionnels et personnels. A deux jours, le cerveau des nouveau-nés traite les sonorités de la langue comme les adultes traitent la parole, et ce, à quelques différences près. En effet, bien que les mêmes aires cérébrales soient activées, l'hémisphère droit, en charge de la prosodie[37] est plus activé que le gauche, alors que c'est l'inverse chez les adultes qui traitent le contenu sémantique du discours. Par ailleurs, les bébés réagissent à la voie naturelle et à la voie fredonnée, mais pas à la voix d'automate, montrant que l'attention à la mélodie de la langue est sélective dès la naissance. Enfin, la connectivité entre les neurones au sein de chaque hémisphère n'est pas du tout développée en comparaison du cerveau adulte, ce qui indique que l'exposition au langage est absolument nécessaire pour développer la fonction linguistique.

Ces compétences néonatales ont été interprétées comme l'évidence d'une canalisation épigénétique : certaines compétences sont héritées biologiquement, et orientent l'attention du nouveau-né vers certaines stimulations (e.g., la prosodie), quand toutes les autres se développent avec l'expérience et sur la base des premières acquisitions. Bien que ces deux études concernent la perception du langage, elles soulignent deux caractéristiques cruciales sur

[36] A l'aide d'une technique d'IRM (imagerie par résonnance magnétique) fonctionnelle qui permet de localiser les zones du cerveau où un surcroît d'activité se produit lors d'une exposition à un stimulus donné.

[37] Aspects mélodiques de la langue comprenant les variations de tons, d'accent, d'expression, d'émotion…

le plan évolutionnaire : (i) la dimension audiovisuelle de la perception de la parole, et (ii) la sensibilité aux contours prosodiques qui indique un traitement 'musical' de la langue.

Cette canalisation prend également la forme d'une période sensible pour apprendre à parler, à comprendre, mais surtout à extraire les règles grammaticales de la langue maternelle et se les approprier de manière à devenir capable de produire une infinité de phrases jamais produites avant (Pinker, 1986). Les exemples quasi-expérimentaux, et bien qu'anecdotiques, des *enfants sauvages* (Malson & Itard, 1985) montrent qu'une fois cette période passée, les enfants ne parviennent pas à développer un langage normal. Victor de l'Aveyron, recueilli et éduqué par le médecin Jean Itard (1801, Dans: Malson & Itard, 1985) a appris à s'exprimer et à se faire comprendre à l'aide de lexigrammes, il a aussi maitrisé la compréhension, mais n'a jamais parlé, ni maitrisé la grammaire du français. En résumé, la canalisation épigénétique renvoie au degré de résistance qu'exprime une fonction développementale face aux variations de l'environnement (Jouen & Molina, 2007).

2.2.2.2. Linguistique développementale évolutionnaire

Locke (2009) pose les bases d'un cadre théorique transdisciplinaire permettant d'étudier les apports développementaux à l'évolution du langage et les apports évolutionnaires au développement du langage. Son propos s'appuie sur les fonctions adaptatives que peuvent remplir les compétences langagières à chaque stade de trait d'histoire de vie – et non comme c'est le cas usuellement, uniquement chez les individus adultes. En particulier, il propose une théorie d'acquisition de la flexibilité vocale basée sur une sélection parentale (Locke, 2006). Son propos s'insère aisément dans les théories de communication dirigée vers les apparentés (*kin communication models*) tel que formulé par Fitch (2010). Locke (2006) compare la dynamique des pleurs des nourrissons avec le pattern des infanticides recensés aux Etats-Unis. Les pleurs ont des effets physiologiques sur les receveurs adultes (e.g., augmentation du taux de testostérone, accélération de la fréquence cardiaque) qui incitent à l'action (activation de la réponse au stress) et au soin envers le nourrisson. Les pleurs peuvent aussi devenir un réel facteur de stress pour l'adulte s'ils sont anormalement fréquents. Ils possèdent des caractéristiques acoustiques spécifiques si l'enfant est molesté, en détresse ou malade (Locke, 2017), comme cela est le cas chez les macaques (Maestripieri et al., 2000) et très vraisemblablement chez les babouins (observation personnelle). Locke (2006) montre que la fréquence des pleurs diminue à deux reprises au cours de la première année de vie : à 3 et 7

mois, de manière concomitante à la raréfaction des infanticides (Shapiro-Mendoza et al., 2006) et des décès naturels (Murphy et al., 2013). A deux mois, la fréquence des pleurs commence à diminuer et s'accompagne d'un changement de nature « expressive » à « communicative »[38]. A partir de 3 mois, l'enfant gazouille et entame une phase d'exploration vocale au cours de laquelle il produit des sons de plus en plus divers, qui acquièrent une flexibilité fonctionnelle de plus en plus complète (i.e., sont détachés du contexte affectif dans lequel ils sont produits) et qui sont peu à peu utilisés pour attirer l'attention (Oller et al., 2013, 2016). A partir de 7 mois, la majorité des nourrissons produit généralement ses premiers babils, alternant voyelles et consonnes de manière à former des syllabes (Oller et al., 2013, 2021). Ainsi, les deux épisodes de raréfaction des infanticides sont associés à deux jalons du développement vocal du nourrisson, à savoir : les gazouillis et le babillage. Locke (2006) propose que ces deux compétences vocales procurent des avantages sélectifs aux nourrissons en signalant aux parents qu'ils sont en bonne santé et qu'il n'est donc pas inutile de poursuivre l'investissement parental. A ce titre, Locke (2006) propose une synthèse des études montrant que le babillage induit de l'engagement parental, stimule les soins apportés aux nourrissons et est « plaisant » pour les parents. Sur le plan évolutionnaire, l'auteur propose que les fonctions de ces différentes vocalisations et les réponses, notamment vocales, induites chez les parents aient augmenté avec l'impéritie (*helplessness*) des nourrissons après le dilemme obstétrical, et aient contribué à l'émergence de la flexibilité vocale dans la lignée humaine. A la fin du stade nourrisson, toutes les dimensions *structurelles* du langage ont commencé à se développer (e.g., phonologiques, lexicales, morphologiques et syntactiques), de même que ses dimensions sociocognitives (e.g., intentionnalité, pragmatique) (Locke & Bogin, 2006). Il est dès lors possible de constater les corrélats socioculturels de l'acquisition linguistique – une autre dimension adaptative indéniable – avec des écarts de l'ordre de 6 mois de développement entre les enfants des catégories socio-professionnelles supérieures et inférieures (e.g., Fernald et al., 2013).

Durant l'enfance, une continuelle amélioration de la phonologie, du lexique, de la longueur des phrases et de la syntaxe est observée. De même, le discours devient plus fluide et l'enfant commence à utiliser le langage de manière créative, par exemple avec le développement de l'humour (Locke & Bogin, 2006). Avec l'émergence de la théorie de l'esprit et de la compréhension du temps (Spreng & Grady, 2010; Suddendorf, 1997; Suddendorf & Corballis, 2007), l'enfant commence également à utiliser le langage pour s'exprimer à propos des faits et

[38] C'est-à-dire adressée à un destinataire.

des entités qui ne sont pas immédiatement perceptibles – une caractéristique centrale du langage dénommée *displacement* (Hockett, 1960). C'est durant la phase juvénile que le jeune humain développe non pas les aspects structurels du langage mais ses usages, y compris l'usage en contextes appropriés, les compétences narratives et discursives qui élargissent la dimension pragmatique de la langue (Cf. section 2.2.3.). La juvénilité procure ainsi une opportunité de développer les usages probants et attractifs de la langue et préparent à l'usage 'compétitif' de la langue observé à l'adolescence (Locke & Bogin, 2006). L'adolescence est ainsi marquée par des changements profonds de l'usage de la langue. La fluidité et la rapidité d'élocution augmentent drastiquement, le contenu sémantique évolue avec un renouvellement du lexique, et l'incorporation d'idiomes ou encore d'argot. L'adolescent peut également modifier son matériel linguistique acquis précocement, comme la phonologie. Mais l'usage du langage prend une fonction sociale cruciale dans l'adaptabilité de l'individu à son environnement : il est utilisé pour le commérage de manière particulièrement exacerbée, de même que pour moquer les amis ou tromper, molester, négocier, persuader, le tout s'accompagnant de l'apparition du sarcasme (Locke & Bogin, 2006). Ces dimensions essentiellement pragmatiques du langage ne peuvent être investies avant de nombreuses années d'acquisition, alors même qu'elles sont cruciales pour le statut social et reproducteur des individus. Une hypothèse avancée par les auteurs est que l'achèvement d'un haut statut social et d'un bon succès de reproduction pourrait être lié à la capacité à attirer l'attention, chez les primates non humains, comme chez les humains (Chance, 1967; Locke & Bogin, 2006).

Un nombre croissant d'études vient également étayer l'hypothèse que les compétences langagières sont associées au comportement sexuel (cour, attractivité, choix du partenaire, nombre de partenaires), et plus marginalement à des succès de reproduction différentiels (nombre d'enfants et succès de reproduction des enfants) (synthèse dans Locke & Bogin, 2006). Les études tendent à montrer que les femmes jugent les hommes avec des tons de voix graves plus attractifs, en particulier dans des contextes de cour, et d'autant plus durant la période d'ovulation (Hughes et al., 2004; Locke, 2017). Dans certaines populations humaines, le ton grave de la voix prédit le succès reproducteur des hommes (Apicella et al., 2007). Inversement, les hommes jugent plus attractives les voix de femmes lorsqu'elles sont enregistrées en période de plus haute fertilité qu'en période de basse fertilité ; les caractéristiques acoustiques de la voix variant au cours du cycle ovarien (Shoup-Knox et al., 2019). De plus, les 'ornements' discursifs et narratifs accroissent l'attractivité des hommes pour les femmes et inversement (Lange, 2011), tandis que les imperfections langagières présentent la tendance inverse (Locke,

2009). Par ailleurs, l'usage énergique, volontaire et extensif du langage oral, par exemple chez les acteurs ou chez les avocats qui plaident, sont associés à des taux de testostérone élevés (Locke & Bogin, 2006). Bien que la plupart de ces études restent muettes sur la descendance engendrée en fonction des compétences langagières, et donc échouent à pointer le réel succès de reproduction, elles convergent sur les liens existants entre compétences vocales et langagières d'une part et la physiologie et le comportement reproducteur d'autre part. Dans le cadre de ce travail, retenons que l'entièreté du développement est nécessaire au jeune humain afin de maitriser les aspects langagiers complexes, en particulier ses usages sociaux et sexuels, dans lesquels la dimension pragmatique est indispensable à l'intercompréhension.

2.2.2.3. Les apports du cas particulier de l'extinction des langues

Un autre aspect adaptatif du langage concerne la connaissance partagée par une communauté de locuteurs. Deux études récentes permettent d'établir un lien entre les langues menacées de disparition et la connaissance biologique des plantes et des animaux, mais aussi de leurs vertus médicinales (Cámara-Leret & Bascompte, 2021; Kik et al., 2021). En comparant la connaissance des plantes médicinales dans trois régions du monde, l'Amérique du nord, le nord-est de l'Amazonie et la Papouasie Nouvelle-Guinée, les auteurs montrent que 75% des 12 495 plantes médicinales décrites à ce jour ne sont connues que dans une seule langue – une absence de redondance caractéristique de l'héritage bio-ethno-culturel. Les analyses montrent notamment que les langues menacées soutiennent la connaissance de 86% des plantes connues dans une seule langue en Amérique du nord et 100% en Amazonie[39] (Cámara-Leret & Bascompte, 2021). La disparition de ces langues menace donc d'affecter profondément la connaissance que nous avons des plantes médicinales. La Papouasie Nouvelle-Guinée héberge plus de 10% des langues naturelles parlées par de petites communautés de locuteurs – i.e., des langues menacées. Dans une étude de grande ampleur menée sur plus de 6000 élèves de la Papouasie Nouvelle-Guinée, les auteurs ont documenté la connaissance linguistique de la langue indigène et évalué les connaissances ethnobiologiques associées à chaque population. Les résultats montrent que seulement 58% des participants sont en pleine maitrise de la langue indigène (comparé à 91% de leurs parents) et que leurs connaissances des oiseaux locaux et de l'usage traditionnel des plantes est corrélative de la maitrise de la langue. La plupart des noms

[39] De manière ironique, les plantes médicinales décrites dans une langue menacée ne sont pas, en tant qu'espèces de plante, menacées d'extinction !

de plantes et d'oiseaux manquent de mots équivalents en anglais, rendant l'accès à la connaissance indigène dépendante de la langue indigène (Kik et al., 2021).

Ces deux études soulignent admirablement un des bénéfices du langage qui dépasse l'adaptation de l'individu seul. En effet, c'est tout une communauté de locuteurs qui détient et maintient une connaissance indigène, condition de son partage avec d'autres communautés linguistiques. Cet exemple d'intelligence ou de cognition distribuée souligne la nécessité d'inclure les sciences sociales dans la compréhension des phénomènes … qui s'expriment à l'échelle sociale (Servais, 2012). Le langage en est un, indéniablement. A l'inverse de la biologie ou des sciences cognitives qui conçoivent la performance au niveau de l'individu, et même « *dans* » l'individu – dans ses *adaptations* pour la biologie ou dans son *cerveau* pour les sciences cognitives, les sciences sociales proposent une vision de la performance *distribuée* à l'échelle de plusieurs individus – la dyade, la famille, la société (Hinde, 1983; Servais, 2012). La conception du langage défendue dans ce travail entre pleinement dans la vision sociale et distribuée du phénomène linguistique et du comportement langagier. De plus, l'approche « sociologique » permet d'approcher la complexité sociale des phénomènes, mais aussi des groupes (Strum, 2012), soulignant la nécessité et l'intérêt d'étudier la pragmatique, dans une perspective développementale, évolutionnaire mais également sociale – ce qui sera développé dans la section suivante, et dans les perspectives à la fin du Chapitre 4.

2.2.3. La pragmatique développementale dans une perspective comparée

> *« Si répondre est de la part d'un enfant un comportement d'insolence qui peut susciter une punition, ne pas répondre est une faute plus difficilement repérable. Ne pas répondre accorde un délai, permet de s'habituer à l'idée de la défaite, et surtout, de faire durer de quelques minutes supplémentaires le temps de jouer. Vécu dans la mauvaise foi, mais aussi dans l'exaltation du sursis, cet entre-deux d'une surdité feinte est l'apprentissage d'une liberté non de l'affrontement mais de l'écart, du non-dit : le pire en regard de la morale des familles, de sa volonté de mainmise. L'affrontement, aussi violent soit-il est une façon d'adhérer, de reconnaitre une autorité. »*
>
> *Chantal Thomas (1998, p.30)*

Etudier la communication chez différentes espèces et cultures requiert une 'standardisation' des échanges sociaux et communicatifs à des fins comparatives. La Théorie des Actes de Langage (*Speech Act Theory*) initialement développée par John Austin (1962), puis reprise notamment par John Searle (1962) a ouvert le champ à l'étude de la pragmatique et plus généralement à l'étude des *intentions* et des *effets* que véhiculent les énoncés de communication. Le propos

central de la théorie des actes de langage est que l'analyse logique traditionnelle de certains énoncés (en termes de vérité) est inadéquate pour capturer les *effets* de ces énoncés dans le monde extérieur. Au contraire, certaines phrases ne sont pas simplement des descriptions de l'état du monde mais des états ou des évènements en elles-mêmes (e.g., faire une promesse, formuler une requête). Ainsi, il est parfois possible *d'agir avec des mots* et Austin (1962) propose que chaque énoncé puisse être caractérisé par trois *actes de langage* distincts : locutoire, illocutoire et perlocutoire. L'acte locutoire est la signification sémantique logique de l'énoncé. L'acte illocutoire est l'intention communicative véhiculée par l'énoncé (et l'énonciation) et l'acte perlocutoire est l'effet que cet énoncé et cette énonciation engendrent dans la réalité. Ainsi dans l'exemple : « Regarde papa le chien », l'acte illocutoire est une requête attentionnelle, ou une demande de partage et l'acte perlocutoire renferme les effets probables tels que : le père regarde le chien, le père est content que son enfant souhaite lui montrer quelque chose, le père répond quelque chose à l'enfant.

Toutefois, les philosophes du langage intéressés par la pragmatique négligent ses aspects développementaux (Bernicot, 1994). C'est Jérôme Bruner (1975) dans *Ontogenesis of speech acts*, qui adapte les concepts de la pragmatique à l'étude du développement langagier. Il s'inspire explicitement de la théorie de la *zone proximale de développement* de Vygotski selon laquelle l'asymétrie de compétence entre l'adulte et l'enfant permet à l'enfant d'atteindre une performance supérieure que s'il était seul (Bernicot, 1994; Vygotsky, 1962). Il développe les concepts d'étayage (*scaffolding*), de tutorat (*tutoring*) et de formats d'interaction (*interaction format*) qui permettent d'étudier, chez l'enfant développant le langage, l'émergence progressive des fonctions pragmatiques (perlocutoires, puis illocutoires) et de focaliser l'étude sur les *usages* du langage, en lieu et place des *formes* de langage (Bernicot, 1994; Bruner, 1975).

De manière intéressante, les trois actes de langage (i.e., locutoire, illocutoire, perlocutoire) ne se développent pas au même rythme chez l'enfant (Bates et al., 1975). Encore plus intéressant est le fait que l'acte locutoire est en fin de compte le dernier des trois à émerger. Dans la lignée de Piaget (1936), Bates et al. (1975) proposent que les signaux de communication puissent être utilisés intentionnellement à partir du cinquième stade sensori-moteur. Toutefois, les signaux du bébé ont des effets factuels (perlocutoires) sur l'entourage dès la naissance (Austin, 1962; voir aussi Oller et al., 2016, fig.2, p.387). L'effet perlocutoire a été décrit indépendamment par Wallon (1936) dans la manière dont les nouveau-nés utilisent leur tonus postural et leurs mouvements pour entrer en interaction avec les adultes environnants, puis plus tard, comme une manière d'exprimer leurs émotions. En accord avec cette vision, Iverson & Thelen (1999)

et Trevarthen & Aitken (2001) suggèrent que l'incarnation du langage débute bien avant l'émergence du langage lui-même, en s'enracinant dans les premiers liens main-bouche de la phase sensori-motrice, les échanges multimodaux entre un nouveau-né et sa mère caractérisés par la rythmicité et la musicalité d'une (proto-)conversation. Ainsi, ces actes perlocutoires, puis illocutoires sont enchâssés dans les routines interactionnelles entre l'enfant et l'adulte, lesquelles préfigurent et étayent la transition de la période pré-linguistique à linguistique (Bruner, 1975).

Dans une étude pionnière, Bates et al. (1975) étudient le développement des actes perlocutoires et illocutoires chez l'enfant préverbal. En accord avec la théorie des actes de langage, elles définissent comme *perlocutoires* les actes qui ont des effets sur le receveur sans que ceux-ci résultent d'une convention entre 'locuteurs'. Les nourrissons ne commençant à produire des gestes de pointage que vers 10 mois, les auteurs concluent qu'ils ne produisent que des actes perlocutoires avant les premiers pointage (Bates et al., 1975; mais voir Rodríguez et al., 2015 pour un contre-argument). Les actes *illocutoires* ont pour caractéristique d'être produits de manière sélective, c'est-à-dire adressés à un receveur particulier, avec l'intention d'obtenir un effet sur ce dernier. Autrement dit, ils sont produits *intentionnellement* pour leur *valeur de signal*. L'étude de Bates et al. (1975) détaille comment le nourrisson passe d'un stade perlocutoire à illocutoire à locutoire, laissant supposer un phénomène transitoire d'un type d'acte de langage à l'autre, alors qu'il s'agit plus vraisemblablement d'un phénomène additif comme proposé ci-dessus. Ces études ont ouvert la voie à un champ entier d'investigation des signaux communicatifs de l'enfant préverbal, dont le pointage est certainement l'exemple canonique.

L'autre point fécond de la théorie des actes de langage est que l'intention communicative peut être décomposée en différentes natures selon la réponse de l'adulte qui satisfait le mieux l'enfant. Bates et al. (1975) proposent que la fonction *proto-déclarative* émerge quand le nourrisson cherche à obtenir des réponses attentionnelles différentes de la proximité physique. Le nourrisson découvre que *l'attention de l'adulte*, ici caractérisée par le sourire, le rire, le contact visuel et la parole, peut être obtenue, maintenue ou même regagnée en s'exhibant ou en exagérant ses comportements. Peu après, le nourrisson incorpore un troisième élément à cette stratégie, généralement un objet qu'il montre sans intention de le donner, mais générant une dynamique attentionnelle *triadique* à l'origine de l'attention conjointe (Bates et al., 1975; Camaioni, 1997; Carpenter et al., 1998; Rodríguez et al., 2015). Les premiers gestes de pointage émergent peu après et sont généralement de nature *proto-impérative*, c'est-à-dire utilisant

l'adulte pour obtenir un objet, avant d'être *proto-déclaratif*, c'est-à-dire utilisant un objet pour obtenir l'attention de l'adulte (Camaioni, 1997). Ce cadre analytique a été massivement investi par les psychologues du développement et les comparativistes (primatologues) afin d'étudier la nature intentionnelle des gestes de pointage, leur émergence, leur combinaison avec les premiers mots et leur fonction déictique dans les premières phrases (e.g., Aureli et al., 2017; Bourjade, 2016; Bourjade et al., 2020; Camaioni, 1997; Guidetti, 2002, 2005; Leavens, 2004). Des distinctions plus récentes permettent par exemple de séparer les pointages déclaratifs en pointages *informatifs* et pointages *expressifs*, selon qu'ils visent à donner une information à l'adulte ou à recevoir l'attention de ce dernier (Colonnesi et al., 2010; Tomasello, Carpenter, et al., 2007), bien que les éléments observables à disposition du chercheur pour réaliser cette catégorisation sont parfois minces ou biaisés (Bourjade et al., 2020). Certaines de ces études sur le pointage seront succinctement évoquées dans le paragraphe suivant pour leur pertinence en regard de la question centrale de notre travail.

En résumé, la théorie des actes de langage articulée à la théorie vygotskienne par Jérôme Bruner, a donné naissance à la *pragmatique développementale* comme champ d'étude (Bernicot, 1994; Bruner, 1975; Vygotsky, 1962). Ce champ a par la suite donné lieu à des études pionnières sur le développement de l'enfant, qui ont servi de base à un grand nombre d'études sur la nature intentionnelle des signaux de communication chez l'enfant et le primate non humain. C'est également de cette théorie que découle l'opérationnalisation des concepts de *communication intentionnelle* (Leavens, 2004; Townsend et al., 2017), *d'attention sociale* (Bourjade, 2017; Franco & Butterworth, 1996) et de *sensibilité à l'attention sociale* qui seront présentés avec la problématique de cette recherche (section 2.3.3.).

2.2.4. Origines développementales du langage humain

Il est souvent dit que le langage ne fossilise pas, empêchant l'étude empirique de son histoire évolutive. Mais n'étant pas inscrit dans les gènes ni hérité de la génération précédente, il possède des origines développementales qu'il est possible d'étudier pour comprendre les mécanismes sociaux, comportementaux, et cognitifs qui permettent son acquisition. De plus, les gestes communicatifs jouent un rôle fondamental dans l'acquisition du langage, ce qui permet d'établir des parallèles avec la communication gestuelle des primates non humains et de fournir un support empirique aux hypothèses gestuelles de l'évolution du langage. Sans rentrer dans les innombrables détails du processus d'acquisition langagier au cours des

premières années de vie, je dresse ci-dessous un portrait succinct des grandes étapes de ce développement, en indiquant certains cadres théoriques connexes qui peuvent être utilisés à chaque phase, et en illustrant chacune des étapes avec quelques études empiriques choisies visant à souligner les apports de la psychologie développementale et de la psycholinguistique à l'étude de l'évolution du langage.

2.2.4.1. La communication dyadique du nourrisson

Après une exposition aux stimuli sociaux environnants durant leur vie fœtale, les nouveau-nés sont capables d'orienter leur tête ou leur corps vers diverses stimulations sociales – la réponse *d'orientation sociale* est utilisée comme indicateur de leurs perceptions et d'un développement typique (Als et al., 1977; Chevallier et al., 2012). Dès deux mois, les bébés engagent des interactions dyadiques avec les adultes par des regards mutuels et montrent des signes clairs d'engagement en répondant à l'adulte par des regards, mouvements de sourcils et sourires (Beebe et al., 2016; Lavelli & Fogel, 2013). Les études des dyades mère-enfant de cet âge-là montrent que le comportement de face-à-face est très présent, avec la mère qui s'approche du visage de l'enfant de très près. Les nouveau-nés sont sensibles à la temporalité des interactions. Plus généralement, *le modèle de la synchronie bio-comportementale* de Feldman (2012) propose un cadre théorique permettant d'explorer les relations de contingences temporelles observées entre les mères et leurs nourrissons. La synchronisation est comportementale, mais elle prend également effet au niveau physiologique, au niveau des battements cardiaques et de l'activité cérébrale (Bell, 2020). Sur le plan comportemental, les bébés, sont sensibles à la *temporalité* des interactions dès 2 mois (Beebe et al., 2016; Bell, 2020). Par exemple, des nourrissons de 2 mois sourient davantage à leur mère ou à un adulte montrant la même temporalité de réponse (Bigelow & Rochat, 2006). Dès 4 mois, les nourrissons sourient plus dans un jeu de *« coucou caché »* organisé que désorganisé (les différentes séquences présentées dans le désordre), suggérant que les bébés de cet âge forment des attentes à propos de la temporalité et la structure des interactions (Rochat et al., 1999).

A cet égard, *le modèle de l'intersubjectivité* de Trevarthen (2011) propose que l'orientation sociale, la sensibilité à la subjectivité d'autrui et l'intention communicative soient des compétences innées du nouveau-né qui permettraient de rendre compte de l'organisation, de la rythmicité et de la musicalité des interactions dyadiques entre la mère et l'enfant. A cet âge-là, le nourrisson produit des mouvements corporels, de tête et de bras et/ou jambe caractérisés par

la répétition et le rythme et qui permettent progressivement d'entrer en interaction (Iverson & Thelen, 1999; Thelen, 1981) et de communiquer des émotions (Trevarthen, 2011). Il produit également des pleurs, des *grunts* végétatifs[40], et des protophones qui ne sont pas socialement orientés (Locke, 2006; Oller et al., 2013, 2016). Les protophones, qui sont des vocalisations hors pleurs composées uniquement de voyelles, n'expriment pas de valence émotionnelle particulière (i.e., flexibilité fonctionnelle) et constituent la principale expression vocale du bébé de 3-4 mois qui semble explorer sa voix (Oller et al., 2013, 2021). Les parents répondent de manière contingente à 30-50% de ces vocalisations pré-linguistiques (hors pleurs), et le taux de réponse est corrélé au taux de vocalisations et influence le développement phonologique (Goldstein et al., 2003; Goldstein & Schwade, 2008). Lorsque l'adulte chante ou parle, le bébé de 5 à 9 mois tend à réduire la quantité de ses mouvements corporels (Jover et al., 2019; Nakata & Trehub, 2004; Scola et al., 2015). De même, de récentes études montrent que les bébés de cet âge tendent à éviter de vocaliser quand la mère parle, et ce, dès l'âge de 3 mois (Bourjade et al., soumis; Hilbrink et al., 2015; Northrup & Iverson, 2020). Un nombre croissant d'études souligne que la coordination des échanges entre mère et enfant prend la forme de proto-conversations caractérisées notamment par des tours de parole et par la musicalité communicative de la voix de la mère, rythmée par les mouvements de l'enfant alignés sur la prosodie maternelle (Esteve-Gibert & Guellaï, 2018; Gratier et al., 2015; Gratier & Trevarthen, 2008; Rochat et al., 1999; Trevarthen, 1999).

Selon *la théorie de l'attachement* (Bowlby, 1969), cette contingence temporelle entre le parent et l'enfant participe de la formation du lien d'attachement, par le biais notamment d'une *responsivité* parentale (Ainsworth, 1963; Ainsworth et al., 2015) adéquate associée à un développement socio-émotionnel ultérieur 'positif' (Mesman et al., 2009). Ces mécanismes ont été explorés entre autres avec les expériences dites du *still-face* (visage impassible) initialement développées pour montrer que l'enfant jouait un rôle actif dans les échanges communicatifs dès le plus jeune âge (Tronick et al., 1978; Tronick & Cohn, 1989). Dans le paradigme du *still-face*, le bébé est exposé à une série de conditions durant lesquelles la mère, ou l'adulte, alterne une communication face-à-face naturelle et des périodes de visages neutres (*still face*) et de non responsivité. Les marqueurs émotionnels importants sont observés au cours de la phase de visage neutre, et dans la comparaison avec la phase de communication naturelle qui lui précède ou lui succède. Généralement, les bébés tendent à éviter le visage neutre en regardant ailleurs,

[40] Liés à l'effort (McCune et al., 1996).

et les comportements positifs diminuent pendant la période de non responsivité, tandis que les comportements négatifs augmentent (Mesman et al., 2009). La manière dont les auteurs classent les comportements positifs et négatifs, de même que les variantes du *still-face* qui sont utilisées génèrent une quantité de résultats difficiles à comparer. Toutefois, il est intéressant dans le cadre de cette synthèse, de souligner que des bébés de 5 mois augmentent la quantité de vocalisations hors pleurs durant la période de non responsivité alors qu'ils réduisent la quantité de sourires et que cette sensibilité à la rupture interactionnelle (et attentionnelle) est corrélée au nombre de mots et de phrases comprises à 13 mois (Goldstein et al., 2009).

2.2.4.2. Emergence des interactions triadiques et de l'intentionnalité

A l'inverse des études sur le développement communicationnel au cours des six premiers mois de vie qui sont relativement multimodales, les études portant sur les acquisitions faites entre 7 et 14 mois tendent à se concentrer sur une seule des deux modalités, auditive ou visuelle (avec des exceptions toutefois, e.g., Murillo et al., 2021). Les raisons peuvent être théoriques, méthodologiques, ou techniques, il est important de souligner que cela participe du manque de connaissance sur la *réorganisation modale* du signalement intentionnel qui s'opère entre 10 mois et l'entrée dans le langage, vers 14 mois en moyenne (Guidetti et al., 2014; McCune & Zlatev, 2015).

De nombreuses divergences théoriques entourent la question de l'intentionnalité des signaux de communication, et plus particulièrement des pointages, de leur nature ostensive et coopérative, et de la capacité à trianguler l'attention. Dans la lignée de Trevarthen (2011), les *approches innéistes* considèrent que la communication du bébé humain est intentionnelle dès la naissance. Dans cette perspective, il n'y a pas de communication qui ne soit pas intentionnelle et la question de l'intentionnalité des signaux est une non question[41]. Ce n'est pas ce cadre théorique qui sera retenu ici, pour des raisons évidentes de manque de comparabilité avec les primates non humains. Toutefois, même s'il est impossible de le savoir, la communication *pourrait* être intentionnelle dès la naissance. Cela dit, ce n'est qu'entre 8 et 12 mois que le nourrisson exhibe tous les marqueurs d'une communication intentionnelle sur le plan *comportemental* (Bates et al., 1975; Carpenter et al., 1998; Liszkowski et al., 2004). Les

[41] De manière collatérale, il n'y a pas non plus de critères d'intentionnalité associés à la production de signaux, rendant la théorie inutilisable à des fins de comparabilité avec les primates non humains sur cette question.

approches cognitivistes proposent que la communication intentionnelle apparaisse après la révolution cognitive des 9 mois au cours de laquelle émerge la compréhension de l'état attentionnel et intentionnel d'autrui (Tomasello, 1999). Les *approches gradualistes* promeuvent au contraire que l'émergence des signaux comme les pointages précède l'émergence de leur usage intentionnel, référentiel[42], ostensif[43] et représentationnel[44] (Bates et al., 1975; Cochet & Vauclair, 2010b; Franco & Butterworth, 1996; McCune & Zlatev, 2015). Ce positionnement théorique est compatible avec les *approches non mentalistes* d'un apprentissage progressif par l'enfant (Franco & Butterworth, 1996; Leavens, 2004; Matthews et al., 2012) et avec les *approches socio-constructivistes* qui promeuvent une appropriation progressive des outils de communication via la médiation de l'adulte/de la culture (Salomo & Liszkowski, 2013; Vygotsky, 1978). La même analyse peut être appliquée aux interactions triadiques, c'est à dire impliquant un objet entre l'enfant et l'adulte ; la coordination de l'attention avec un adulte vers une entité extérieure (i.e., *l'attention conjointe*) émerge sur le plan comportemental à la fin de la première année (Bakeman & Adamson, 1984; Carpenter et al., 1998; Carpenter & Call, 2013), mais il est possible d'en observer les précurseurs tout au long de la première année (Bakeman & Adamson, 1984; Rodríguez et al., 2015).

Les approches gradualistes et cognitivistes s'opposent principalement sur la relation de *causalité* entre le comportement et l'état de compréhension du monde social qui le motive. Une illustration de ce débat peut être donnée avec le concept d'ostension (discuté dans Bourjade, 2016). Tomasello et collaborateurs (Tomasello et al., 2005; Tomasello & Carpenter, 2007) proposent une ligne de partage entre les pointages humains et non humains basés sur l'intentionnalité partagée. Selon cette proposition, le pointage humain est *ostensif* et ne peut émerger que lorsque l'enfant maitrise la capacité d'intentionnalité partagée (*shared intentionality*), qui permet une intention communicative de deuxième ordre (Dennett, 1971), c'est-à-dire une intention de communiquer qui s'accompagne de la croyance que cette intention sera perçue comme telle (Grice, 1957; Tomasello, 2008). Du point de vue gradualiste, les pointages émergent de *quelque chose*, et leur utilisation ostensive se développe

[42] Pour faire référence à une entité extérieure, généralement distale.

[43] C'est-à-dire dans l'intention de se faire comprendre et avec la croyance que cette intention est perçue et partagée par le receveur. C'est une dimension coopérative de la communication qui ne porte pas sur la véracité du contenu sémantique mais sur les intentions partagées des locuteurs de se comprendre mutuellement (Moore, 2016; Tomasello et al., 2005). Noter qu'il s'agit d'une acception quelque peu différente de celle qu'en font Rodriguez et ses collaborateurs, sur les gestes impliquant un objet (Moreno-Núñez et al., 2020; Rodríguez et al., 2015).

[44] C'est-à-dire qui désigne des entités extérieures non visibles ou abstraites.

progressivement. Par exemple, Rodriguez et ses collaborateurs (Moreno-Núñez et al., 2020; Rodríguez et al., 2015) proposent que les gestes impliquant un objet soient des précurseurs développementaux de la fonction déictique distale associée aux pointages. De ce point de vue, il est possible d'observer des pointages qui ne sont pas encore utilisés de manière ostensive. Ces divergences ont des conséquences importantes sur le plan comparatif, discutée à différentes reprises (Bourjade, 2016; Bourjade et al., 2020). Du point de vue de Tomasello (2006, 2008), les grands singes ne pointent pas par manque d'intentionnalité partagée (ou de second ordre). Or, d'un point de vue comportemental, ils pointent, et la question de savoir si ces pointages sont ostensifs ou ne le sont pas est une question empirique (Graham et al., 2020; Moore, 2016). Le présent travail reste agnostique sur la relation causale entre subjectivité et comportement et s'inscrit dans une démarche comportementale, non mentaliste, où le gradualisme des apprentissages individuels fournit un *mécanisme* à l'approche socio-constructiviste des influences environnementales, sociales et historico-culturelles (Bourjade, 2016; Bourjade, Gaunet, et al., 2019; Bourjade et al., 2020; Leavens et al., 2017; Strum, 2012; Vygotsky, 1978).

Sur le plan factuel, les nourrissons de six mois commencent à alterner le regard entre un objet distal et l'adulte et opèrent ainsi une transition progressive vers les *interactions triadiques* (Bakeman & Adamson, 1984; Carpenter & Call, 2013). Entre 8 et 12 mois, les nourrissons commencent à diriger l'attention de l'adulte vers des entités externes par le moyen des alternances de regards entre l'adulte et l'objet, puis des gestes de pointage (Beuker et al., 2013; Camaioni et al., 2004; Carpenter et al., 1998; Liszkowski et al., 2012). L'attention conjointe, définie comme la capacité à coordonner sa propre attention avec l'attention d'autrui vers une entité externe (Bourjade, 2017; Carpenter et al., 1998; Klein et al., 2009; Scaife & Bruner, 1975), est initialement proposée par l'adulte et l'enfant y adhère en suivant le regard de l'adulte vers l'objet (Beuker et al., 2013; Shepherd, 2010). C'est ensuite l'enfant qui va prendre l'initiative de diriger le regard de l'adulte vers un objet, au moyen d'un pointage, ou d'une vocalisation (Beuker et al., 2013). Cet engagement triadique est considéré comme un marqueur de la communication référentielle (Cochet & Vauclair, 2010b; Colonnesi et al., 2010; Liszkowski et al., 2007; Tomasello, Carpenter, et al., 2007) bien que la transition vers la représentation (c'est à dire la référence à un objet non visible ou à une abstraction) nécessite un développement ultérieur (McCune & Zlatev, 2015; Murillo & Casla, 2021). Les gestes de pointage sont universellement produits à travers les cultures humaines (Liszkowski et al., 2012; Salomo & Liszkowski, 2013) et permettent, de même que les interactions triadiques, de prédire les premières acquisitions langagières ultérieures (Cochet & Byrne, 2016; Colonnesi et al.,

2010; Kita, 2003; Liszkowski et al., 2012), avec toutefois des différences entre cultures (Kita, 2003; Mastin & Vogt, 2016; Salomo & Liszkowski, 2013).

Vers 6-7 mois, les nourrissons produisent également leurs premiers babils, généralement accompagnés de mouvements rythmiques temporellement alignés avec le babillage, et représentant des précurseurs prosodiques (Esteve-Gibert & Guellaï, 2018; Locke, 2006). A ce stade, les vocalisations des bébés ne sont pas produites en regardant l'adulte et pour cette raison, ne sont pas considérées comme intentionnelles (Oller et al., 2013). La coordination temporelle des vocalisations et des regards vers l'adulte émerge progressivement au cours de la première année de vie. Les expériences de *still-face* indiquent par exemple que les bébés de moins de 9 mois détournent le regard face au visage neutre et augmentent les vocalisations négatives (Mesman et al 2009). Toutefois, dans des conditions plus naturelles d'interaction, les bébés de 10 mois ne coordonnent pas leur attention visuelle avec leurs vocalisations (D'Odorico et al., 1997; Donnellan et al., 2019; Harding & Golinkoff, 1979), mais commencent à associer leur vocalisation avec des regards vers l'adulte autour de 11-12 mois (D'Odorico et al., 1997; Donnellan et al., 2019; Wu & Gros-Louis, 2014, 2017). La synchronisation des vocalisations et des regards est considérée comme le marqueur classique de la communication intentionnelle chez l'enfant (Csibra, 2010). Toutefois, une étude ayant catégorisé les vocalisations comme *communicatives* si elles survenaient durant une interaction avec un adulte, ou étaient associées à un pointage ou à un geste ostensif ou bien encore étaient associées à un regard, parviennent à une conclusion différente (Esteve-Gibert & Prieto, 2013). Avec cette conception plus large de la dimension communicative, fondée sur la communication visuelle, les auteurs montrent que le ton des vocalisations est significativement plus haut et que leur durée est plus courte quand les vocalisations sont communicatives que lorsqu'elles sont exploratoires. Les contours prosodiques permettent également de distinguer différentes intentions communicatives (e.g., le mécontentement, la requête, la satisfaction) dans les vocalisations des enfants de 7 mois (Esteve-Gibert & Prieto, 2013). Par la suite, les séquences associant pointages et vocalisations chez les enfants de 10 à 14 mois peuvent être différenciées sur la base des contours prosodiques des vocalisations en fonction de leurs intentions communicatives (Aureli et al., 2017; Grünloh & Liszkowski, 2015). Ces études éclairent la nature *intermodale de l'intention communicative* et du développement langagier qui a été sous-estimé jusqu'ici. Les approches multimodales présentées dans le paragraphe suivant tentent de réparer cet écueil.

2.2.4.3. Multimodalité et sensibilité à l'attention

Les approches multimodales de l'acquisition du langage et de la communication chez l'enfant sont sans conteste les meilleures alliées théoriques de la résolution du problème de la transition des modalités mis en évidence à l'échelle évolutionnaire. Trois grandes approches sont présentées ci-dessous et seront reprises dans la section 2.3. pour leurs apports au problème de la transition.

La théorie de la dynamique des systèmes offre une vision incarnée (*embodied*) des activités perceptives, motrices et cognitives qui émanent toutes trois des *propriétés corporelles* du sujet (Jouen & Molina, 2007; Smith & Thelen, 2003). Appliquée au développement du langage, elle s'accommode parfaitement avec la vision multimodale portée par McNeill (2012) selon laquelle les gestes et la voix sont pour le langage, deux manifestations siamoises émanant d'une même source. Iverson & Thelen (1999) distinguent quatre étapes développementales permettant de relier les activités de la main avec les activités de la bouche. Les liens initiaux (*initial linkages*) s'observent de la naissance à 6 mois et consistent en des liens réflexes d'activation de la main et de la bouche (e.g., *Babkin reflex*), les mouvements de la main en direction de la bouche sont très courants et sont peu à peu orientés vers un but (réaction circulaire secondaire, puis tertiaire selon la théorie piagétienne). Entre 6 et 8 mois, l'émergence du contrôle moteur (*emerging control*) s'observe dans les comportements rythmiques du nourrisson : c'est l'apparition du babillage vocal sur lequel s'alignement les mouvements de bras, de tête et de corps, comme des marqueurs prosodiques gestuels. Entre 9 et 14 mois, le bébé entre dans la phase de couplage flexible (*flexible coupling*) des gestes et des premiers mots : les gestes précèdent les mots, gestes et mots n'ont pas les mêmes référents et lorsqu'ils sont produits ensemble, ils le sont de manière séquentielle. Dans une étude récente, Murillo et ses collaborateurs (2021) montrent par exemple que les mouvements rythmiques des enfants de 9 mois produits de manière chevauchante avec des vocalisations, permettent de prédire l'émergence des premiers pointages proximaux (*reaching)* à 12 mois. Enfin, dans la phase de couplage synchrone (*synchonous coupling*), entre 16 et 18 mois, les bébés produisent des mots et des gestes de manière coordonnée et synchrone. L'analyse des enfants qui bégayent ou de ceux qui montrent des délais pour entrer dans le langage oral montre que ces enfants présentent également des délais sur les gestes : à l'échelle de la phrase pour le bégaiement (les gestes sont suspendus et 'attendent' les mots) et à l'échelle du développement individuel pour les enfants parleurs tardifs (*late talkers*) (Iverson & Thelen, 1999).

Une autre proposition théorique offre des mécanismes à la formation du lien (*linkage*) entre les modalités communicatives. Selon McCune & Zlatev (2015), la transition vers la référence représentationnelle est un phénomène multimodal. Les premiers gestes de pointage précèdent de quelques semaines les premiers mots articulés verbalement. McCune & Zlatev (2015) décrivent ces mots comme étant limités au contexte, c'est-à-dire désignant des entités directement perceptibles pour l'enfant. Autrement dit, le *displacement* n'est pas encore en place. Les auteurs s'intéressent à la *référence représentationnelle*, c'est-à-dire à l'usage d'un signe faisant référence à une entité non directement perceptible, ou abstraite, comme la plupart des mots de la langue. De manière intéressante, la *transition modale* entre le pointage *visuel* et l'usage *verbal* de la référence représentationnelle (i.e., les mots non directement liés au contexte perceptible) pourrait s'étayer sur la dimension représentationnelle des *grunts communicatifs* (McCune, 2021). Le couplage entre *l'intention* et *la modalité* communicative s'exprime d'abord avec les gestes de pointage, puis l'articulation des premiers mots contextuels (Guidetti et al., 2014; Iverson & Thelen, 1999; McCune & Zlatev, 2015). A ce stade, l'enfant n'a pas encore stabilisé ses schèmes moteurs vocaux[45]. McCune & Vihman (2001) observent que ces schèmes vocaux se stabilisent *après* l'apparition des premiers mots contextuels, mais *avant* la généralisation à la référence représentationnelle et l'usage des *grunts* communicatifs[46]. Noter que les g*runts* sont répandus chez tous les mammifères et ont été largement étudiés chez les primates non humains pour leur prévalence dans les contextes sociaux de diverses natures (Dunbar, 1998; Ey et al., 2009; Silk et al., 2018). Chez l'enfant, le *grunt* prend d'abord la forme d'un son végétatif associé à l'effort (*effort grunt*), puis d'un son associé à la concentration et à l'attention (autodirigée) portée à une tâche (*attention grunt*), avant d'être utilisé par l'enfant à des fins communicatives (*communication grunt*) manifestées par des regards, et/ou des gestes adressés à l'adulte (McCune et al., 1996). Ces *grunts* communicatifs permettent de prédire l'apparition des premiers mots référentiels (au sens représentationnels) chez l'enfant (McCune et al., 1996; McCune & Zlatev, 2015). La proposition de McCune (2021) est que le *grunt,* d'abord produit dans l'effort physique, puis dans la concentration, soit associé pour l'enfant, à un état interne *signifiant* pour lui. Lorsque le *grunt* commence à être utilisé « pour ses effets sur

[45] McCune & Vihman (2001) développent une méthodologie permettant d'identifier la stabilisation phonologique de la production vocale des enfants. Il s'agit d'observer 10 répétitions de sons de consonnes supra glottales (voisées ou non) durant une période de 3 à 4 mois jusqu'à répétition exacte du même pattern moteur et sonore.

[46] McCune et collaborateurs (1996) définissent le *grunt* comme une vocalisation qui résulte de la fermeture de la glotte par l'activation des muscles du larynx suivie de l'émission soudaine d'une voyelle – avec la bouche fermée ou ouverte, mais sans intervention d'articulateur supra-glottal.

autrui », c'est-à-dire en tant que signal communicatif, l'association entre une signification (*meaning*) et une vocalisation (le *grunt*) est déjà en place pour l'enfant. De plus, la signification valable pour l'enfant de ses propres *grunts* initialement autocentrés relève d'une forme de représentation puisqu'elle ne désigne pas un objet perceptible à l'extérieur. L'auteur propose que le *grunt* communicatif facilite la transition vers le couplage entre intention communicative et vocalisation.

Dans la perspective de la théorie des actes de langage, les gestes et les mots sont deux véhicules pour communiquer une intention communicative, de la forme d'une assertion ou d'une requête par exemple (Searle, 1962). Cinq fonctions communicatives sont habituellement considérées : l'assertion (de la vérité de l'énoncé), la requête ou l'injonction (faite à autrui), l'expression (de son état psychologique), la promesse (faite à autrui), et la déclaration (à propos d'un état du monde). Les fonctions communicatives associées aux gestes de pointage ont été extensivement étudiées (e.g., Bates et al., 1975; Liszkowski et al., 2004; Tomasello, Carpenter, et al., 2007) mais il est intéressant de constater que le couplage des vocalisations permet d'identifier l'intention communicative (e.g., Grünloh & Liszkowski, 2015). Plusieurs études mentionnent le couplage du pointage avec des vocalisations et rapportent que ces séquences multimodales sont généralement de bons prédicteurs des compétences langagières quelques mois plus tard (e.g., Aureli et al., 2017; Donnellan et al., 2019; Igualada et al., 2015; Murillo & Casla, 2021). Aureli et al. (2017) rapportent que les enfants qui pointent à 12 mois produisent des pointages multimodaux à 15 mois et qu'à 18 mois, leur prosodie (i.e., durée, fréquence fondamentale, amplitude de changement de ton, contour final) permet de différencier les intentions communicatives associées aux pointages. Dans une étude récente sur le chevauchement multimodal entre gestes et vocalisations, Murillo & Casla (2021) rapportent que les gestes déictiques (pointer, montrer, donner) sont multimodaux dès le départ tandis que les gestes iconiques et conventionnels sont produits dans la seule modalité visuelle jusqu'à 24 mois. Entre 24 et 30 mois, la production des gestes iconiques et conventionnels (appelés représentationnels par les auteurs) augmente et permet de prédire les compétences de vocabulaire et de morphologie évaluées indépendamment. Dans une étude sur les gestes conventionnels, Guidetti (2005) rapporte que les enfants âgés de 16 à 30 mois combinent gestes et mots pour exprimer un accord ou un refus. Les enfants de 16 mois présentent une prédominance de gestes sur les mots, et la tendance s'inverse à 24 et 30 mois, tandis que la combinaison geste-mot reste stable. Ces résultats soulignent de nouveau le fait que geste et mot sont issus du même système communicatif et suggèrent que la transition des modalités est en réalité une *réorganisation* des

modalités dans la mesure où il y a une continuité fonctionnelle entre les deux et que les deux modalités continuent de coexister (Guidetti, 2005; Guidetti et al., 2014; Iverson & Thelen, 1999; McNeill, 2012).

D'autres sources d'évidence soulignent l'usage des vocalisations associées aux gestes de pointage par les enfants de 12 à 24 mois. Dans une étude sur la sensibilité à l'attention d'autrui, des enfants de 12 et 18 mois étaient observés en train de communiquer avec un adulte dans trois conditions expérimentales : l'adulte attentif à l'enfant et à la cible du pointage, l'adulte attentif à l'enfant seulement, l'adulte inattentif (Liszkowski et al., 2008). A 18 mois, les enfants augmentaient leur fréquence de vocalisations quand l'adulte était inattentif et après les pointages inefficaces. Cette étude a donné lieu à diverses réplications sur des enfants de 10 à 15 mois. Wu & Gros-Louis (2017) rapportent qu'à dix mois, les enfants augmentent leurs vocalisations seulement quand l'adulte les regarde au lieu de regarder la cible de leur pointage ; ce qui est également rapporté chez les enfants de 12 mois par Igualada et al (2015). Dans une étude plus écologique prenant en compte l'attention du parent avant le pointage et la réaction au pointage de l'enfant, Gros-Louis & Wu (2012) ont montré que les enfants de 12 mois vocalisent plus quand le parent ne les regarde pas, ou quand il ne regarde pas la cible du pointage, ou quand il répond au pointage en commentant sur autre chose que la cible du pointage. Ces emphases vocales associées aux pointages sont de bons prédicteurs des compétences langagières ultérieures (Wu & Gros-Louis, 2014). Ces études seront discutées plus en avant en regard du contexte attentionnel favorisant la multimodalité de la communication (section 3.4.1.2.).

2.3. Origines du couplage intention – modalité (vocale) de communication

2.3.1. Le problème de la transition des modalités et le contrôle vocal

L'état de l'art présenté ci-dessus souligne les apports des études portant sur la communication des primates non humains et sur l'acquisition du langage chez l'enfant à la compréhension de l'évolution du langage. Comme l'a pointé Fitch (2010), les modèles théoriques souffrent cependant d'une insuffisance explicative concernant l'aspect résolument *multimodal* du langage humain. Aucun modèle n'explique de manière satisfaisante le problème de la transition des modalités gestuelles/visuelles et vocales/auditives selon leurs usages dans le langage moderne – d'où le terme dédié de *problème de la transition des modalités* (Hewes, 1973; Kendon, 2011; Orzechowski et al., 2016). En effet, si un protolangage purement gestuel avait

préexisté au langage actuel, comment expliquer l'apparition du langage oral ? Inversement, si un protolangage purement vocal avait préexisté au langage actuel, comment expliquer la gestuelle et l'émergence du contrôle vocal dans la seule lignée humaine ? A cet égard, il est intéressant de noter que l'imitation vocale est présente chez les oiseaux chanteurs (e.g., Snowdon & Hausberger, 1997) et chez les cétacés (e.g., Rendell & Whitehead, 2001) mais absente chez les primates, mis à part chez l'humain. Des études neuroanatomiques et fonctionnelles comparées ont permis de mettre en évidence que la production vocale était sous contrôle subcortical chez la plupart des mammifères et des oiseaux, humains compris (Fitch, 2010). Toutefois, il existe une différence majeure chez l'humain au niveau des connexions cortico-laryngiennes qui permettent le contrôle volontaire des vocalisations. En effet, l'humain possède des connexions directes entre le cortex moteur et des neurones moteurs du tronc cérébral impliqués dans le contrôle du larynx. Les dommages orchestrés sur ces zones cérébrales provoquent la perte de la capacité vocale chez l'humain alors qu'elle n'a pas d'effet chez le macaque rhésus ou le singe écureuil (Jürgens & von Cramon, 1982). De même, des lésions au niveau de l'équivalent de l'aire de Broca dans la partie latérale inférieure du cortex prémoteur ne trouble pas la capacité vocale des singes tandis que cette région est impliquée dans le contrôle vocal volontaire chez l'humain (Jürgens, 2002).

A ce titre, il est intéressant de noter que les chimpanzés captifs capables de produire des sons non vocaux mais impliquant l'action des articulateurs supérieurs comme les lèvres, adressent ces sons volontairement et intentionnellement à un humain *inattentif* (Hopkins et al., 2007). Les chimpanzés qui produisent ces sons présentent une asymétrie anatomique et fonctionnelle plus importante dans l'homologue de l'aire de Broca au sein de l'hémisphère gauche (Taglialatela et al., 2008, 2011) indiquant un contrôle cortical (Fitch, 2010; Meguerditchian et al., 2014). Ainsi Fitch (2010) émet l'hypothèse que le contrôle cortical du *larynx* spécifiquement, et non du tractus vocal dans son entièreté, est absent chez les primates non humains. Les observations de proto-consonnes et la formation de dialectes de sons non-voisés chez les orangs-outans supportent cette hypothèse (Lameira, 2018). En revanche, l'observation d'un jeune orang-outan produisant des babils (Lameira et al., 2016) tend à la remettre en question. Dans l'ensemble, ces travaux suggèrent qu'il est nécessaire d'approfondir nos connaissances sur les capacités de contrôle vocal des primates non humains, incluant leurs performances et leurs déterminismes. Il est également nécessaire de mieux documenter l'émergence de ce contrôle chez l'enfant humain et chez les autres modèles animaux (oiseaux chanteurs, cétacés) capables d'imitation vocale. Enfin, à la suite de Fitch (2010), il est clair qu'il devient incontournable d'identifier des

forces sélectives plausibles responsables de cette émergence et de la réorganisation modale qui s'en suit. Je propose à cet égard d'examiner les contextes favorables à l'usage de la modalité vocale et de documenter systématiquement son association (synchrone, chevauchée ou alignée) à la modalité visuelle dans la communication des PNH et des enfants développant le langage.

A ce titre, les travaux sur l'acquisition du langage chez l'enfant humain nous sont d'une aide précieuse. Premièrement, la théorie de la dynamique des systèmes qui met en évidence les liens physiologiques et comportementaux entre geste et voix dès la naissance est une bonne candidate pour dépasser le problème de la transition des modalités (Iverson & Thelen, 1999). Deuxièmement, l'observation des enfants développant le langage éclaire la nature multimodale de la communication non verbale et le fait qu'il s'agisse plus d'une *réorganisation* des modalités communicatives que d'une transition *per se* puisque les modalités vocale et gestuelle continuent de coexister (Guidetti et al., 2014). Troisièmement, les enfants présentent un décalage temporel court entre la production de gestes déictiques et conventionnels et l'oralisation des premiers mots. Ce décalage est souvent expliqué comme un décalage de maturation entre le contrôle moteur des mouvements corporels et celui des articulateurs vocaux et supérieurs (Iverson & Thelen, 1999). Il a également été suggéré que ce décalage puisse venir du développement plus lent de la fonction représentationnelle associée aux sons (McCune, 2021) alors que cet élément est incontournable à la double articulation du langage oral et à la nature arbitraire de la relation entre signe et signification.

A cet égard, ce décalage offre une fenêtre d'investigation idéale sur la mise en place du *couplage entre intention communicative et modalité de communication*, qui peut être investiguée par les contextes dans lesquels ce couplage émerge. Suivant la proposition de Locke (2006), il est par exemple possible d'examiner l'effet des vocalisations émises par les nourrissons sur le comportement de leurs parents ; et les études empiriques tendent à montrer que les parents répondent plus à la communication multimodale (Donnellan et al., 2019; Gogate et al., 2001) et qu'ils s'engagent plus volontiers dans l'interaction quand l'enfant produit des babils que d'autres vocalisations (Locke, 2006, 2017). A ce stade, l'étude de l'usage des modalités gestuelles/visuelles et vocales/auditives au cours du développement pré-langagier s'impose pour comprendre les mécanismes du couplage entre intention communicative, modalité de communication et référence représentationnelle. De telles études permettraient d'éclairer le rôle de la coexistence des deux modalités dans l'attention conjointe et dans le maintien et le regain de l'attention parentale par les enfants préverbaux. De même, l'émergence des structures linguistiques pourraient être plus précoces dans la modalité visuelle que dans la

modalité auditive comme c'est le cas pour le lexique et la prosodie (Esteve-Gibert & Guellaï, 2018; Tomasello, 2008), sans que les mécanismes soient bien compris à ce jour. Il semble donc fructueux de mener plus en avant les travaux sur les enfants préverbaux afin d'identifier des contextes favorisant l'usage de la modalité vocale, mais également l'usage différentiel de chaque modalité dans l'émergence des structures linguistiques et des fonctions pragmatiques, i.e., émergence de la relation forme/référent/fonction, grammaticalisation d'un geste conventionnel ou d'un idéophone, prosodie gestuelle et vocale.

2.3.2. Etudier la flexibilité du couplage intention – modalité de communication

Les travaux passés, en cours et à venir présentés dans cette note de synthèse ont pour enjeu principal d'élucider comment les individus appartenant à différentes espèces aujourd'hui bien vivantes résolvent le problème d'intégration des modalités communicationnelles, visuelles et auditives, au cours du développement communicatif et à l'échelle des différentes structures linguistiques. Etudier le développement vocal ou gestuel séparément semble être une impasse à cet égard. L'approche proposée ici est focalisée sur la *communication intentionnelle*, où *l'intention communicative* peut être exprimée par différentes *modalités sensorielles* de manière synchrone ou asynchrone, pour ses effets[47] sur les receveurs. Le receveur est donc alternativement celui dont on souhaite altérer le comportement, mais aussi celui qui est responsable des effets (au sens des conséquences) de la communication, y compris de son interprétation (au sens sémantique et pragmatique). Les travaux et projets présentés dans ce document explorent deux passerelles comportementales entre le « geste » et la « voix » constituant les deux axes principaux du programme de recherche.

Axe 1. La première passerelle consiste à étudier l'attention du receveur et ses effets sur la communication. Il s'agit notamment d'examiner l'effet des ruptures attentionnelles sur l'usage différentiel des modalités communicatives, et plus généralement la *communication sensible à l'attention* au cours du développement et dans des situations expérimentales choisies. Cet axe a été investigué initialement avec l'étude de la communication gestuelle multimodale du babouin olive dans des situations expérimentales en captivité. Ces travaux réalisés depuis 2011 en postdoctorat et après, sont synthétisés dans le **Chapitre 3**. Ils soulignent notamment

[47] Perlocutoires, illocutoires et locutoires

l'importance des ruptures de l'attention visuelle dans la diversification des modalités communicatives. Sur cette base, j'ai développé l'hypothèse de la communication sensible à l'attention selon laquelle les forces sélectives responsables de l'émergence du contrôle vocal s'ancreraient dans *l'économie de l'attention* entre nourrissons 'dépendants' et parents 'partagés' (avec d'autres enfants et préoccupations). Cette hypothèse fait l'objet d'un projet de recherche en cours (**Projet 1.1.**) et sera présentée au **Chapitre 4**, avec des prédictions testables et les premières études en cours permettant d'étayer les prédictions. Le deuxième projet de recherche (**Projet 1.2.**) permettant d'investiguer le rôle de l'attention visuelle sur la communication sensible à l'attention porte principalement sur l'attention conjointe et ses modalités sensorielles pour explorer le couplage intention – modalité de communication durant les triangulations attentionnelles. Un volet de ce projet sera étendu à l'étude d'un langage émergent visant à étudier le couplage signe – référent selon les modalités visuelle ou auditive[48] et en présence ou en l'absence de contraintes attentionnelles de la part du receveur.

Axe 2. La deuxième passerelle consiste à étudier l'émergence des structures linguistiques avec une approche résolument multimodale et multidimensionnelle[49] qui soit détachée du primat vocal de la linguistique formelle. Il s'agit d'appliquer des formalismes issus de la linguistique à la communication d'un modèle animal, en l'occurrence le babouin olive. Cet axe également développé dans **le Chapitre 4**, comporte deux projets alternativement orientés vers la modélisation phonologique (**Projet 2.1.**) et vers l'analyse sémantique et pragmatique de la communication multimodale (**Projet 2.2.**). Il vise à explorer les contraintes motrices et exécutives d'une part qui contraignent la production de signaux visuels et auditifs, alignés ou chevauchants, mais également les contraintes sociales qui s'exercent sur la production de ces signaux, dans une société naturelle de babouin olive, socialement très complexe (Strum, 2012). Le **Projet 2.1.** examinera les contraintes phonotactiques de séquences de menace, impliquant la main, le corps et des mouvements orofaciaux. Le **Projet 2.2.** permettra d'explorer la flexibilité de la signification des signaux de communication chez les babouins en développant une approche 'lexicale' suivie d'une approche pragmatique mettant en concurrence différentes interprétations possibles en tenant compte des éléments contextuels.

[48] Ce projet de recherche est à l'interface des deux axes puisqu'il consiste à la fois à étudier l'émergence d'une structure linguistique fondamentale (le couplage forme – sens), et à étudier l'effet de la contrainte attentionnelle.

[49] Au sens de Forrester (2008) ; ce qui correspond à la communication multicomposante selon d'autres auteurs (Aychet, Blois-Heulin, & Lemasson, 2021).

L'objectif ultime de ces travaux est de documenter l'émergence du contrôle volontaire et intentionnel des signaux de communication en tenant compte de leur modalité sensorielle, et des règles syntaxiques et sociales gouvernant leur usage, à l'échelle de la microgénèse, de l'ontogénèse et de la phylogenèse. Ce programme de recherche devrait permettre de proposer des forces sélectives plausibles à l'émergence du contrôle vocal permettant aux humains de parler en gesticulant alors que tous les autres primates gesticulent sans parler. Dans cette quête, il permettra d'éclairer un certain nombre de mécanismes aujourd'hui inconnus régissant le couplage intention – modalité de communication et sa sensibilité au contexte attentionnel et aux contraintes sociales. La section suivante présente les définitions opérationnelles utilisées dans les travaux empiriques des Chapitres 3 & 4.

2.3.3. Opérationnalisation des construits

2.3.3.1. La communication intentionnelle

La communication intentionnelle relève de la capacité à utiliser des signaux de communication dans le but de modifier l'état mental ou le comportement du receveur (Bard, 1992; Premack & Woodruff, 1978). Cette définition théorique s'accorde avec la fonction illocutoire de la communication étudiée par Bates et ses collaborateurs (Bates et al., 1975). A la suite de ces études pionnières, les critères utilisés par les chercheurs pour inférer une intention communicative n'ont cessé d'être raffinés et rediscutés (e.g., Leavens, 2004) jusqu'à l'obtention de ce qui ressemble à un paradigme scientifique permettant l'étude systématique de l'usage intentionnel des signaux de communication (Townsend et al., 2017). Dans la lignée de ce consensus, j'étudie la communication intentionnelle selon les critères développés par Townsend et ses collaborateurs (2017) en en croisant parfois les nécessités et les fonctions, mais toujours dans le respect qu'il n'y a intentionnalité que lorsqu'il est possible d'établir un lien entre une action et ses conséquences. En d'autres termes, l'intentionnalité émane de la nature orientée (vers un but) du comportement communicatif. Cette *orientation vers le but* s'observe par (i) la persévérance du signaleur lorsque la première instance de communication échoue ou reste insatisfaisante, ou (ii) une modification ou réponse comportementale de la part du receveur qui soit répétable et satisfaisante pour le signaleur. La *persévérance* peut s'observer par la répétition du même signal ou par l'usage de signaux supplémentaires ou différents du signal initial ; ce qui est dénommé *l'élaboration*. Une *réponse satisfaisante* est obtenue lorsqu'elle induit une cessation de la communication chez le signaleur. Le corollaire de cette

orientation vers le but relève d'un contrôle exécutif de la part du signaleur. En effet, pour que le signal communicatif permette efficacement de modifier le comportement du receveur, il est considéré que le signal doit (i) être orienté vers une audience ou un receveur particulier, (ii) être produit dans une modalité sensorielle qui peut être perçue par le receveur, (iii) être utilisé pour regagner l'attention visuelle d'un receveur visuellement inattentif[50], et (iv) être accompagné de regards vers le receveur et/ou une cible externe (alternance de regards). Les critères (ii) et (iii) relèvent de la communication sensible à l'attention du partenaire et seront développés dans la section 2.3.3.3.. Noter qu'une séquence communicative peut parfois satisfaire à la fois le critère d'élaboration et le critère de regain d'attention.

La nature *ostensive* des signaux communicatifs (qui a été discutée dans la section 2.2.4.2.) peut être opérationnalisée au travers de l'étude des échecs de communication (Moore, 2016) ou des approches pragmatiques de la communication. En effet, s'il s'agit d'établir un lien entre la production d'un signal intentionnel et le travail d'interprétation différentielle de la part du receveur, la dimension coopérative de la communication ostensive semble être satisfaite. A cet égard, l'attention différentielle portée aux signaux communicatifs émis par des jeunes ou des adultes (Bouchet et al., 2017) ou par des individus de haut rang hiérarchique (Chance, 1967; Shepherd et al., 2006) est une piste à considérer. Les travaux présentés dans cette note de synthèse n'ont pas considéré l'ostension comme un critère nécessaire à l'intentionnalité ; en conséquence il s'agit d'une intentionnalité de premier ordre (Dennett, 2009; Graham et al., 2020). Toutefois, la nature ostensive des signaux sera examinée dans l'approche pragmatique de la communication du babouin olive présentée dans le Chapitre 4.

De la même manière, la nature *représentationnelle* des signaux n'étant pas un prérequis de leur usage intentionnel, ce critère n'a pas été retenu pour identifier les signaux de communication étudiés dans les travaux rapportés ici.

2.3.3.2. L'attention sociale

L'attention est un construit de la psychologie référant à l'orientation de l'activité de l'individu vers des stimuli particuliers de l'environnement. L'attention opère à la fois sur la perception sélective (et donc plus efficace) des stimuli que sur la préparation à l'action. L'attention portée

[50] Ce critère fait l'objet de controverses qui seront abordées dans la section 2.3.3.3.. En l'état, considérer qu'il s'agit d'un critère facultatif.

à n'importe quelle caractéristique d'un autre individu peut être considérée comme de *l'attention sociale*, même si les neurosciences cognitives ont tendance à restreindre leur définition du concept d'attention sociale à l'attention portée à la cible du regard d'autrui (Bourjade, 2017). L'attention sociale est étudiée au travers du comportement d'orientation (du corps, du visage ou du regard) vers un stimulus social et peut être de nature réflexe ou volontaire. Les études rapportées dans cette note de synthèse s'appuient sur les deux acceptions du concept d'attention sociale, mais distinguent plusieurs sous catégories qui permettent de lever les ambiguïtés. *L'attention sociale* est donc utilisée pour désigner l'attention portée à un stimulus social quel qu'il soit. Lorsqu'il s'agit de savoir si un individu peut détecter, suivre, exploiter ou comprendre l'orientation du regard d'un autre individu, l*'attention médiée socialement* (Bourjade, 2017) s'appuie sur la détection de la direction du regard d'autrui approximée par l'orientation du regard, de la tête ou des yeux (Emery, 2000). Une fois détecté, le regard d'autrui induit un shift attentionnel chez l'individu focal : A suit la ligne du regard de B. Le *suivi du regard* (*gaze following*) survient quand A suit la ligne du regard de B dans l'espace ; *l'attention conjointe* (*joint attention*) survient quand A porte son attention sur la même entité externe que B ; l'attention partagée (*shared attention*) survient quand A et B en situation d'attention conjointe ont une compréhension mutuelle de leur focalisation partagée (Emery, 2000).

Le *suivi du regard* est une compétence très conservée au cours de l'évolution et concerne tous les primates (Shepherd, 2010). *L'attention conjointe* est également observée mais les comportements peuvent différer entre humains et non-humains. Ici, l'acception large du concept tel que proposé ci-dessus est retenue, mais il est important de noter que (i) il existe des situations fortuites où deux individus regardent dans la même direction sans que ce soit de l'attention conjointe (ex : un bruit fort survient et fait converger les regards), (ii) de nombreux auteurs incluent dans l'attention conjointe chez l'enfant les gestes de pointage quand d'autres considèrent la triangulation des regards comme suffisantes (Bard et al., 2014; Beuker et al., 2013; Leavens et al., 2015), (iii) aucune étude à ma connaissance n'a à ce jour distingué les situations d'attention conjointe uniquement visuelle de celles qui implique la parole (alors que la référence représentationnelle peut être interprétée comme une attention conjointe auditive vers une entité abstraite[51], (iv) la distinction entre l'initiation de l'attention conjointe et la réponse à l'initiation d'autrui permet de prédire l'entrée dans le langage (Cochet & Byrne,

[51] Je remercie à cet égard Kim Bard pour des échanges fructueux sur ces questions d'attention conjointe, d'orientation sociale et de perceptions sensorielles chez les personnes atteintes de déficiences visuelles ou auditives.

2016). L'attention conjointe rapportée dans mes travaux passés, en cours et futurs, repose à minima sur (i) *l'alternance des regards* entre le partenaire social et une cible externe, et/ou (ii) la focalisation sur la même cible externe en excluant les perturbations environnementales. *L'attention partagée* n'est pas distinguable de l'attention conjointe sur le plan comportemental et n'est donc pas retenue dans ces études (Bourjade, 2017).

L'étude des ruptures attentionnelles développée dans le Chapitre 4 opérationnalise l'attention que les parents portent à leurs enfants de deux manières. L'attention *perceptive* relève de ce que le parent peut voir (attention visuelle), entendre (attention auditive) ou sentir (attention tactile). Les ruptures attentionnelles sont considérées modalité par modalité, excepté pour l'attention auditive (non contrôlée dans ces études). L'attention *active* réfère aux comportements parentaux comme le portage (aux bras, sur les genoux), les soins tactiles (caresses, épouillage), la communication adressée à l'enfant, le face-à-face ou encore l'attention conjointe. L'attention *détournée* (*diverted*) réfère à la focalisation de l'attention parentale sur d'autres stimuli que l'enfant, comme un autre individu ou un écran de téléphone ou de télévision. L'*inattention* survient lorsqu'aucune forme d'attention n'est présente. Toutefois, l'attention auditive parentale ne pouvant être contrôlée, l'inattention complète n'est jamais attestée. Finalement la situation « hors de vue » qui survient lorsque le parent humain change de pièce ou que la mère babouin ou chimpanzé est hors de vue de son enfant est la forme d'inattention la plus complète à laquelle ces études ont accès.

2.3.3.3. La sensibilité à l'attention d'autrui

L'étude de la sensibilité à l'attention d'autrui proposée dans cette note de synthèse réfère exclusivement aux situations communicationnelles où le *signaleur* est sensible (ou pas) à l'attention du *receveur*. La communication sensible à l'attention (du receveur) s'opérationnalise en termes de *concordance modale* entre les signaux communicatifs émis et l'attention du receveur. Le signal concorde avec l'attention du receveur quand ce dernier peut le percevoir. Le signal ne concorde pas avec l'attention du receveur si ce dernier ne peut le percevoir (ex : un geste orienté vers un partenaire visuellement inattentif). Les individus présentent une communication *sensible à l'attention* lorsque la non-concordance est absente ou en quantité significativement inférieure à la concordance modale (Dafreville et al., 2021). La communication strictement visuelle est la seule qui requiert une attention visuelle préalable à l'émission du signal pour fonctionner. Les études présentées dans le Chapitre 3 ont investigué

la discrimination de l'attention visuelle d'un receveur humain par des babouins adultes en situation expérimentale. Le projet 1.1. portant sur l'effet des ruptures attentionnelles sur les modalités de communication repose également sur la discrimination de l'attention visuelle, de même que tactile.

L'attention visuelle peut être manipulée, c'est-à-dire modifiée, par l'émission d'un signal dans une autre modalité. Initialement, les gestes à composantes auditives et tactiles étaient pensées comme des *attracteurs attentionnels* capables de manipuler l'attention visuelle du receveur (Tomasello et al., 1994). Cependant, la question de la *manipulation de l'attention* fait débat pour différentes raisons : (i) parce qu'aucune étude n'est parvenue à mettre en évidence une séquence « geste tactile/audible suivi de geste visuel » qui pourrait correspondre à la définition initiale (Genty et al., 2009; Hobaiter & Byrne, 2011b; Liebal, Call, & Tomasello, 2004), (ii) une étude a montré que les chimpanzés préféraient se repositionner à la vue d'un receveur initialement inattentif (Liebal, Call, Tomasello, et al., 2004), et a été répliquée chez les mangabeys et les macaques à Bonnet (Aychet et al., 2020; Deshpande et al., 2018), (iii) un attracteur attentionnel ne devrait servir qu'à attirer l'attention sans véhiculer d'information supplémentaire (Hobaiter & Byrne, 2011b), et (iv) les signaux à composante tactile ou auditive émis envers un individu visuellement inattentif n'entrainent pas la modification de son attention visuelle dans la seule étude qui l'a examiné (Dafreville et al., 2021). A ce jour, il n'existe pas d'évidence de *manipulation* de l'attention per se, bien que cet aspect n'ait pas été examiné de manière systématique. En revanche, les exemples de communication sensible à l'attention sont nombreux et seront détaillés dans les chapitres suivants.

Récemment, nous avons opérationnalisé deux descripteurs de la communication sensible à l'attention (Dafreville et al., 2021). *L'ajustement unimodal* réfère à la capacité du signaleur à éviter les situations de non-concordance modale, c'est-à-dire à produire aucun ou moins de signaux dans une modalité donnée quand celle-ci n'est pas perçue par le receveur que lorsqu'elle l'est (ex : moins de signaux strictement visuels vers un receveur visuellement inattentif qu'attentif). *L'ajustement intermodal* réfère à la capacité de l'individu à favoriser les situations de concordance modale, c'est-à-dire à déployer des signaux d'une modalité qui est perçue par le receveur en évitant la modalité non perçue (ex : produire plus de signaux audibles que visuels quand le receveur est visuellement inattentif).

Noter que de nombreuses études sur le sujet de la sensibilité à l'attention présupposent une *attribution d'attention* au sens de l'attribution d'un état mental (attentionnel) à autrui. Or, à ce jour, nous ne savons pas si les primates non humains perçoivent l'attention visuelle d'autrui

comme (i) un ensemble d'indices visuels associés à des contextes particuliers, (ii) une focalisation, un but ou un à-propos de l'attention d'autrui, (iii) une expérience subjective de *voir*, qui peut être à l'origine d'une connaissance sur le monde, ou (iv) l'intention d'un autre individu de se focaliser sur une partie de l'environnement (Bourjade, 2017). Face à cet état de fait équivoque, certains chercheurs adoptent une posture mentaliste par défaut, comme par exemple Frans De Waal (1999) avec son concept d'anthropodéni, quand d'autres priorisent une posture non-mentaliste par défaut comme David Leavens (2017) avec sa présentation de la mal-mesure de l'intelligence des grands singes. Dans mes travaux, j'adopte prioritairement une posture non-mentaliste comme étant la plus parcimonieuse en vertu d'une décision sous incertitude (Sober, 2005), que j'articule parfois avec des construits qui 'capturent' une partie des phénomènes non explicables autrement.

2.4. Les apports d'une approche comparative

2.4.1. Des inférences évolutives à partir des comparaisons entre espèces ?

L'évolution des compétences cognitives et comportementales est massivement étudiée au travers de l'approche comparative entre espèces (et entre cultures, voir ci-dessous) du fait que ces 'traits' ne fossilisent pas (Apicella & Barrett, 2016). Jusqu'à ce jour, la psychologie comparée a utilisé la classification des relations de parentés entre espèces établies sur la base de caractères biologiques (e.g., morphologiques, moléculaires) pour inférer des scénarios évolutifs de continuité ou de discontinuité. A ce titre, la psychologie comparée se comporte comme une 'utilisatrice' des arbres phylogénétiques produits par les biologistes de l'évolution. Il n'est cependant pas inintéressant de considérer les traits comportementaux et cognitifs comme des marqueurs de l'évolution susceptibles d'être hérités d'ancêtres communs, i.e., comme des *homologies* (Cap, 2015) – la réticence vient toutefois du caractère non matériel de ces traits et de leur héritabilité extra-génétique dont on méconnait encore la stabilité au fil du temps (Danchin & Pocheville, 2014). L'approche 'utilisatrice' est celle adoptée dans les travaux rapportés dans cette note de synthèse ; toutefois l'approche de phylogénie comportementale sera abordée dans les perspectives en fin de Chapitre 4.

Traditionnellement, la psychologie comparée s'est focalisée sur le chimpanzé du fait de sa proximité phylogénétique avec l'humain. En conséquence, la très grande majorité des études et des connaissances générées dans ce domaine porte sur cette espèce. De manière quelque peu hâtive, les ressemblances et les différences observées entre le chimpanzé et l'humain sont

alternativement considérées comme des évidences de continuité ou de discontinuité, respectivement. Toutefois, si les deux espèces diffèrent pour un même trait comportemental, il semble hâtif d'inférer le scénario de discontinuité sans que l'on sache s'il s'agit d'un caractère apparu uniquement dans la lignée humaine (i.e., une autapomorphie) ou d'un caractère disparu uniquement chez le chimpanzé (i.e., une réversion de caractère). Ne pas inférer le scénario de discontinuité n'implique pas que l'on accepte le scénario de continuité pour autant[52]. En réalité, il est impossible de départager les deux scénarios dans la majorité des cas. Le deuxième cas de figure le plus commun est l'observation de ressemblances chez deux espèces phylogénétiquement éloignées qui sont interprétées comme des cas de *convergence évolutive*. Là également, si certains cas sont évidents, comme l'imitation vocale chez les oiseaux chanteurs et les humains par exemple, d'autres le sont beaucoup moins du fait de la méconnaissance d'un trait donné chez les différentes espèces du clade considéré.

Par exemple, l'observation d'un trait semblable chez le babouin du genre *Papio spp*. et chez l'humain ne permet pas d'inférer un scénario de convergence évolutive ni d'héritage d'un ancêtre commun tant que la grande majorité des genres du clade des catarrhiniens n'est pas échantillonnée, soit 30 genres d'après la phylogénie moléculaire de Perelman et al. (2011) (voir Figure 1). De plus, ce raisonnement s'appuie sur le modèle de *parcimonie* selon lequel l'évolution *minimise* le nombre de changements requis (somme des apparitions et disparitions). Or, le modèle de parcimonie n'est qu'un modèle (logique) parmi d'autres (statistiques). Aujourd'hui les biologistes de l'évolution travaillent avec des modèles statistiques complexes tenant compte de la probabilité d'apparition et de disparition des caractères, qui n'est pas nécessairement équiprobable et peut varier selon que le caractère est un trait discret (qualitatif, ex : le contrôle du larynx au niveau cortical) ou continu (quantitatif, ex : la taille du cerveau) (Harmon, 2018). Pour toutes ces raisons, les travaux rapportés ici restent précautionneux quant à la possibilité de faire des inférences évolutives *directes* sur la base des résultats rapportés. L'approche défendue dans ce travail est de documenter l'état (ou la quantité) d'un trait donné chez les espèces étudiées afin de pouvoir s'appuyer sur cette connaissance ultérieurement pour

[52] Une part importante des travaux de psychologie comparée cherche à caractériser l'unicité de l'être humain (*human uniqueness*) en inférant une discontinuité au niveau phylogénétique qui s'appuie sur une « rupture » dans les caractères observés chez l'humain et les non humaines (e.g., l'intentionnalité partagée de Tomasello). Certains primatologues considèrent à l'inverse que la continuité (graduelle au sens darwinien) est le modèle de raisonnement le plus parcimonieux (e.g., Frans de Waal).

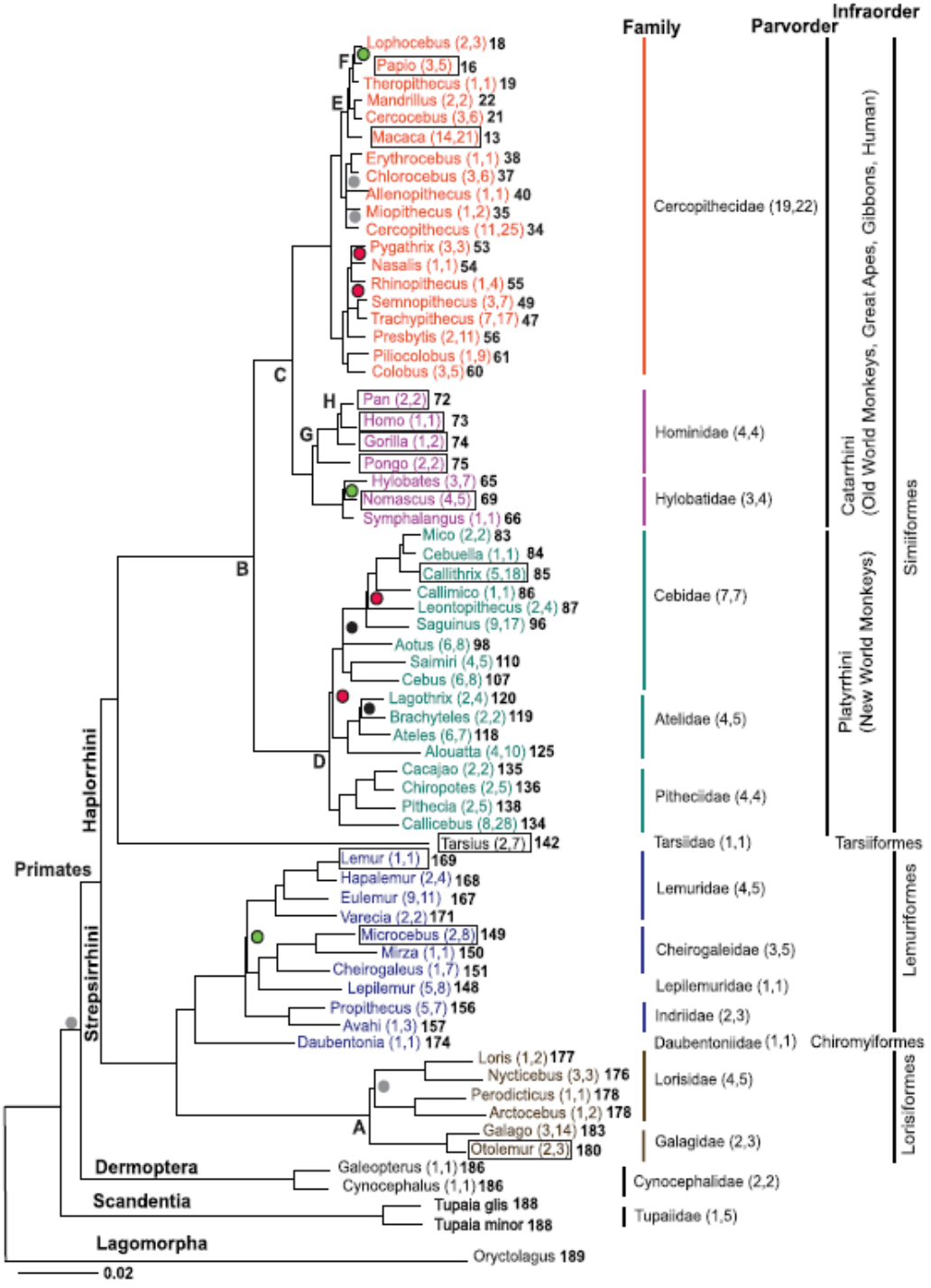

Figure 1. Phylogénie moléculaire de 61 genres de primates. Les dermoptères et les scandentiens sont des petits mammifères arboricoles considérés apparentés aux primates. Les lagomorphes (e.g., lapins, lièvres) sont utilisés comme groupe extérieur. Les encadrés indiquent les genres pour lesquels les génomes ont été séquencés. Les nombres entre parenthèses à côté de chaque genre représentent le nombre d'espèces utilisées dans l'étude, suivis du nombre total d'espèces décrites pour ce genre. Les nombres entre parenthèses à côté des familles représentent le nombre de genres utilisés dans l'étude, suivis du nombre total de genre décrits. Les lettres majuscules représentent l'émergence de chaque phylum indexée sur les données fossiles : A) Séparation Galagidae-Lorisidae 38–42 MA, B) Emergence des Simiiformes 36–50 MA, C) Emergence des Catarrhini 20–38 MA, D) Emergence des Platyrrhini 20–27 MA, E) Emergence de la tribu Papionini 6–8 MA, F) Emergence des Theropithecus 3.5–4.5 MA, G) Emergence de la famille Hominidae 13–18 MA, H) Séparation Homo-Pan 6–7 MA. **Adapté de Perelman et al. (2011).**

renseigner les modèles théoriques et computationnels d'évolution (voir section 4.3.1. pour un développement ultérieur).

Les travaux qui font l'objet de cette note de synthèse s'appuient essentiellement sur les modèles babouin, chimpanzé et humain. Le babouin olive (*Papio anubis*) est une espèce subsaharienne de la famille des cercopithécidés apparue au miocène ancien il y a environ 20 MA, et dont la divergence avec la lignée hominoïde remonte elle, à environ 35 MA - c'est-à-dire que les plus anciens fossiles connus d'ancêtres des cercopithécoïdes et des hominoïdes datent de 34 MA ; (Lecointre & Le Guyader, 2001). Le chimpanzé commun (*Pan troglodytes*) est une espèce subsaharienne de la famille des hominidés qui émerge il y a 13 MA (Lecointre & Le Guyader, 2001). On ne connait pas de chimpanzé fossile mais la divergence avec la lignée des hominines est estimée à 6-7 MA (Perelman et al., 2011). Les deux espèces diffèrent sur de nombreux points qui présentent des intérêts comparatifs.

Les babouins vivent dans des milieux de savane semi-arborée relativement ouverts et qui s'apparentent au paléoenvironnement des premiers humains - la mosaïque de savane arborée date d'environ 15.5 MA au Kenya (Cerling et al., 2011), et qui présentent des avantages pour investiguer l'effet des ruptures d'attention visuelle sur la communication. A l'inverse, les chimpanzés vivent dans des milieux forestiers denses qui ne présentent pas les mêmes contraintes sur les moyens de communication. En effet, le son se propage moins loin dans les milieux fermés qu'ouverts et la visibilité au sol est beaucoup moins bonne, contraignant ainsi la communication à la fois visuelle et acoustique (Hardt & Benedict, 2021). Les babouins forment également des sociétés complexes multi-niveaux structurées par des liens d'affinités très forts entre femelles adultes. Les femelles sont philopatriques[53] tandis que les mâles dispersent à maturité sexuelle (Smuts et al., 2008). Les interactions envers les enfants sont extrêmement fréquentes, de même que les comportements allo-parentaux et l'établissement de liens affiliatifs entre des mâles adultes et de jeunes enfants (Strum, 2012). Au contraire, les chimpanzés forment des sociétés de fission-fusion avec des mouvements réguliers mais imprévisibles d'individus dans des sous-groupes ; les sous-parties sont majoritairement constituées d'une femelle adulte et de sa progéniture et les opportunités sociales sont bien moins fréquentes que chez les babouins (Nishida et al., 1999). Les mâles, et non les femelles, sont philopatriques et il est possible que cela influence l'investissement maternel (Pusey, 1983). Sur

[53] Qualifie les individus qui restent dans leur groupe natal pour s'y reproduire ; généralement un des deux sexes, tandis que l'autre disperse vers d'autres groupes pour se reproduire.

la base de ces différences, la prévalence de la communication gestuelle sur la communication vocale qui est communément rapportée chez les chimpanzés pourrait s'avérer être un caractère dérivé chez cette espèce plutôt qu'un héritage ancestral à tous les hominidés et encore moins à tous les PNH. En d'autres termes, étudier uniquement les compétences communicatives des chimpanzés en regard des compétences communicatives humaines ne permet pas d'établir un quelconque scénario évolutif ; un groupe extérieur est nécessaire pour raisonner sur les inférences évolutives.

En captivité, les chimpanzés et les babouins diffèrent également sur de nombreux points qui relèvent du mode d'hébergement et de ses conséquences. Les babouins olive de la station de primatologie de Rousset (CNRS) sont hébergés dans de larges groupes et n'ont que peu d'interactions avec les soigneurs humains. A l'opposé, les chimpanzés captifs sont souvent maintenus isolés ou en très petits groupes et sont plus choyés par les soigneurs humains, bien que cela dépende des conditions d'hébergement (Bard et al., 2014; Bard & Leavens, 2014). Dans l'ensemble, les babouins captifs ont moins d'interaction avec les humains que les chimpanzés captifs, ce qui permet de contrôler leurs apprentissages et leur « histoire expérimentale » ; différence qu'il est possible de capitaliser pour étudier l'effet de l'expérience individuelle sur la communication sensible à l'attention (Cf. Chapitre 3). L'ensemble de ces raisons rend plus que féconde la dimension comparative entre ces trois espèces dans ce programme de recherche.

2.4.2. La nécessaire prise en compte de l'évolution culturelle chez l'humain

L'évolution a modelé les comportements dans des milieux écologiques donnés de sorte que les populations que l'on observe sont généralement adaptées à leur environnement[54] (Nettle, 2009). Chez l'humain, la transmission culturelle est un facteur additionnel d'adaptation qui modèle les comportements et peut générer des phénomènes de microévolution (biologique). La transmission culturelle opère verticalement, mais aussi horizontalement et obliquement de sorte qu'elle peut générer une variation très importante en un temps beaucoup plus court que l'évolution biologique (Micheletti, 2020). En conséquence, savoir si la variation observée entre les populations humaines résulte de l'adaptation aux conditions locales (microévolution) ou de

[54] A condition que la fréquence des changements environnementaux n'excède pas le temps requis à la sélection naturelle pour modeler des adaptations. Ce raisonnement peut s'appliquer à l'évolution culturelle et sociétale également (Rosa, 2019).

l'évolution culturelle ou les deux est une question empirique qu'il convient d'examiner (Nettle, 2009). Du point de vue théorique, les modèles mathématiques d'évolution culturelle sont conçus dans la lignée des modèles d'évolution biologique, ce qui génère des controverses, mais aussi des avancées significatives (Micheletti, 2020; Nettle, 2020; Thornhill & Fincher, 2013). Du point de vue empirique toutefois, l'étendue de la variation inter-populations au sein de l'espèce humaine n'a fait qu'être effleurée en surface malgré les mises en garde répétées à propos du biais d'échantillonnage vers les populations occidentales, dénommées WEIRD[55] (*Western, Educated, Industrialized, Rich and Democratic*) par Henrich et al. (2010b, 2010a) alors qu'elles ne représentent que 12% de la population mondiale (Nielsen et al., 2017). Or, le peu d'études disponibles met en évidence des différences entre populations humaines concernant le développement de l'enfant et la manière dont les parents interagissent avec leurs enfants (Keller et al., 2004, 2009; Little et al., 2016; Nielsen et al., 2017). Ceci pose des problèmes de validité scientifique concernant la généralisation des résultats obtenus sur les populations WEIRDs à l'ensemble de l'humanité. Cela pose également des problèmes de validité scientifique à toute science comparative impliquant des données issues de populations humaines. Dans le cadre des travaux rapportés dans cette note de synthèse, comparer humains et non humains sans savoir estimer la variation humaine pour un trait donné ne paraissait pas conclusif.

Sur cette base, l'Axe 1. visant à étudier l'effet des ruptures attentionnelles maternelles sur le développement de la communication de l'enfant a été entrepris dans une démarche de comparaison entre espèces, mais également entre populations humaines (Cf. Projet 1.1., section 4.1.2.). Le comportement parental humain varie selon les cultures et les environnements (Keller, 2008). Les parents possèdent des conceptions différentes de ce qu'ils souhaitent valoriser dans le développement (Carra et al., 2013; Keller et al., 2005) constituant des modèles parentaux différents, sous influence culturelle (Keller et al., 2006, 2009). Or, ces modèles parentaux sont susceptibles de modifier les interactions entre parents et enfants (Bader et al., 2019; Bader & Fouts, 2018; Carra et al., 2013) et ainsi, d'affecter l'étayage du développement langagier que ces interactions procurent (Bruner, 1975; Vygotsky, 1978). Par exemple, les mères occidentales s'engagent dans plus d'interactions triadiques impliquant des objets que les mères sub-sahariennes (Bornstein et al., 2017; Keller, 2008), tandis que les interactions distales sont moins communes et le contact physique plus fréquent chez ces dernières (Keller et al.,

[55] Ce qui signifie également bizarre, étrange ou encore curieux en anglais.

2005). D'autre part, les adultes n'adressent pas la même quantité de discours aux enfants selon les populations : les adultes parlent très peu aux enfants tsimane de moins de trois ans, de même qu'aux enfants mayas ou papous (Casillas et al., 2020, 2021; Cristia et al., 2019). Du point de vue de l'enfant, des différences entre populations ont été mises en évidence concernant la fréquence des gestes de pointage, la quantité ou la multimodalité des séquences de gestes de pointage (Callaghan et al., 2011; Salomo & Liszkowski, 2013) ou encore les fréquences d'interactions triadiques et leur valeur prédictive de l'acquisition ultérieure de vocabulaire (Mastin & Vogt, 2011, 2016). Il est intéressant de noter que ces deux dernières études ont été réalisées au Mozambique dans deux environnements différents ; l'un rural et l'autre urbain. La prise en compte des influences environnementales permet de tenir le *facteur culturel*[56] constant dans les comparaisons entre populations et ainsi, d'évaluer son influence (Apicella & Barrett, 2016). Mastin & Vogt (2016) ont par exemple montré que les prédicteurs du développement langagier chez les enfants mozambicains variaient selon l'environnement rural ou urbain et qu'à plusieurs égards, les enfants mozambicains urbains étaient plus semblables aux enfants néerlandais (urbains) qu'aux enfants mozambicains ruraux.

[56] Bien qu'il ne soit jamais défini de manière conclusive. Dans le cas présent, la culture pouvait être approximée par le fait qu'il s'agisse de deux populations parlant la même langue et habitant le même pays.

Chapitre 3. La communication gestuelle multimodale

Dans ce chapitre, je présente le contexte théorique et expérimental qui a donné lieu à une série d'expériences visant à tester la nature intentionnelle des gestes communicatifs chez le babouin olive en captivité. Ces études postdoctorales ont révélé la nature multimodale et multidimensionnelle de la communication gestuelle du babouin olive, de même que l'importance de l'accès au visage pour discriminer les états attentionnels, développer des comportements d'attention conjointe et acquérir une communication sensible à l'attention du receveur. Cette référence (*baseline*) expérimentale a ensuite été utilisée comme grille de lecture pour interpréter la communication intraspécifique, en captivité (ci-dessous et Projet 2.1.) et dans la nature (Projets 1.1. et 2.2.).

3.1. Approche expérimentale de la communication intentionnelle

3.1.1. Le paradigme de requête alimentaire

Le paradigme de requête alimentaire est un dispositif de communication interspécifique qui consiste à présenter un item alimentaire attractif à un sujet tout en le maintenant hors de portée. Dans le cadre des expériences sur les primates non humains, l'expérimentateur tient le plus souvent la récompense alimentaire dans la main - bien que d'autres systèmes aient été testés (Tempelmann et al., 2011) et se tient face au singe, à environ 1 mètre de distance face à la cage. Les expériences sont généralement constituées de plusieurs essais contenant des essais 'test' présentés au milieu d'essais dits 'de motivation'. Durant les essais de motivation, la nourriture est présentée au sujet et elle lui est donnée s'il effectue un geste de requête consistant, chez les babouins étudiés ici, à étendre un bras en dehors de la cage (Figure 2). Selon les espèces et les situations, la requête peut prendre la forme d'un pointage, d'une quémande ou encore être effectuée avec le doigt, la main, le bras, les lèvres ou encore les yeux (singes: Bourjade, 2016; chiens: Gaunet & Deputte, 2011; cochons: Nawroth et al., 2013, chèvres: 2015; chevaux: Proops & McComb, 2010). Les essais 'test' sont utilisés pour tester un certain nombre de prédictions sur la nature communicative et intentionnelle des gestes de requête alimentaire.

Les recherches menées au *Yerkes National Primate Research Center* sur des chimpanzés captifs ont sans doute constitué la base de ces investigations. Leavens et al. (1996) ont initialement rapporté que les gestes de pointage de trois chimpanzés adultes étaient associés à des

vocalisations et des alternances de regards entre le visage de l'expérimentateur et la nourriture. Dans une étude plus complète menée sur 115 chimpanzés, Leavens et Hopkins (1998) rapportèrent que la moitié des chimpanzés produisaient des gestes de requête face à la nourriture, tandis que 27 d'entre eux produisaient des gestes et des vocalisations, et 8 d'entre eux seulement des vocalisations. Plus spécifiquement, une étude ultérieure montra que les premiers signaux étaient majoritairement des gestes ou des gestes accompagnés de vocalisations, mais que si l'expérimentateur donnait la banane à un individu tiers, les sujets vocalisaient (Leavens et al., 2004). Ces études ont permis d'éclairer la nature multimodale de la communication entre humain et chimpanzé dans une situation de requête alimentaire. Toutefois, elles n'indiquaient pas précisément si les sujets réagissaient à l'attention visuelle de l'expérimentateur ou à son attention détournée socialement, ou encore à la frustration.

Figure 2. Gestes de requête alimentaire unimanuel par un mâle (a) et bimanuel par une femelle (b).

De nombreuses études ont investigué la capacité des chimpanzés et plus généralement des grands singes à ajuster leur comportement de requête alimentaire à l'état attentionnel de l'expérimentateur humain ; ce que nous avons appelé la *communication sensible à l'attention*. Typiquement, les essais 'test' contenaient des essais où l'expérimentateur ne pouvait pas voir le singe, e.g., avec les yeux fermés ou un seau sur la tête (Kaminski, Call, & Tomasello, 2004; Povinelli et al., 1996), et/ou des essais qui variaient les indices d'attention disponibles, e.g., orientation du corps, de la tête, du regard (Hostetter et al., 2007; Kaminski, Call, & Tomasello, 2004). Plusieurs études ont ainsi investigué quels indices attentionnels étaient utilisés par les sujets pour ajuster leur communication, et ont mené à des résultats très mixés. La majorité des

études montrent que les singes réalisent un ajustement *unimodal* en produisant moins de requêtes alimentaires quand l'expérimentateur n'est pas attentif ; les grands singes utilisent l'orientation du corps (Hostetter et al., 2001; Kaminski, Call, & Tomasello, 2004; Povinelli et al., 1996) ou du visage (Tempelmann et al., 2011), mais rarement l'état des yeux (Kaminski, Call, & Tomasello, 2004; Povinelli et al., 1996; Theall & Povinelli, 1999) pour discriminer l'état attentionnel de l'expérimentateur. Les petits singes n'étaient quasiment pas étudiés à l'époque mais quelques études mentionnaient des résultats similaires (Hattori et al., 2010; Maille et al., 2012; Meunier, Prieur, et al., 2013), qui ont été largement répliqués depuis (e.g., Aychet et al., 2020; Canteloup et al., 2015; Deshpande et al., 2018). De plus, deux études rapportaient un ajustement *intermodal* réalisé par des chimpanzés qui favorisaient la modalité acoustique quand l'expérimentateur n'étaient pas visuellement attentif (Hostetter et al., 2007; Leavens et al., 2010), sur la base de l'aperture de l'œil (ouvert ou fermé) (Hostetter et al., 2007). C'est dans ce contexte, que nous avons étudié la capacité du babouin olive à ajuster ses requêtes alimentaires à l'état attentionnel d'un expérimentateur humain.

3.1.2. Sensibilité à l'attention d'autrui

La première étude que nous avons menée consistait explicitement à tester la sensibilité à l'attention du partenaire chez des babouins olive captifs (Bourjade, Meguerditchian, et al., 2014). L'étude s'appuyait sur 16 individus adultes initialement entrainés à produire un geste de requête alimentaire par un expérimentateur humain. Les sujets étaient ensuite exposés à 4 conditions expérimentales mélangées à des essais de motivation dans un ordre contrôlé suivant un carré latin. Dans la condition Yeux Ouverts (YO), l'expérimentateur se tenait face à la cage les yeux ouverts en regardant le sujet sans le fixer du regard ; dans la condition Yeux Fermés (YF), l'expérimentateur se tenait face à la cage, le morceau de banane à la main, les yeux fermés ; dans la condition Dos Tourné (DT), l'expérimentateur se tenait dos à la cage avec le morceau de banane tenu dans le dos de sorte qu'il soit visible du sujet ; dans la condition Dehors (D), le morceau de banane était déposé devant la cage et l'expérimentateur quittait la zone de test. Chaque essai 'test' durait 30 secondes durant lesquelles l'expérimentateur maintenait la posture, puis se terminait en donnant le morceau de banane au sujet indépendamment de son comportement durant l'essai.

Les babouins ont massivement produit des gestes de requête alimentaire lorsque l'expérimentateur était attentif, les yeux ouverts. A cet ajustement *unimodal*, ils ont associé des

alternances de regard entre la banane et le visage de l'expérimentateur, plus fréquemment quand elle avait les yeux ouverts que dans toutes les autres conditions. Les babouins n'ont pas produit de requête alimentaire quand l'expérimentateur n'était pas présent (condition D) et ils ont répété leurs requêtes jusqu'à l'obtention de la récompense, montrant ainsi une persévérance. De plus, ils ont présenté un ajustement *intermodal* en produisant des gestes sonores (taper sur la cage ou sur un plexiglas) spécifiquement lorsque l'expérimentateur ne pouvait pas les voir (conditions DT et YF).

Ces résultats ont intrigué les experts relecteurs de l'article, de même que les éditeurs[57] car la plupart des études sur les chimpanzés échouaient à montrer un ajustement de la communication à l'attention de l'humain sur la base de l'aperture de l'œil ; en d'autres termes sur la base de son attention visuelle *effective*. Ces résultats négatifs avaient donné lieu à une controverse sur la nature de la *compréhension* que les chimpanzés ont de la vision : reconnaissent-ils des indices auxquels ils s'ajustent ou comprennent-ils qu'avoir les yeux ouverts permet de *voir* et donc de rassembler des connaissances sur l'état du monde ? Par ailleurs, les babouins avaient-ils ces mêmes compétences ? Ils semblaient faire un usage tactique de leurs modalités de communication qui n'avait pas été démontré par un grand nombre d'études également. Sur la base des critiques reçues, j'ai exploré deux développements de cette première étude. Premièrement, j'ai cherché à savoir comment ces babouins étaient parvenus à discriminer si finement l'état attentionnel de l'expérimentateur. Deuxièmement, j'ai cherché à examiner l'ajustement intermodal et l'usage de la modalité acoustique dans d'autres situations. Je me suis appuyée dans les deux cas sur la manipulation de l'expérience individuelle dont l'approche est présentée ci-dessous.

3.2. Effet de l'expérience individuelle sur l'acquisition de l'intentionnalité

Comme cela a été introduit plus haut, les singes anthropoïdes et les petits singes ne sont souvent pas hébergés en captivité dans des conditions similaires. Une des conséquences de l'exposition des grands singes aux interactions avec les humains est qu'ils apprennent spontanément à pointer, quémander, voire échanger des objets contre de la nourriture (observation personnelle

[57] Baboons know when to be noisy. Nature: Research Highlights, 504, 190. https://doi.org/10.1038/504190a

au Wolfgang Köhler Primate Research Center, Leipzig), tandis que les petits singes n'en ont pas l'opportunité. Par conséquent, les requêtes alimentaires des petits singes tendent à être considérées comme des *conditionnements* tandis que celles des grands singes seraient des instances de communication intentionnelle, ce qui vient implicitement tirer une ligne de partage dans l'intelligence que l'on attribue à ces espèces (Bourjade, 2016; Bourjade et al., 2020). Dans les études que j'ai menées pour quantifier l'effet de l'expérience individuelle sur le comportement communicatif, j'ai tiré parti de deux spécificités initialement conçues comme des 'faiblesses' : (1) si les petits singes ne développent pas spontanément les gestes de requêtes alimentaires mais qu'il est possible de les y entrainer, alors il est aussi possible de *contrôler* et même de *manipuler* l'entraînement ; et (2) s'il est impossible de contrôler les différents déterminismes dans le développement de l'enfant, et semble-t-il dans celui du chimpanzé, l'utilisation d'un *modèle adulte naïf* que l'on entrainerait devrait permettre de séparer l'effet de l'expérience individuelle, de celui de la maturation ou des influences socio-culturelles usuellement en synergie dans le développement des jeunes individus. Trois études ont reposé sur cette logique et sont présentées ci-dessous.

3.2.1. Coordination des gestes et des regards

Dans la première étude, nous avons cherché à obtenir une base de référence des compétences gestuelles des babouins captifs (Bourjade et al., 2015). Nous avons impliqué 9 nouveaux sujets (3 mâles et 6 femelles) qui n'avaient pas été entrainés à produire le geste de requête et nous les avons directement exposés à une série de 4 essais 'test' effectués dans la condition Yeux Ouverts (YO). Ces sujets témoins ont ensuite été comparés aux sujets entrainés de Bourjade et al. (2014) pour la quantité de gestes produits, la fréquence des alternances de regards et la coordination entre les deux.

Nous avons montré que 5 des 9 sujets naïfs produisaient spontanément des gestes de quémande (entre 1 et 12 gestes observés durant les 4 essais) sans avoir été entrainés à les produire. Toutefois, les sujets entrainés en produisaient dix fois plus. Les deux groupes présentaient une quantité comparable d'alternances de regards entre le visage de l'expérimentateur et le morceau de banane. En revanche, les individus naïfs passaient plus de temps à regarder le visage de l'expérimentateur que la banane et inversement pour les sujets entrainés. Seuls les babouins entrainés coordonnaient temporellement leurs alternances de regards et leurs gestes, suggérant que l'entrainement avait permis d'établir cette association. Ces résultats sont intéressants à

plusieurs égards. Premièrement parce que les babouins 'naïfs' présentaient spontanément les deux comportements clefs, bien que de manière non coordonnée, et deuxièmement parce que l'entrainement a permis aux sujets qui en ont bénéficié de passer d'une attention dyadique (regard vers le visage de l'expérimentateur) à une attention triadique (alternance entre le visage de l'expérimentateur et la banane) ; une étape développementale clef du développement sociocognitif chez l'enfant (Bakeman & Adamson, 1984; Beuker et al., 2013; Carpenter & Call, 2013). De surcroit, ces sujets ont coordonné leurs alternances de regards avec leurs gestes, ce que les enfants humains développent progressivement au cours de la deuxième année de vie (Franco & Butterworth, 1996). Ces résultats ont motivé l'étude suivante qui s'est intéressée précisément à l'acquisition de ces comportements d'attention conjointe au cours des phases d'entraînement.

3.2.2. Acquisition des comportements d'attention conjointe

Dans les deux études qui ont suivi nous avons manipulé l'expérience des babouins au cours de la phase d'entrainement au geste de requête alimentaire (Bourjade, 2016). Un deuxième groupe de 12 babouins adultes naïfs a été entrainé à produire le geste de requête alimentaire dans des conditions différentes des premiers. Les 16 premiers individus avaient été entrainés par un humain de face, les regardant, ci-après le groupe *Face*. Les 12 babouins supplémentaires ont été entrainés par un expérimentateur se tenant de profil, aidé par un complice invisible du sujet, qui donnait un signal quand le sujet tendait le bras/produisait une requête alimentaire, ci-après le groupe *Profil*.

Sur la base des résultats précédents, nous supposions que les babouins développaient la coordination geste-regard pendant l'entraînement. De plus, les babouins du groupe *Profil* ne recevant pas les mêmes stimulations sociales que les babouins du groupe *Face*, et notamment l'accès au regard de l'humain, nous avons testé l'hypothèse selon laquelle l'accès aux indices d'attention humains favorables à la communication (visage de face, yeux ouverts, regard dirigé) était nécessaire pour que cette coordination apparaisse. Nous avons d'abord étudié l'émergence des gestes et de la coordination avec les alternances de regards chez 11 sujets, six d'entre eux appartenant au groupe *Face*, les cinq autres au groupe *Profil* (Lamaury et al., 2019). Nous avons ainsi testé l'effet de l'état attentionnel de l'expérimentateur sur l'acquisition des gestes, l'émergence des alternances de regards et la coordination des deux. Les babouins des deux groupes ont développé les gestes de requêtes. Ils produisaient également autant d'alternances

de regards au début de l'entrainement. Seulement, à partir de la deuxième étape de l'entraînement, c'est-à-dire quand la récompense devient hors de portée et que les gestes de préhension commencent à se ritualiser en gestes de « requête », les babouins du groupe *Profil* cessèrent progressivement de produire des alternances de regards tandis que les babouins du groupe *Face* en produisirent de plus en plus (Bourjade, 2016; Lamaury et al., 2019). De surcroit, les individus du groupe *Face* ont progressivement associé leurs gestes à des alternances de regards alors que les babouins du groupe *Profil* n'ont pas développé ce comportement d'attention conjointe. Cette étude souligne l'importance de l'accès au visage de l'expérimentateur pour développer le comportement coordonné d'attention conjointe.

3.2.3. Développement de la reconnaissance de l'attention visuelle

Dans l'étude suivante, nous avons investigué l'effet des deux conditions d'entrainement sur les capacités des babouins des deux groupes à discriminer les indices attentionnels de l'expérimentateur humain (Bourjade, Gaunet, et al., 2019). Le groupe *Profil*, contrairement au groupe *Face* n'avait pas accès au corps de face, au visage de face, ni aux yeux de l'expérimentateur (pour une illustration de la méthode, voir: Bourjade, 2016). Nous avons testé la capacité des babouins à discriminer des indices d'attention signifiants sans avoir été entrainés à le faire. En effet, l'entrainement au geste de requête opéré sur les babouins du groupe *Face*, de même que tous les essais de motivation réalisés pendant la phase de test[58], représentent des situations d'apprentissage associatif entre une récompense alimentaire et un indice quelconque de l'environnement. Les babouins auraient pu apprendre par conditionnement que pour fonctionner, le geste devait être adressé à un humain positionné à 1 mètre, ou encore à un humain fermant la bouche. Cette étude a donc cherché à tester systématiquement l'hypothèse d'un apprentissage associatif responsable des résultats obtenus sur le groupe *Face*. En étudiant les compétences de discrimination du groupe *Profil*, nous pouvions ainsi : (1) tester leur capacité spontanée à discriminer des indices attentionnels signifiants hors entrainement explicite, (2) évaluer l'influence de l'entrainement en comparant la réponse aux indices entrainés et non entrainés, (3) tester l'effet additif ou exclusif de l'entrainement explicite

[58] Dans les études présentées ici, mais également dans toutes les autres ; nous avons argumenté sur le fait que l'apprentissage associatif était présent dans toutes les études, y compris chez les grands singes n'ayant pas subi d'entrainement « explicite » mais étant exposés aux essais de motivation.

comparé à une connaissance implicite des indices d'attention signifiants, et (4) tester l'effet de récence et la mémoire à long terme de ces discriminations.

Nos résultats ont montré que les babouins du groupe *Profil* testés immédiatement après l'entrainement (1) ne discriminent pas les indices attentionnels signifiants comme les yeux ouverts, (2) répondent davantage aux indices entrainés que non entrainés, (3) ne répondent pas plus aux conditions mélangeant des indices entrainés et des indices naturellement signifiants (ex : corps de profil et visage de face avec les yeux ouverts), (4) présentent un effet de récence fort avec une modification significative de leurs performances une année plus tard. En effet, lors d'une deuxième phase de test différée d'une année, les individus du groupe *Profil* (1) discriminent spontanément l'orientation du corps de l'expérimentateur sans y avoir été entrainés, (2) ne répondent pas plus aux indices qui avaient été entrainés qu'à d'autres indices, et (3) répondent plus aux indices non entrainés naturellement signifiants qu'auparavant. Par ailleurs, aucun individu du groupe *Profil* n'utilise de geste sonore dans les conditions d'inattention visuelle de l'expérimentateur.

Ces résultats indiquent que l'apprentissage associatif est présent dans les phases d'entrainement aux gestes de requête alimentaire mais ne permet pas d'expliquer l'ensemble des comportements observés. Sous l'hypothèse de l'apprentissage associatif, nous attendions que les babouins produisent plus de gestes dans les situations les plus vraisemblables à celle de l'entrainement (où l'association geste – indices de l'environnement – nourriture a été réalisée). Cette hypothèse est corroborée par les résultats obtenus immédiatement après l'entrainement sur les babouins du groupe *Profil*, et par le comportement gestuel des babouins du groupe *Face*. Toutefois, elle ne permet pas d'expliquer (1) que les babouins du groupe *Profil* aient discriminé l'orientation du corps sans y avoir été entrainés, (2) que ces mêmes babouins répondent plus aux indices naturellement signifiants et moins aux indices entrainés en test différé, (3) que les babouins du groupe *Face* développent des comportements d'attention conjointe coordonnant leurs gestes avec des alternances de regards, et (4) que ces mêmes babouins utilisent des gestes sonores quand l'humain est visuellement inattentif sans y avoir été entrainés (Table 1).

3.3. Intentionnalité de la communication intraspécifique

3.3.1. Répertoire de gestes communicatifs

Dans l'étude suivante, nous avons élargi l'investigation de la communication gestuelle à l'ensemble du répertoire de gestes, en situation de communication intraspécifique (Molesti et al., 2020). L'étude a été menée sur un groupe captif de 47 individus observés au cours d'une année. Nous avons répertorié 67 gestes visuels, audibles et tactiles utilisés de manière flexible dans différents contextes fonctionnels (agonistique, affiliatif, reproduction, …). Notre définition du geste communicatif[59] n'impliquait pas a priori un usage intentionnel, mais permettait d'en étudier la possibilité. En nous appuyant sur les critères publiés dans la littérature, nous avons étudié pour chaque geste observé, s'il était (i) associé à un comportement visuel en direction du receveur, (ii) suivi d'une attente de réponse dans le cas où une réponse immédiate n'était pas observée, et (iii) destiné à un receveur visuellement attentif (Hobaiter & Byrne, 2011a; Liebal et al., 2006; Townsend et al., 2017).

Comme cela est classiquement observé chez les grands singes (Byrne et al., 2017), nous avons observé que la taille du répertoire de gestes diminue avec l'âge, les juvéniles[60] étant la catégorie d'âge présentant la plus grande diversité de gestes ; et les gestes silencieux-visuels étant représentés en plus grandes proportions que les gestes audibles et tactiles chez les subadultes[61] et les adultes[62]. Les babouins produisent 90.5% de leurs gestes accompagnés d'un suivi visuel vers le receveur et 87% sont suivis d'une attente de réponse. Dans les deux cas, les enfants[63] diffèrent significativement des autres catégories d'âge en présentant une proportion inférieure de gestes accompagnés d'un suivi visuel ou suivis d'une attente de réponse, indiquant une acquisition développementale. Nous avons également montré que 81.2% des gestes étaient dirigés vers un receveur visuellement attentif (estimé par le fait qu'il avait le regard ou la tête dirigé vers le signaleur). La proportion de gestes tactiles augmentait significativement quand le receveur n'était pas visuellement attentif comparée aux proportions de gestes visuels et audibles qui diminuaient. Ces résultats suggèrent que les babouins ont une communication sensible à

[59] Un geste était défini comme un mouvement corporel, manuel ou facial destiné à un partenaire ou une audience spécifique par le biais de l'orientation du corps, le contact visuel ou le contact tactile.

[60] Dans cette étude, les individus âgés de [1 à 4[ans.

[61] Dans cette étude, les individus âgés de [4 à 7[ans.

[62] Dans cette étude, les individus de 7 ans et plus.

[63] Dans cette étude, individus âgés de moins d'un an.

l'attention en dehors du dispositif expérimental de requête alimentaire mais que la modalité favorisée est la modalité tactile, et non la modalité auditive comme les expériences présentées ci-dessus l'ont montré. L'ensemble de nos résultats suggère également une acquisition de l'usage intentionnel et social des gestes au cours du développement.

3.4. Importance des ruptures attentionnelles et de l'accès au visage

3.4.1.1. Synthèse des travaux sur la communication gestuelle multimodale

Ces travaux expérimentaux et observationnels menés chez les babouins olive en captivité apportent de nouvelles connaissances sur l'acquisition et l'évolution de la communication multimodale. Dans un premier temps, nos études ont permis de montrer que la communication gestuelle, bien qu'essentiellement visuelle devenait multimodale dans les situations de ruptures attentionnelles. Les babouins impliqués dans le dispositif expérimental de la requête alimentaire produisaient des gestes sonores face à un humain visuellement inattentif (Bourjade, Meguerditchian, et al., 2014). En situation de communication intraspécifique, ils utilisaient un geste tactile pour s'adresser à un congénère visuellement inattentif (Molesti et al., 2020). Ces résultats suggèrent que la communication audible/vocale intentionnelle pourrait être étayée sur l'intention d'attirer l'attention des autres quand la modalité tactile n'est pas disponible (i.e., en situation de communication distale stricte). La communication sensible à l'attention et l'usage des gestes sonores pouvant représenter des précurseurs potentiels du langage multimodal (Bourjade, Meguerditchian, et al., 2014; Hopkins et al., 2007; Leavens et al., 2004). Le rôle des ruptures attentionnelles dans l'acquisition d'une communication intentionnelle multimodale sera investigué plus en avant dans l'Axe 1 du programme de recherche proposé ci-après.

Ces travaux ont clairement montré que l'attention visuelle du receveur était essentielle au développement d'une communication gestuelle intentionnelle, de même que l'accès aux indices attentionnels d'autrui (Table 1). En effet, la manipulation des expériences individuelles des babouins indique que l'accès au visage est un facteur essentiel du développement de l'attention conjointe, de la coordination des gestes et des regards, et de l'ajustement des gestes à l'état attentionnel du destinataire (Bourjade, Gaunet, et al., 2019; Lamaury et al., 2019). Plus particulièrement, il est intéressant de rappeler que les babouins naïfs (non entrainés) présentaient des comportements gestuels rudimentaires et des regards vers le visage de l'expérimentateur, caractéristiques d'un engagement dyadique (contact visuel mutuel). L'entrainement a permis aux sujets qui en ont bénéficié de passer d'une attention dyadique

(regard vers le visage de l'expérimentateur) à une attention triadique (alternance entre le visage de l'expérimentateur et la banane et plus de temps passé à regarder la banane) ; une étape clef du développement sociocognitif chez l'enfant (Bakeman & Adamson, 1984; Beuker et al., 2013; Carpenter & Call, 2013).

Table 1. Synthèse des travaux postdoctoraux et plus récents sur la communication gestuelle du babouin olive

Expérience individuelle	Type de communication	Attention conjointe	Indices attentionnels	Attracteurs attentionnels	Communication intentionnelle	Références
Accès au visage humain	Interspécifique	Oui	Direction du regard	Gestes auditifs	Regard; Effet d'audience; Persévérance; Elaboration	Bourjade et al 2014; Bourjade et al 2015
Pas d'accès au visage humain	Interspécifique	Non	Orientation du corps	Aucun	Persévérance	Bourjade 2016 ; Lamaury et al 2019; Bourjade et al 2019
Vie sociale de l'espèce	Intraspécifique	Oui	Orientation de la tête[1]	Gestes tactiles	Regard; Effet d'audience; Attente de réponse[2]	Molesti et al 2019

[1] L'orientation de la tête était l'indicateur utilisé par l'observateur mais il est possible que les babouins utilisent des indices plus précis (comme la direction du regard).

[2] Les critères de persévérance et d'élaboration n'ont pas encore été analysés sur ce jeu de données.

L'ensemble de ces travaux réalisés sur le babouin en captivité souligne la nature résolument *développementale* de l'usage intentionnel des signaux de communication des différentes modalités. La comparaison des groupes *Face* et *Profil* permet d'éclairer les *mécanismes* développementaux à l'œuvre dans ce processus. *L'apprentissage associatif* semble former une base permettant aux babouins d'associer la production du geste de requête avec l'obtention d'une récompense alimentaire sous certaines conditions qui doivent être respectées (et qui variaient selon le mode d'entrainement). Une fois acquis, ce geste peut être utilisé de manière relativement flexible permettant aux babouins de groupe *Profil* de l'utiliser face à des indices non entrainés, indiquant de manière fiable l'attention visuelle d'autrui. Les babouins du groupe *Face* qui ont eu accès au visage de l'humain ont développé des comportements d'attention conjointe en coordonnant leurs gestes avec des alternances de regards, et une communication sensible à l'attention, reposant sur un ajustement unimodal et intermodal avec l'usage de gestes sonores (Table 1). Ces acquisitions ont été interprétées comme le passage d'un 'outillage initial' (apprentissage associatif du geste et des indices attentionnels) sur lequel est venu s'étayer le déploiement de moyens communicatifs élaborés (attention conjointe, ajustement intermodal) qui signalent une compréhension plus globale de la situation communicative (Bourjade, Gaunet,

et al., 2019). Une interprétation vygotskienne de ces résultats serait que les babouins ont appris le geste dans la relation inter-psychologique avec l'humain, puis se sont appropriés de manière intra-psychologique le potentiel communicatif (et la signification) du geste de sorte qu'ils peuvent l'utiliser ensuite de manière flexible et ajustée à des situations perçues dans leur globalité.

3.4.1.2. La communication sensible à l'attention dans la littérature

Les études présentées ci-dessus sur les babouins olive ne sont pas les seules à rapporter une communication sensible à l'attention du receveur. Comme introduit plus haut, Leavens, Hopkins et leurs collaborateurs ont rapporté l'usage par les chimpanzés captifs d'attracteurs attentionnels composés de gestes audibles, de sons oraux non vocaux et de vocalisations atypiques (Hopkins et al., 2007; Leavens et al., 1996, 2004, 2010; Leavens & Hopkins, 1998). De même, une étude pionnière chez les enfants humains a permis de mettre en évidence l'usage des vocalisations par les enfants de 18 mois envers un adulte inattentif (Liszkowski et al., 2008). Dans l'ensemble, ces études concourent à montrer que le canal auditif peut être investi pour regagner l'attention d'un partenaire inattentif, de différentes manières selon les espèces : par les gestes sonores chez les babouins, par les vocalisations, les sons oraux et les gestes sonores chez les chimpanzés et par les vocalisations chez les enfants humains (Figure 3).

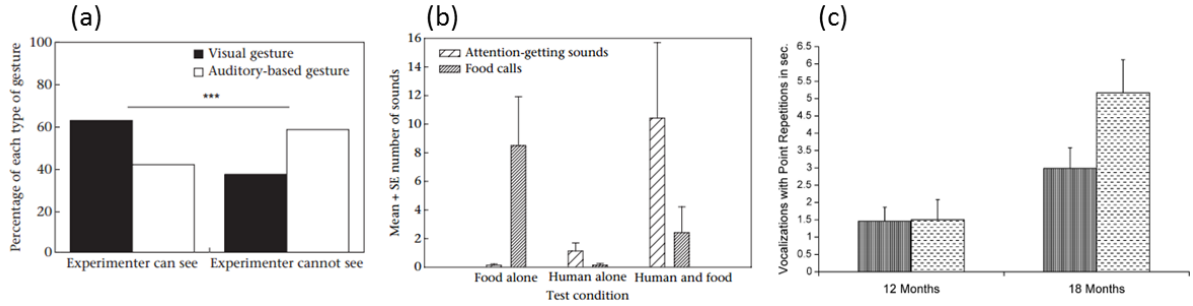

Figure 3. Communication sensible à l'attention chez les babouins, les chimpanzés et les enfants humains. (a) gestes acoustiques manuels des babouins, adaptés de (Bourjade, Meguerditchian, et al., 2014), (b) bruits non vocaux des chimpanzés, adaptés de (Hopkins et al., 2007), (c) durée moyenne des vocalisations des bébés de 18 mois au cours de pointages répétés, lorsque l'adulte est attentif (gris foncé) ou inattentif (gris pâle), adapté de (Liszkowski et al., 2008).

Une étude plus approfondie de la littérature suggère que la communication sensible à l'attention est un phénomène répandu chez les PNH en captivité, de même que dans la nature, et chez les enfants préverbaux entre 12 et 18 mois également (Bourjade & Hobaiter, *en préparation*). Les études expérimentales chez les PNH montrent que les chimpanzés en situation de requête

alimentaire utilisent plus souvent les vocalisations et les attracteurs attentionnels quand l'humain a le dos tourné que lorsqu'il est de face (Hostetter et al., 2007; Hostetter et al., 2001; Leavens et al., 2010). Chez les singes non anthropoïdes, les vocalisations et gestes sonores sont généralement produits de manière très marginale et ne sont pas plus fréquents quand l'expérimentateur est visuellement inattentif (Aychet et al., 2020; Canteloup et al., 2015; Maille et al., 2012; Meunier, Prieur, et al., 2013) excepté dans les études rapportées ci-dessus. En revanche, chimpanzés, mangabeys et macaques viennent se repositionner dans la ligne de vue d'un expérimentateur humain quand celui-ci n'est pas visuellement attentif (Aychet et al., 2020; Deshpande et al., 2018; Liebal, Call, Tomasello, et al., 2004), ce qui est une forme de communication sensible à l'attention avec un ajustement unimodal visuel.

En situation de communication intraspécifique, les PNH semblent favoriser la modalité tactile lorsque le receveur est visuellement inattentif, bien qu'une étude plus approfondie soit nécessaire à ce sujet. En effet, 60% des gestes tactiles des grands singes observés en captivité sont adressés à un partenaire visuellement inattentif tandis que 80% des gestes silencieux-visuels sont adressés à un receveur attentif (Pika et al., 2005). Nous avons répliqué ce résultat chez les babouins (Molesti et al., 2020) mais dans une récente étude développementale chez le chimpanzé sauvage, nous avons mis en évidence un changement de modalité préférentielle en cas d'inattention visuelle : les juvéniles délaissent progressivement la modalité tactile pour la modalité audible quand leur mère est visuellement inattentive (Dafreville et al., 2021). Une autre étude sur la même communauté de chimpanzés a récemment mis en évidence l'utilisation de vocalisations combinées à des gestes après un échec de communication vocale (Hobaiter et al., 2017). Dans l'ensemble ces résultats indiquent que la communication sensible à l'attention peut emprunter alternativement le canal tactile ou le canal audible et il parait maintenant nécessaire de clarifier les contextes et les déterminants de l'usage de chacune des deux modalités.

Chez l'enfant, la situation expérimentale de référence est celle du geste de pointage vers des cibles imagées lors de la lecture d'un livre avec un parent ou un expérimentateur. Typiquement, l'adulte se comporte de différentes manières selon des conditions expérimentales faisant varier : (i) son attention visuelle vers l'enfant ou vers la cible du pointage, (ii) la présence/absence de réponse aux pointages, et (iii) l'adéquation de la réponse à la cible pointée par l'enfant. Toutes les études montrent que les enfants produisent des vocalisations associées à leurs pointages mais la distribution des vocalisations dépend de l'âge. Les enfants de 10 mois vocalisent davantage quand les adultes les regardent (Wu & Gros-Louis, 2017), tandis qu'à 12 et 18 mois

ils vocalisent plus quand l'adulte ne les regarde pas, ou qu'il ne regarde pas la cible pointée ou qu'il répond en parlant d'autre chose que de la cible pointée (Gros-Louis, 2012; Igualada et al., 2015; Liszkowski et al., 2008).

Cette première analyse de la littérature nous a permis de soulever plusieurs écueils et d'envisager des directions pour de futures études. En effet, les études réalisées chez les PNH montrent un biais 'gestuel' dans l'approche de la multimodalité. Les gestes sonores sont échantillonnés mais la plupart du temps, les vocalisations ne le sont pas, appelant à réaliser des études *réellement* multimodales. Chez les enfants, le même biais existe, mais cette fois il est vocal. Les gestes sonores ne sont pas échantillonnés dans ces études et seuls les pointages et les vocalisations sont considérées. Il est donc nécessaire là aussi de mener des études permettant de quantifier de manière systématique l'usage de chacune des modalités et leurs combinaisons. Par ailleurs, les études mélangent souvent différents contextes ou conditions expérimentales qui ne permettent pas toujours de savoir si l'usage des modalités tactiles ou audibles est choisi préférentiellement dans les situations d'inattention visuelle ; en d'autres termes, les contextes permettant de faire émerger l'usage intentionnel des différentes modalités est sous-spécifié. Enfin, pour répondre et dépasser la controverse sur les attracteurs attentionnels, il serait nécessaire de qualifier les différentes significations des gestes à composantes tactiles et audibles, de même que celles des combinaisons gestes-vocalisations, par exemple en utilisant la méthode de l'ASO (*Apparently satisfying Outcome*) (Hobaiter & Byrne, 2014, 2017) afin de lever le doute sur leur sémantique – à savoir si ces attracteurs attentionnels ne servent qu'à attirer l'attention ou s'ils sont porteurs d'information véhiculée par un autre canal sensoriel. Cette question adresse plus largement la problématique de la redondance et de la complémentarité dans la communication multimodale ; une question qui pourrait être abordée sous un angle nouveau en considérant ces deux options sous l'angle de la pragmatique. Ces voies d'investigation sont pour l'essentiel, reprises dans le chapitre suivant qui présente mon programme de recherche à court, moyen et long termes.

Chapitre 4. Les passerelles entre geste et voix

Sur la base des travaux empiriques présentés dans le Chapitre 3 et des travaux théoriques mettant en correspondance les concepts des différentes disciplines impliquées dans ce questionnement scientifique (Bourjade, 2016, 2017; Bourjade et al., 2020), je développe ci-dessous les éléments structurants de mon programme de recherche pour les 10 à 15 années à venir. Les éléments de continuité (Axe 1) sont évidents puisque la fondation du questionnement autour du rôle des ruptures attentionnelles dans l'émergence du contrôle vocal s'ancre directement dans les travaux réalisés entre 2011 et 2017 sur les babouins olive. Les éléments d'innovation qui ont émergé plus récemment portent principalement sur des conceptions multimodales et multidimensionnelles des structures linguistiques et s'inscrivent dans le champ de la *linguistique animale* (Axe 2). Deux grandes questions vont permettre de guider l'ensemble de ce programme de recherche et d'en délimiter les étapes et projets :

1. Les ruptures attentionnelles permettent-elles de faire émerger un usage différentiel et/ou combiné des modalités communicationnelles ? Quels rôles jouent ces ruptures attentionnelles (de la part du receveur) au cours du développement communicatif, dans le cadre des triangulations attentionnelles, ou encore dans l'émergence de la relation forme/référent ?
2. Quelles contraintes motrices, exécutives et sociales façonnent l'émergence de structures linguistiques multimodales dans la communication du babouin olive ?

Ces deux questions constituent deux passerelles entre les actions gestuelles (corporelles, manuelles ou orofaciales) et les actions vocales (activation des cordes vocales au sens strict) qui seront explorées à l'aide de différents projets structurés sous deux axes principaux. Différentes méthodologies empruntées à la psychologie développementale, à l'éthologie, à la linguistique, ou encore à l'anthropologie et aux neurosciences seront utilisées pour répondre à ces questions. Ces travaux permettront d'éclairer les mécanismes de couplage intention-modalité de communication à différentes étapes du développement de l'enfant. Ils fourniront également un soutien ou une réfutation empirique à certaines hypothèses théoriques d'évolution du langage. Les perspectives qui en découleront et qui sont présentées dans la section 4.3. seront développées selon deux mouvements opposés. L'un visera à stimuler une recherche plus théorique et fondamentale qui passera par la mathématisation des scénarios évolutifs ; l'autre s'orientera vers une recherche translationnelle qui saura mettre à profit les connaissances

nouvellement obtenues sur les liens entre attention, intention et modalité(s) sensorielle(s) pour améliorer l'évaluation et la prise en charge de situations cliniques comme la dépression post-partum, ou encore la prématurité.

4.1. Axe 1 : Rupture attentionnelle et communication multimodale

4.1.1. Les modèles théoriques de conflit attentionnel

> *"The interpretation of tomb paintings – that parents regarded their children as essential members of the family – is corroborated by the prominence given their children in the official art of the heretic pharaoh Akenaten and his wife Nefertiti (ca. 1378-1362 BCE). King and queen have their children with them at state ceremonies, hold them on their laps, and kiss and embrace them warmly. There is no mistaking the fondness of these royal parents for their children – or at least as part of the public message conveyed by these official representations: conjugal and parental affection is to be emulated."*
>
> *Valerie French (2019, p.292)*

La parentalité est un phénomène remarquablement indissociable de la notion *d'animal vivant* si on la considère dans son acception large. La nature de l'investissement que les parents réservent à leur progéniture est classiquement séparée en deux stratégies bien distinctes. Dans la stratégie *r*, les parents donnent naissance à un grand nombre de descendants qu'ils abandonnent à la naissance ; l'investissement est majoritairement prénatal. Dans la stratégie *k*, à laquelle répondent les mammifères en général, les parents produisent moins de descendants, mais s'investissent auprès d'eux pour maximiser leur survie (MacArthur & Wilson, 1967). Les primates, et spécifiquement les humains, présentent un investissement parental fort qui répond à la fois à la théorie de la sélection de parentèle ; i.e., l'idée d'augmenter sa valeur reproductive en investissant dans ses descendants et apparentés (Hamilton, 1964), et à celle, associée, du conflit parental (Trivers, 1974). La sélection de parentèle est efficiente pour expliquer les comportements altruistes envers les individus apparentés. Suite aux travaux d'Hamilton (1964), Trivers (1974) étend le raisonnement à la relation parent-enfant en proposant un conflit d'investissement entre parents et enfants. Du point de vue du parent, il s'agit d'investir suffisamment vers *tous* les descendants pour qu'ils survivent et se reproduisent, sans affecter sa propre survie. Du point de vue de l'enfant, il s'agit de maximiser sa propre valeur adaptative en obtenant l'investissement maximal du parent. De cette économie, résulte un conflit où parents et enfants divergent sur la limite de l'investissement : la sélection naturelle favorise chez l'enfant une demande supérieure à ce que le parent peut donner.

Plusieurs propositions théoriques sur l'évolution du langage s'inscrivent admirablement bien dans le cadre de ce conflit parental, même si elles n'y font pas toutes explicitement référence. Par exemple, Fitch (2010) propose que l'échange d'information entre parents et enfants soit un mécanisme clef de l'évolution du langage permettant de résoudre le problème de la communication coopérative et son maintien au cours de l'évolution. De même, Falk (2004) souligne l'importance de la communication mère-enfant dans l'émergence des substrats d'un protolangage. Elle propose que l'émergence de la bipédie ait entrainé des modifications importantes de la proximité et de l'interaction mère-enfant liées à des modifications anatomiques : la bipédie se généralisant, le pied s'est rigidifié de sorte qu'il n'est plus possible pour le jeune de s'agripper à la mère comme le font les autres primates, le bassin de la mère plus étroit mène à des naissances précoces et en conséquence les bébés sont nidicoles (*altricial*) et immatures sur le plan locomoteur (entre autres). La bipédie a ainsi rendu obsolète la résolution du conflit d'investissement parental observée chez les quadrupèdes. En effet, chez les autres primates, le jeune peut s'agripper précocement à se mère et il est responsable du maintien de la proximité (Bard, 2019). Falk (2004) propose que les mères humaines aient peu à peu développé le chant et le « parler bébé » (*motherese*) pour rassurer vocalement les enfants immobiles et posés au sol au cours de leurs activités (e.g., cueillette, feu, soin à un autre enfant etc.).

Dans la même lignée, Mehr & Krasnow (2017) proposent que ce conflit parental pour l'attention (au sens large) ait donné lieu à une « course aux armements », c'est-à-dire un raffinage parallèle des moyens mis en œuvre pour influencer le comportement d'autrui à son avantage. Dans le cadre du conflit attentionnel parent-enfant, les auteurs proposent que les parents aient progressivement déployé des stratégies de plus en plus efficaces (et de moins en moins couteuses) pour rassurer leurs enfants à distance, comme le chant ou le « parler bébé ». Ils proposent également que les enfants aient raffiné leurs stratégies pour demander, maintenir ou regagner l'attention du parent. Cette proposition est compatible avec l'hypothèse de Locke (2006, 2017) selon laquelle la sélection aurait favorisé les nourrissons qui produisaient des babils de bonne qualité. Le babillage aurait pris progressivement la valeur d'un indicateur de la valeur adaptative de l'individu ; par les babils de haute qualité, les bébés auraient fait la démonstration de leur bonne santé à leurs parents, ce qui aurait contribué à stimuler l'investissement parental (Locke, 2006, 2017; Locke & Bogin, 2006; Oller et al., 2016). Cette proposition est également compatible avec l'hypothèse de la communication sensible à

l'attention (*Attention-sensitive signaling hypothesis*) que j'ai proposée récemment et que j'introduis ci-dessous (Bourjade, 2019).

4.1.1.1. L'hypothèse de la communication sensible à l'attention

Il s'agit d'une articulation des modèles théoriques essentiellement « musicaux/vocaux » présentés ci-dessus avec les arguments empiriques et théoriques provenant de la communication gestuelle/visuelle des enfants préverbaux et des primates non humains (Bourjade, *en préparation*). A cet égard, Tomasello (2008) propose que le langage ait évolué sur la base d'une motivation coopérative propre à l'espèce humaine. Il s'appuie sur la nature *déclarative* de la communication gestuelle des enfants préverbaux qui pointent pour *partager leur intérêt* à propos de quelque chose avec quelqu'un (Liszkowski et al., 2004; Tomasello, Carpenter, et al., 2007). Plus précisément, il propose que la *communication coopérative* repose sur la capacité *d'intentionnalité partagée* à l'œuvre notamment dans *l'attention conjointe partagée* et dans la capacité des enfants à lire, utiliser et diriger l'attention visuelle d'autrui vers des cibles externes. Parmi les arguments mobilisés, on trouve *l'hypothèse de l'œil coopératif* selon laquelle la morphologie de l'œil humain se prête admirablement à cette fonction d'utilisation/direction de l'attention visuelle par ses contrastes de couleurs entre la sclérotique blanche et l'iris foncé (Kobayashi & Kohshima, 2001; Tomasello, Hare, et al., 2007). A ce stade, il convient de préciser que le caractère unique du contraste de l'œil humain a récemment été remis en question puisque les grands singes possèdent un contraste équivalent avec des luminances inversées entre la sclérotique et l'iris (Perea-García et al., 2019). De plus, si les hypothèses de Tomasello évoquées ci-dessus ont le mérite de souligner des aspects importants de la cognition sociale humaine, elles ne sont malheureusement pas formulées dans un langage opérationnel permettant de les tester. Ces mises en garde réalisées, l'utilisation pervasive de l'attention socialement médiée par les humains, de même que les capacités endogènes de détection du regard et leur signature neurale propre (Böckler et al., 2014; Bourjade, 2017; Klein et al., 2009; Shepherd, 2010; van der Wel et al., 2018) incitent à ne pas éliminer trop rapidement les hypothèses évolutionnaires donnant un rôle majeur au regard ou au visage. Ainsi, j'articule les modèles présentés ci-dessus, avec l'importance de l'attention visuelle telle que Tomasello la souligne dans ses différents travaux (Tomasello, 2006, 2008, 2014; Tomasello & Carpenter, 2007).

L'hypothèse de la communication sensible à l'attention (*Attention-Sensitive Signaling Hypothesis* ; ASSH) propose que l'émergence de la bipédie ait modifié la proximité physique

entre mère et enfant, rendant obsolète la résolution du conflit attentionnel d'investissement parental opéré par les quadrupèdes, i.e., le jeune nidifuge (*precocial*) peut se rapprocher de sa mère, grimper et s'agripper à elle en cas de détresse/de besoin de réassurance (Bard, 2019; Falk, 2004). La communication distale adressée à l'enfant, sous forme de « parler bébé » (Falk, 2004), de chant (Mehr & Krasnow, 2017) ou de regards, sourires et gestes, i.e., la *communication distale multimodale*, aurait donc commencé à être utilisée par les mères pour rassurer leurs enfants à distance. Ce faisant, les nourrissons extrêmement dépendants de leurs parents pour assurer leur survie auraient été sous de fortes pressions de sélection pour signaler leurs besoins et recevoir de l'attention parentale ; l'usage des vocalisations aurait permis aux nourrissons de capturer l'attention du parent, même à distance. Selon Locke (Locke, 2006, 2017), certaines vocalisations auraient ainsi acquis une valeur indicatrice de la bonne santé de l'enfant. Selon l'hypothèse de la communication sensible à l'attention, la rupture de la proximité physique d'avec la mère aurait progressivement exposé les enfants au *visage* de leur mère et à ses *indices d'attention visuelle*. Les enfants devenant de bons « lecteurs » de visage et d'attention socialement médiée auraient peu à peu discriminé plus efficacement l'état attentionnel de la mère, leur permettant d'ajuster leur communication visuelle et vocale. En conséquence, les enfants plus audibles/vocaux en situation d'*inattention* visuelle de la mère auraient bénéficié d'un avantage sélectif sur les autres du fait qu'ils recevaient plus d'attention et de soins (Bourjade, 2019).

Cette hypothèse n'est pas mutuellement exclusive avec celles qu'elle propose d'articuler mais propose un mécanisme permettant d'expliquer l'émergence de la préférence vocale en situation de rupture attentionnelle chez l'espèce humaine. Elle offre un rôle inhabituel aux micro-contextes sociaux, comme la variation d'attention maternelle, de sorte qu'il convient de tester sa validité par des plans expérimentaux intra-sujets avant de tester l'effet général de la production vocale inter-sujets[64]. Elle fournit des forces sélectives plausibles à l'émergence du contrôle vocal dans la lignée humaine, de même qu'à l'usage combiné des modalités visuelles et audibles des signaux de communication.

[64] Ce qui serait équivalent à tester l'hypothèse de Locke (2006) bien que celle-ci porte plus spécifiquement sur le babillage.

4.1.1.2. Prédictions testables

Il est possible de formuler des hypothèses opérationnelles et des prédictions testables à partir de ces théories du conflit attentionnel (Bourjade, *en préparation*). Voici les grandes hypothèses qui pourront être testées dans cet axe de recherche (et qui seront déclinées dans le projet 1.1.) :

(H1 : Falk)* La communication distale (vocale ou gestuelle) est négativement corrélée à la proximité physique entre mère et enfant (e.g., quantité de temps en contact).

(H2 : alternative à H1)* La communication distale (vocale ou gestuelle) est privilégiée en cas de nidicolie, d'immaturité locomotrice ou de contention du jeune/du sujet.

(H3 : Locke)** # Les babils des nourrissons induisent plus d'investissement parental (taux de réponse, responsivité, positivité de la réponse) que les autres vocalisations.

(H4 : Locke)** Les enfants augmentent leur fréquence de babillage envers des individus moins apparentés que les parents (frères/sœurs, tantes etc.) pour stimuler l'investissement attentionnel.

(H5 : Bourjade)* Le temps passé exposé au visage de la mère est inversement corrélé au temps passé en contact physique direct.

(H6 : Bourjade)# L'exposition au visage de la mère accroit la sensibilité à l'attention (et l'inattention) du receveur dans le développement communicatif.

(H7 : Bourjade)** Les ruptures attentionnelles (essentiellement visuelles) favorisent la diversification des modalités de communication et l'usage de la communication vocale/audible.

(H8 : alternative à H7)** La violation des attentes (e.g., le parent ne regarde pas la cible du pointage ou ne satisfait pas une requête) favorise la diversification des modalités de communication et l'usage de la communication vocale.

* Hypothèses qui pourront être testées sur la base du comportement maternel ;

** Hypothèses qui seront testées sur la base du comportement communicatif de l'enfant en intra-sujet ;

Hypothèses qui seront testées sur la base du comportement communicatif de l'enfant, en inter-sujet.

Ces hypothèses, formulées dans un langage encore sous-spécifié, en particulier sur le plan développemental, seront précisées et testées à l'aide de comparaisons interspécifiques et interculturelles (i.e., entre différentes populations humaines et différents modes de vie ; Projet 1.1. ci-dessous).

4.1.2. Projet 1.1. Développement d'une communication sensible à l'attention

Ce projet vise à examiner le rôle des ruptures attentionnelles dans la diversification des modalités de communication utilisées par les enfants de différentes espèces et populations envers leur mère. Il permettra de tester les huit hypothèses proposées ci-dessus en les déclinant de manière précise selon les comparaisons considérées. La première étape consistera à développer une base de données de signaux visuels, acoustiques et multimodaux produits par les nourrissons et attribués aux différents modes d'attention de la mère (tels que définis dans la section 2.3.3.2.), en utilisant des observations quantitatives sur six populations différentes : babouins et chimpanzés sauvages, enfants français urbains, enfants Gamo ruraux, et enfants mozambicains, ruraux et urbains. Cette étape permettra de calculer la fréquence des modes d'attention maternelle en fonction de l'espèce, et des facteurs éco-culturels chez l'humain. Cette analyse générera de nouvelles connaissances sur la variabilité des modes attentionnels des mères, étape nécessaire à l'établissement d'un lien entre l'attention maternelle et l'émergence de la communication multimodale chez les nourrissons.

Des paires focales mère-enfant sont en cours d'échantillonnage dans six populations ; les babouins sauvages (*Papio anubis*) sont observés sur le plateau de Laïkipia au Kenya (projet Uaso Ngiro, en collaboration avec Shirley Strum). Les chimpanzés sauvages de la communauté Sonso (Budongo, Ouganda) sont observés à partir d'un corpus vidéo préexistant partagé par Catherine Hobaiter. Les enfants français sont recrutés via le *babylab* de la plate-forme CCU (MSHS-T, UT2J / CNRS, Toulouse) et ont été filmés par Mawa Dafreville dans le cadre de sa thèse sur le projet (co-encadrement Michèle Guidetti). Les enfants Gamo (Ethiopie) seront observés à partir d'un corpus vidéos préexistant partagé par Lauren Bader (Bader & Fouts, 2018), et les enfants mozambicains urbains et ruraux seront observés à partir d'un corpus vidéo préexistant partagé par Paul Vogt (Mastin & Vogt, 2016). La taille des échantillons sera comprise entre [14; 30] paires mère-enfant pour chaque population.

Dans une étude expérimentale réalisée sur des bébés de six mois, nous avons pu montrer qu'à cet âge, le nourrisson réagissait plus à l'interaction de sa mère qu'à son état d'attention visuelle (Bourjade et al., soumis). Les bébés de cet âge ne présentent pas de communication sensible à l'attention mais évitent de vocaliser en même temps que leur mère. Sur la base de cette référence, nous avons recruté des enfants de 7 à 18 mois pour l'étude présentée ici. Les périodes de développement vont de 4 à 12 mois chez les babouins, et de 9 à 36 mois chez les chimpanzés, permettant pour chaque espèce un échantillonnage de l'avant/pendant/après émergence de la communication intentionnelle.

Ce projet a déjà donné lieu à une première thèse en cours (Mawa Dafreville, co-encadrement Michèle Guidetti) et à deux mémoires de master 2 soutenus en 2021 (Claire Lerbet, Université Toulouse 3 ; Noémie Pantel, Université de Rennes 1). Deux mémoires de master 2 supplémentaires pourront être dirigés sur ce sujet et le recrutement d'un post-doc et d'un assistant de recherche est envisagé afin de mener à bien l'étude comparative. Il s'agit également d'un projet reposant sur plusieurs collaborations internationales, rappelées ci-dessous.

Objectif : Quantifier les liens entre modalités communicationnelles des nourrissons et modalités d'attention maternelle chez différentes espèces et populations humaines.

Méthode : Observation en conditions naturelles et approche comparative.

Collaborateurs principaux : Michèle Guidetti & Mawa Dafreville (doctorante sous ma direction et celle de Michèle Guidetti ; Université Toulouse Jean Jaurès) - Paul Vogt (Université de Tilburg) ; Catherine Hobaiter (Université de St Andrews) ; Shirley Strum (Université de Californie, San Diego) ; Lauren Bader (IAST, Toulouse) ; Céline Scola & Marianne Jover (Aix-Marseille Université).

Financements : Bourse doctorale de la région Réunion ; programme de soutien à la mobilité de l'InSHS du CNRS ; Financement de la Maison des Sciences de l'Homme et de la Société de Toulouse ; Institut Universitaire de France.

Etat d'avancement : en cours et jusqu'à 5 ans ; perspectives à moyen terme

4.1.2.1. La comparaison entre espèces

La thèse de Mawa Dafreville porte sur l'usage des modalités de communication selon l'attention visuelle de la mère dans une perspective comparée. Deux articles en cours d'écriture examineront les préférences modales et les capacités d'ajustement unimodal et intermodal d'enfants français âgés de 7 à 18 mois en fonction de l'attention de la mère et de ses comportements distaux (e.g., discours adressé à l'enfant) et proximaux (e.g., portage). Ils cibleront plus particulièrement les hypothèses issues des théories de Falk (H1, H2), Locke (H3, H4) et celles issues de l'ASSH (H6 et H7). Nous attendons un effet de l'âge sur la capacité des nourrissons à ajuster leur communication à l'attention du parent : l'ajustement devrait être meilleur après l'émergence de la communication intentionnelle (H7). De plus, les bébés en âge de marcher pourraient privilégier un ajustement unimodal visuel en se repositionnant dans la ligne de vue du parent afin de s'adresser à lui, ou un ajustement intermodal tactile (H2). A l'inverse, les enfants maintenus dans une chaise pourraient privilégier la modalité auditive pour s'adresser à un parent visuellement inattentif (H2). De plus, le babillage, plus que d'autres vocalisations, pourrait être adressé à la mère visuellement attentive dans le cadre d'interactions

vocale avec l'adulte (H3) et générer une réponse contingente permettant de prédire le développement langagier ultérieur. Un article publié récemment a déjà permis de montrer que les chimpanzés immatures développaient une communication sensible à l'attention de leur mère au cours de la phase juvénile, et qu'elle devenait complètement effective chez les adolescents, i.e., H7 corroborée (Dafreville et al., 2021). En effet, lorsque la mère est visuellement inattentive, la moitié des juvéniles et tous les adolescents favorisent les gestes audibles et tactiles en comparaison aux gestes strictement visuels ; réalisant ainsi un ajustement *intermodal*. Un quatrième article portera sur la comparaison des comportements maternels chez les mères françaises et chez les mères babouins avec le but explicite de tester les hypothèses issues de la théorie de Falk (H1, H2) et l'ASSH (H5). A ce titre, on s'attend à observer une proportion de communication distale plus importante chez les mères humaines que babouins (H1) de même qu'une covariation inverse de la proximité physique et de la communication distale (H1), une fréquence de communication distale qui diminue avec l'indépendance et donc l'âge (H2), un temps d'exposition au visage de la mère (i.e., mère attentive) plus important chez l'enfant humain que babouin (H5).

Plusieurs mémoires de master 1 et 2 permettront d'étudier finement le développement de la communication sensible à l'attention chez l'enfant babouin et cibleront les hypothèses H3, H6, H7 déjà examinées avec les autres espèces. Le conflit de sevrage chez le babouin permettra d'examiner H8 et donnera possiblement lieu à de nouvelles études et comparaisons chez les enfants humains – qui tendent à vocaliser davantage quand l'adulte ne satisfait pas leurs attentes, par exemple ne regarde pas la cible du pointage (Gros-Louis, 2012; Igualada et al., 2015). L'étude sera finalisée par une analyse comparative du développement de la communication sensible à l'attention chez les trois espèces, en indexant les périodes développementales sur l'espérance de vie de chaque espèce (postdoctorat envisagé). Une nouvelle étude plus complète chez le chimpanzé pourra faire l'objet d'une thèse en collaboration, et la comparaison babouin/chimpanzé pourra être valorisée à cette étape.

4.1.2.2. La comparaison entre populations humaines

Comme cela a été introduit dans la section 2.4.2., la variation observée chez l'humain peut être attribuable à une adaptation locale ou à l'évolution culturelle, ou aux deux. Dans tous les cas, se pose la question des *universaux* dans les comportements d'étayage langagiers, d'autant plus que cet étayage est considéré comme un déterminant majeur du développement humain par les

théories socio-culturelles (Bruner, 1975; Vygotsky, 1978). Ainsi, l'étayage du développement communicatif et langagier par le parent pourrait être universel sans nécessairement être uniforme, ou pourrait relever de spécificités culturelles marquées au-delà de la langue maternelle (Keller et al., 2018; Mesman et al., 2018). Spécifiquement, des modèles parentaux et des buts de socialisation différents caractérisent les populations humaines (Keller et al., 2005, 2006). En parallèle, les styles de vie et les environnements jouent un rôle dans les interactions mère-enfant (e.g., Mastin & Vogt, 2016). Les études menées au sein de ce projet dit *interculturel* visent à tester l'effet de l'environnement et l'effet des modèles de parentalité « distaux » et « proximaux », caractéristiques des sociétés industrialisées et traditionnelles, respectivement (Keller et al., 2009).

Le mémoire de Master 2 de Claire Lerbet que j'ai supervisé en 2021 a permis d'investiguer les comportements maternels distaux et proximaux dans les six populations et de commencer à examiner les hypothèses H1 et H2. Nous attendions une corrélation inverse entre le temps passé à porter l'enfant et la fréquence de communication distale (H1), chez les enfants n'ayant pas encore développé la marche (H2)[65]. De même, nous attendions plus de proximité physique chez les mères babouins qu'humaines (H1), et au sein des humains, nous attendions, à la suite de Keller (2009), une plus grande proximité physique chez les Gamo et les mozambicains que chez les français. Les résultats très préliminaires montrent que les liens entre proximité physique et communication distale varient selon les populations humaines mais ne répondent pas aux styles « proximaux » et « distaux », et les mères babouins ne réalisent pas de comportements maternels distaux (i.e., H2 corroborée). Ce travail d'exploration a surtout permis de mesurer l'impact des méthodes de collection de données et a soulevé des problèmes de comparabilité entre les corpus (e.g., une approche transversale pure chez les mozambicains, et mixtes chez les Gamo et français). De plus, il s'agira de développer une méthode d'appariement des âges développementaux[66] en lieu et place de l'âge réel afin de pouvoir comparer les populations. Finalement, la population Gamo (actuellement le plus petit échantillon) se comporte de manière totalement différente des autres et nous envisageons d'étendre la collecte de données afin de mieux saisir les spécificités de cette population.

---

[65] L'avènement de la marche chez l'enfant étant susceptible de le rendre plus autonome et moins dépendant de la volonté de sa mère pour interagir, nous pouvons attendre un pattern proche de celui du babouin avec sa mère, selon H2.

[66] Possiblement sur le critère de l'évitement des vocalisations mutuelles et/ou de l'ajustement intermodal évalué avec une tâche simple.

Le mémoire de Master 2 de Noémie Pantel supervisé en 2021 également a porté sur le développement de la communication sensible à l'attention chez des enfants Gamo et français de 7 à 11 mois avec pour objectif de tester H7 (les ruptures attentionnelles favorisent la diversification des modalités de communication). L'âge des Gamo ne permettait pas de tester réellement l'hypothèse développementale selon laquelle l'émergence de la communication intentionnelle favorise l'ajustement uni- et intermodal. Les premiers résultats indiquent toutefois que les enfants des deux populations développaient une communication sensible à l'attention avec des stratégies modales différentes : les enfants français favorisaient les signaux visuels quand la mère était visuellement attentive alors que les Gamo privilégiaient toujours les signaux tactiles et audibles indépendamment de l'attention *visuelle* maternelle. Les deux styles de communication répondent aux deux modèles proposés par Keller et al. (2009) : proximal (et tactile) pour les Gamo, distal pour les français. Un prochain encadrement de master 2 permettra d'étudier le développement de la communication sensible à l'attention chez les enfants mozambicains en fonction du style de vie rural et urbain.

Cette ligne de recherche pourra être développée plus en avant en répliquant l'étude sur une population masaï vivant sur le plateau de Laïpikia dans deux environnements, l'un rural, l'autre urbain. Une approche longitudinale sera favorisée et ce travail pourra faire l'objet d'une thèse. De plus, un système automatisé de mesure de l'attention visuelle pourrait être envisagée pour poursuivre sur cette ligne de recherche. Enfin, l'essentiel du travail consistera à tester statistiquement les hypothèses mentionnées ci-dessus de manière systématique, en répliquant les comparaisons entre espèces sur les différentes populations humaines.

4.1.3. Projet 1.2. Approche sémiotique du couplage Signe - Référent

L'approche sémiotique s'appuie sur l'utilisation de signes, identifiés par leurs formes différentes, qui peuvent être associées à divers référents pour transmettre des significations distinctes. L'objectif de cette ligne de recherche est de quantifier l'usage et le rôle des modalités communicatives dans l'émergence de la relation entre la forme du signe (sa modalité en est un aspect) et ce à quoi il fait référence (e.g., l'objet, la signification, l'intention) dans des situations naturelles et expérimentales permettant de tester différents mécanismes de couplage entre intention communicative et modalité du signe utilisé sous contraintes attentionnelles. Ce projet comporte deux grandes orientations distinctes qui pourront donner lieu à deux projets de thèse.

La première ligne de recherche est focalisée sur le rôle des modalités visuelles et auditives dans l'attention conjointe et sa valeur prédictive des premiers mots verbalisés. Une étude observationnelle pourra être complétée par une investigation des corrélats neuronaux en jeu dans ces processus. La deuxième ligne de recherche, expérimentale, associe des jeux économiques de *coordination* à une approche de sémiotique expérimentale (Galantucci & Garrod, 2011) pour tester l'hypothèse que l'utilisation linguistique des signes visuels précède celle des signes acoustiques (Fay, Lister, et al., 2014b) et que les deux varient en fonction de l'attention visuelle du destinataire. Deux situations expérimentales impliquant un jeu de coordination cibleront l'interface forme-sens chez les apprenants d'un langage naturel et dans un système de communication émergent construit par des paires de joueurs d'un jeu économique (voir ci-dessous).

Objectif : Evaluer le rôle différentiel des modalités visuelles et auditives dans l'émergence du couplage entre forme et référent, et entre modalité et intention (compris comme un cas particulier du lien entre forme et sens), en faisant varier les contraintes attentionnelles de la part du receveur.

Méthode : Surtout expérimentale.

Collaborateurs principaux : Paul Vogt (Université de Tilburg) ; Lauren Bader (IAST, Toulouse) ; Mawa Dafreville, Sandra Molesti & Fabio Montermini (Université Toulouse Jean Jaurès) ; Gladys Barragan-Jason (Station d'Ecologie Théorique et Expérimentale, CNRS) ; Nicolas Fay (University of Western Australia).

Etat d'avancement : à venir – court, moyen et long termes

4.1.3.1. L'attention conjointe multimodale en situation écologique

Ce projet a pour objectif de tester l'hypothèse selon laquelle l'attention conjointe multimodale serait le maillon manquant permettant d'expliquer le passage des gestes intentionnels aux premiers mots parlés chez les enfants développant le langage. Pour mémoire, les enfants produisent leurs premiers gestes de pointage peu de temps avant leurs premiers mots, puis présentent des énoncés multimodaux dès le début de leur phase linguistique. Cependant, les mécanismes de réorganisation modale des gestes intentionnels aux mots intentionnels sont encore mal décrits, alors que l'expression multimodale de l'attention conjointe (ACM) n'a jamais été correctement quantifiée. L'objectif principal de cette ligne de recherche sera d'étudier l'ACM en tant qu'étape intermédiaire plausible de l'intégration multimodale soutenant la réorganisation gestuelle-vocale chez les enfants qui développent le langage. En d'autres termes, il s'agit d'investiguer un mécanisme du couplage intention – modalité communicative qui

permettrait de résoudre le problème de la transition/réorganisation des modalités à l'échelle ontogénétique.

Les corpus de vidéos des enfants français, Gamo, et mozambicains introduits ci-dessus seront utilisés pour mener cette étude. L'attention conjointe sera enregistrée chaque fois que la triangulation de l'attention devient manifeste par des regards et/ou des gestes déictiques et/ou un comportement vocal ou verbal de la mère ou de l'enfant. Il s'agira d'enregistrer si l'enfant initie ou répond et s'il parvient ou non à diriger l'attention de la mère vers la cible. L'attention visuelle de la mère sera enregistrée en début de séquence. Les modalités de communication utilisée pour les différentes phases de la séquence d'attention conjointe (e.g., adressage, direction de l'attention, nom de l'objet etc.) seront décrites et permettront de savoir si l'enfant favorise la modalité vocale quand la mère est inattentive. Cette étude permettra également de mettre en lien les modalités de l'attention conjointe avec le niveau de développement verbal de l'enfant et de savoir si l'attention conjointe multimodale (ACM) est un meilleur prédicteur du développement langagier que l'attention conjointe strictement visuelle. Cette étude devra permettre de tester les prédictions suivantes :

(1) Les enfants plus jeune initient davantage d'attention conjointe visuellement que vocalement,
(2) L'attention conjointe multimodale est plus fréquente quand la mère est visuellement inattentive,
(3) La proportion d'ACM augmente avec l'âge jusqu'à l'émergence de la parole,
(4) L'ajustement des modalités d'ACM à l'attention visuelle du partenaire est robuste aux variations éco-culturelles.

Cette première investigation de l'attention conjointe en situation écologique ouvrira des perspectives intéressantes pour tester certaines hypothèses expérimentalement. Cette étude pourra être menée dans le cadre d'une thèse sur l'attention conjointe multimodale. Elle sera réalisée en collaboration avec Mawa Dafreville (Université Toulouse Jean Jaurès), Lauren Bader (IAST) et Paul Vogt (Université de Tilburg) selon les corpus utilisés.

4.1.3.2. Corrélats neuronaux de l'attention conjointe multimodale

Les mécanismes intermodaux de l'ACM devraient jeter un nouvel éclairage sur la relation entre l'ACM et le développement de la parole. Au niveau neurocognitif, nous concevrons une tâche pour les enfants permettant de séparer l'initiation de l'ACM, qui possède un aspect *d'intention*

communicative, de *l'aspect référentiel* de l'ACM qui exige des compétences déictiques ou verbales de référence à un objet. Pour ce faire, nous construirons une tâche appropriée pour tester l'hypothèse selon laquelle l'activité cérébrale diffère entre les aspects d'adressage et de référence (par exemple, la vitesse de recrutement différentiel des régions unimodales et intermodales), ce qui donnerait des informations pertinentes sur les liens entre ACM et langage. La tâche sera conçue pour permettre l'examen séquentiel des mécanismes neuronaux de l'ACM de l'enfant (i) initiées par un adulte par contact visuel, ou en appelant le nom de l'enfant et (ii) suivies d'une attention visuelle vers un objet externe et/ou référence verbale à cet objet. L'ACM pourrait constituer une passerelle intermodale entre les aspects d'adressage déjà acquis par l'enfant vers les aspects référentiels en développement.

Cette ligne de recherche sera à mener en collaboration avec des collègues de neurosciences intéressés par l'attention visuelle et l'attention conjointe (collaborations envisagées avec Nathalie George (CNRS-ICM) ou Laurence Conty (Université Paris Nanterre)).

4.1.3.3. Sémiotique expérimentale des jeux de coordination

Ce projet utilise une tâche de tir à la corde où deux partenaires doivent tirer sur une corde au même rythme pour rapprocher un plateau de manière à ce que chaque partenaire puisse saisir une récompense. Dans une étude récente sur la coordination au sein de paires d'enfants d'âge scolaire, nous avons montré que les enfants coopéraient mieux avec d'autres enfants non-apparentés qu'avec leurs frères et sœurs (Barragan-Jason et al., 2021). La quantité de regards échangés était un comportement clef du succès, alors que les discours n'avaient pas d'effet sur le succès ou l'échec dans la tâche. Ces résultats suggèrent que la communication non verbale adressée au partenaire tient compte de la familiarité avec ce dernier. Deux études menées chez les chimpanzés ont déjà rapporté des influences similaires : en captivité, le chimpanzé Clint produisait plus de séquences multimodales face à un partenaire inconnu (Leavens et al., 1996) et dans la nature, les chimpanzés immatures présentent les comportements de communication intentionnelle envers des partenaires différents de la mère avant de les présenter envers leur mère (Fröhlich, Wittig, et al., 2019c). Les *efforts pour se faire comprendre* dans une situation de coordination se prête particulièrement bien à l'étude du couplage entre intention communicative et modalité de communication, de même qu'à l'émergence de la relation signe-référent sur des signes nouveaux (e.g., langages émergents) ou encore à la conventionnalisation

ou la négociation sociale de la signification d'un signe (Fröhlich et al., 2016; Pika & Fröhlich, 2019).

L'objectif sera d'étudier des enfants de 24 à 30 mois impliqués dans une tâche de coordination manuelle de tir à la corde. A cet âge, l'enfant développe son vocabulaire et associe les mots à des significations de manière spontanée. Il est possible que la tâche instrumentale génère des échanges de mots (connu d'un enfant mais pas de l'autre) et/ou que des relations signe-référent émergent en dehors du champ lexical de la langue maternelle. De plus, cette tâche « force » l'attention conjointe chez les enfants, ce qui nous permettra de quantifier les liens prédicteurs entre attention conjointe, visuelle et multimodale, et les productions langagières.

Cette étude pourra faire l'objet d'un encadrement de mémoire de master 2 ou d'une partie de projet de thèse. Elle sera réalisée en collaboration avec Gladys Barragan-Jason (Station d'Ecologie Théorique et Expérimentale, CNRS) et Sandra Molesti (Université Toulouse Jean Jaurès).

4.1.3.4. Grammaticalisation des signes visuels et auditifs d'un langage émergent

La même logique appliquée à un environnement virtuel impliquera deux adultes qui résoudront un jeu de coordination électronique en utilisant un petit ensemble de signes de propriétés physiques variables pour communiquer. Cette tâche de sémiotique expérimentale offre l'occasion de cartographier l'émergence de propriétés linguistiques avec des performances sociales de coordination entre deux acteurs. Dans cette expérience comme dans celle présentée ci-dessus, nous nous attendons à ce que la communication améliore la coordination et soit essentiellement visuelle à moins que le partenaire ne soit visuellement inattentif (i.e., lorsque des contraintes attentionnelles sont appliquées). Cette approche expérimentale permettra de mieux cerner les contraintes cognitives et comportementales qui affectent la grammaticalisation des signes.

Le jeu de coordination se joue à deux joueurs qui doivent déplacer un agent dans un espace virtuel dans le but manifeste de coordonner leurs mouvements (Attanasi et al., 2016). L'approche consiste à offrir aux joueurs la possibilité d'utiliser un ensemble restreint et fini de 'jetons' virtuels ayant des propriétés acoustiques similaires (par exemple, un son discontinu) et des propriétés visuelles différentes (par exemple, la forme d'un triangle, la couleur bleue) et

vice-versa. Les joueurs se verront présenter quatre conditions différentes successivement proposées : (1) sans jetons, (2) avec des jetons dans des conditions normales, (3) avec des jetons sans retransmission visuelle sur les tablettes mais avec retransmission auditive, (4) avec des jetons sans retransmission visuelle sur une des deux tablettes seulement. Le défi sémiotique repose sur la capacité des joueurs à développer des référents partagés (par exemple, à droite, à gauche) en utilisant des formes de communication partagées (par exemple, le triangle veut dire « à gauche »). Les conditions sont conçues pour inciter les joueurs à communiquer (1 contre 2), à utiliser les propriétés acoustiques du signe (2 contre 3) et à combiner les propriétés visuelles et acoustiques de manière tactique (3 contre 4). Sur la base des travaux précédents, nous nous attendons à ce que les joueurs créent d'abord des référents partagés en utilisant les propriétés visuelles des jetons (Fay, Lister, et al., 2014b). Nous nous attendons également à ce que les propriétés iconiques des signes visuels se conventionnalisent par le biais des répétitions. La combinaison des propriétés visuelles et acoustiques des jetons devrait accroître la complexité des significations.

Cette étude est envisagée dans le cadre d'un projet de thèse et pourra par la suite être adaptée à des enfants de 2 à 3 ans en train de développer le lexique et la syntaxe. Elle sera possiblement réalisée en collaboration avec Nicolas Fay (University of Western Australia), Fabio Montermini (Université Toulouse Jean Jaurès) et Paul Vogt (University of Tilburg).

4.2. Axe 2 : Linguistique animale à l'interface geste-voix

Le deuxième axe majeur de mon programme de recherche s'inscrit dans une dynamique exploratoire qui s'appuie sur les leviers suivants : (i) l'incarnation (*embodiement*) ; (ii) l'approche sémiotique ; (iii) l'usage d'un modèle animal ; (iv) la centration sur des contextes 'urgents' sur le plan évolutionnaire, c'est-à-dire susceptibles d'exercer des pressions de sélections significatives (e.g., les conflits et leurs différentes ritualisations). La conception incarnée du langage permet d'aborder la question de l'émergence des structures linguistiques sous un angle multimodal et multidimensionnel radical (Wilcox & Xavier, 2013). Elle s'inspire en grande partie des avancées réalisées dans l'étude des langues des signes auxquelles ont été appliquées les formalismes de la linguistique traditionnelle (Sandler & Lillo-Martin, 2006). Elle s'ancre également dans les découvertes neurocognitives du fonctionnement langagier présentant des connexions neuronales directes avec des tâches motrices complexes et planifiées comme la gestuelle accompagnant le langage (Iverson & Thelen, 1999). L'approche sémiotique

donne un cadre théorique plus large permettant de comparer des systèmes linguistiques différents sur un même plan. Plus particulièrement, la sémiotique expérimentale a montré avec beaucoup d'efficacité que des structures linguistiques simples apparaissaient dans des langages émergents (Galantucci & Garrod, 2011). La ligne d'investigation de ce second axe de recherche se trouve à la croisée de ces chemins et utilisera le cadre théorique large de la sémiotique pour mettre sur le même plan des unités gestuelles et vocales produites par le babouin olive dans différents contextes expérimentaux ou naturels. La communication animale offre des opportunités uniques de trouver des propriétés linguistiques en dehors du langage humain (Heesen et al., 2019; Semple et al., 2010) et l'application de méthodes linguistiques à la communication animale n'implique pas nécessairement la découverte de similarités langagières avec l'humain ; au contraire, l'usage de la même méthode a généré jusqu'ici des résultats biens différents dans les deux cas (Schlenker et al., 2017). Cet axe de recherche repose sur différents corpus de conflits et d'agression chez le babouin olive. L'agression est inhérente à la vie sociale et structure en grande partie les sociétés de primates par ses conséquences sociales directes : les relations de dominance-subordination, l'émergence d'une hiérarchie de dominance qui en découle et les nombreuses stratégies de ritualisation des conflits qui permettent de pacifier les relations, d'éviter les dommages physiques et de rendre prévisibles les interactions (Smuts et al., 2008; Strum, 2012).

4.2.1. Projet 2.1. Modélisation phonologique de la communication multimodale

Ce projet a pour objectif d'appliquer des formalismes issus de la phonologie à la communication visuelle et multimodale en partant d'un modèle animal (babouin olive). Des recherches comparatives sur les chants d'oiseaux et les vocalisations chez les primates ont fourni de bonnes preuves que ces systèmes fonctionnent selon le principe de la « double articulation » ; ils reposent en effet sur la combinaison d'unités discrètes non signifiantes en des ensembles signifiants (Bergman, 2013; Berwick et al., 2011; Schlenker, Chemla, Schel, Fuller, Gautier, Kuhn, Veselinović, Arnold, Cäsar, & Keenan, 2016). En revanche, les systèmes de communication visuels sont considérés, à tort, comme essentiellement iconiques, avec des relations non équivoques entre un geste et sa signification, et sont donc moins susceptibles d'être analysés en termes de double articulation. Ici, nous appliquerons la modélisation phonotactique telle que définie par la linguistique formelle à des séquences gestuelles de menace (corpus visuel 1) et de salutation (corpus multimodal 2) pour élucider les points suivants : (i) existe-t-il des propriétés combinatoires d'unités discrètes dans les séquences

gestuelles et multimodales ? et (ii) quelles règles phonotactiques régissent ces arrangements combinatoires, i.e., existe-t-il une 'syntaxe phonologique' ?

Objectif : Appliquer des formalismes linguistiques à la communication du babouin olive, dans une perspective multimodale permettant d'établir de nouvelles passerelles entre gestes et voix et d'accroitre la compréhension de la complexité sociale et communicative de l'espèce.

Méthode : Expérimentale ; Observationnelle ; Application de formalismes linguistiques.

Collaborateurs principaux : Fabio Montermini & Basilio Calderone (Université Toulouse Jean Jaurès) ; Gillian Forrester (Université de Birbeck, Londres) ; Shirley Strum (Université de Californie, San Diego).

Financements : Financement de la Maison des Sciences de l'Homme et de la Société de Toulouse ; Institut Universitaire de France.

Etat d'avancement : en cours et jusqu'à 5 ans ; perspectives à moyen terme.

4.2.1.1. Unités motrices des structures linguistiques

Les séquences de menace qui font l'objet du corpus utilisé pour ce projet combinent une expression orofaciale, i.e., *eyebrow raise*, avec un geste manuel, i.e., *hand slap*, permettant d'établir un lien entre geste et voix. Si les expressions faciales de menace ne sont pas souvent vocalisées par les babouins, elles s'accompagnent d'une ouverture de la bouche qui peut parfois laisser passer un soupir audible et fermement expiré, et parfois des *grunts* spécifiques, i.e., *pant grunts*. Depuis Darwin, les expressions orofaciales sont considérées comme des manifestations émotionnelles inflexibles et héritées phylogénétiquement (Davidson et al., 2004; Gainotti, 2012; Wallez et al., 2019). Elles impliquent les muscles du visage, de la bouche et des lèvres pour produire des mouvements d'intensité variable, comportant un pic d'expression. Le contrôle neural des expressions faciales émotionnelles est considéré comme étant différent de celui du langage ou de la communication intentionnelle sous contrôle de l'hémisphère gauche (Davidson et al., 2004; Gainotti, 2012). Des études sur les asymétries orofaciales menées chez les babouins ont pourtant généré des résultats contradictoires quant à la dominance hémisphérique droite des expressions orofaciales de menace (Wallez et al., 2019; Wallez & Vauclair, 2011, 2013). Il est important de souligner qu'une menace n'est pas un comportement mécaniquement effectif et son interprétation repose donc sur une convention partagée. Bien que phylogénétiquement héritée, cette convention pourrait donner davantage de flexibilité à son producteur qu'une agression plus effective. En d'autres termes, la ritualisation de l'agression

incorporée dans une *menace* pourrait réduire l'intensité émotionnelle et favoriser le contrôle volontaire chez le signaleur.

Ce projet en cours de réalisation avait initialement pour objectif la description multidimensionnelle de séquences de menace impliquant la main et le visage afin (i) d'étudier le contrôle hémisphérique de l'expression orofaciale de menace en fonction de la production du geste associé ou de son absence, (ii) d'étudier la coordination entre ces deux éléments clefs du langage, (iii) de décrire l'organisation interne des séquences en faisant émerger leurs contraintes de type phonotactiques. Les séquences de menace ont été collectées expérimentalement par des humains s'approchant d'un babouin adulte dans sa cage et s'engageant dans une des 4 conditions expérimentales ci-après : (A) l'humain menace le babouin en reproduisant l'expression faciale de menace, i.e., *eyebrow raise*, (B) l'humain menace le babouin en reproduisant l'expression faciale de menace et le geste de *hand slap*, (C) l'humain tend une banane au babouin et lui retire promptement dès que le babouin s'en saisit en réalisant une expression faciale de menace et (D) l'humain tend une banane au babouin tout en regardant de côté et lui retire promptement la banane lorsqu'il s'en saisit, le regard et le visage toujours détournés. La différence entre C et D repose sur l'attitude de l'expérimentateur, menaçant intentionnellement (C) ou inattentif à la situation et donc involontairement maladroit (D). Les conditions ont été réalisées en série AB jusqu'à obtenir une première réponse d'agression, puis ACDC sur les sujets ayant répondu à AB, dans le but de tester le contrôle volontaire de l'individu sur ses signaux de menace. Nous attendions une baisse, voire une absence de signaux dans la condition D et une recrudescence dans la deuxième condition C.

Les premiers résultats montrent que l'*eyebrow raise* est sous contrôle bilatéral quand on considère la latence d'activation de chaque paupière (Wallez & Bourjade, non publié). De plus, les babouins répondent par une menace dans la condition C alors qu'ils ne réagissent presque pas dans la condition D (Figure 4), ou inhibent leur réaction dès qu'ils regardent vers l'humain inattentif (Bourjade et al., *en préparation*). Les séquences de la condition D ne comportant que peu de signaux ont été retirées de l'analyse suivante.

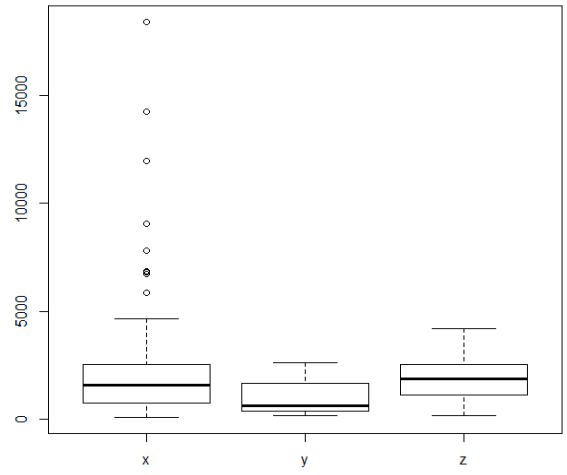

Figure 4. Durée des séquences de menace (millisecondes). x : première condition C ; y : condition D ; z : deuxième condition C.

4.2.1.2. Modélisation phonotactique des corpus de menace et de salutation

L'objectif est d'analyser la structure des séquences gestuelles de menace de babouins et de les décomposer en *unités gestuelles* les plus petites qui soient afin de déterminer si elles sont régies par des principes comparables aux règles phonologiques (Bourjade et al., *en préparation*). La métaphore linguistique employée ici est de considérer chaque séquence comme un *mot* dont on cherche à étudier l'enchainement des unités comme des *phonèmes*.

Corpus 1. Les réactions de menace ont été expérimentalement induites avec la méthode décrite ci-dessus. Les séquences de menace ont ensuite été codées à l'aide de la méthode multidimensionnelle (Forrester, 2008), qui code chaque action sur la base de ses plus petites unités (non sécables) en fonction de critères moteurs et anatomiques. Le corpus résultant est composé de 281 séquences exploitables échantillonnées sur 31 babouins adultes. Huit unités gestuelles ont ainsi été enregistrées concernant les membres, les mains, la bouche, les sourcils et le pelage des babouins.

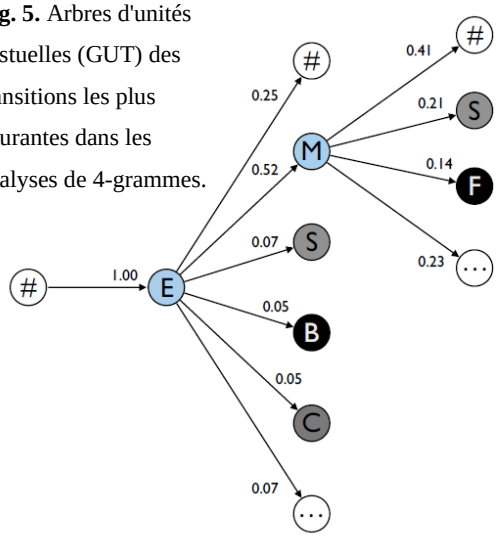

Fig. 5. Arbres d'unités gestuelles (GUT) des transitions les plus courantes dans les analyses de 4-grammes.

Corpus 2. La même approche sera appliquée aux séquences de salutation comportant des composants acoustiques (e.g., 'lipsmacking'). Le corpus sera constitué en enregistrant sur vidéo les salutations de babouins sauvages de la population du plateau de Laikipia au Kenya (en collaboration avec Shirley Strum). Les séquences de salutation seront codées à l'aide de la méthode multidimensionnelle (Forrester, 2008) sans inférence sur la 'sémantique'.

Les données seront analysées en phonologie segmentale, dans le cadre de la théorie des caractéristiques distinctives (Chapman, 2009) et porteront donc sur les caractéristiques des unités gestuelles qui favorisent l'émergence de segments distincts au sein de séquences. Afin de proposer une modélisation de la combinaison d'unités gestuelles en séquences plus grandes, nous allons segmenter les séquences de communication en n-grammes[67] (jusqu'à 4 grammes)

[67] La méthode des *n*-grammes consistent à segmenter les séquences en *n* unités, puis d'en dénombrer la fréquence afin d'étudier l'organisation des unités non signifiantes (phonèmes) dans les ensembles plus larges (mots). Un exemple de phonotaxe du français : le 2 -gramme /p/, /a/ en début de mot est représenté 550 fois dans le lexique (e.g., paradis, parapluie…).

d'unités gestuelles et acoustiques et calculer la probabilité de transition entre celles-ci (Bourjade, Calderone, et al., 2019) (Figure 5). Ensuite, nous identifierons un ensemble de caractéristiques pouvant être utilisées pour décrire toutes les unités d'action des babouins et effectuer des analyses segmentales. Nous nous attendons à ce que les séquences de menace de babouins présentent des propriétés combinatoires au sein de la modalité visuelle. Ce serait la première preuve de combinatorialité dans la communication gestuelle en dehors des langues des signes. Cela jettera donc un nouvel éclairage sur les précurseurs des liens oro-visuels caractérisant le langage multimodal (Kendon, 2011; MacNeilage, 1998).

Ce projet de recherche en cours réalisé en collaboration avec Gillian Forrester (Université de Birbeck), Fabio Montermini et Basilio Calderone (Université Toulouse Jean Jaurès) pourra donner lieu à un travail postdoctoral ou de master 2 afin de mener plus en avant l'analyse segmentale et de tester différents modèles d'organisation internes des séquences (Kershenbaum et al., 2014). Il pourra déboucher sur un travail doctoral portant notamment sur le second corpus.

4.2.2. Projet 2.2. Sémantique et pragmatique des conflits sociaux

> *"Baboons don't use or make material tools in the wild but they do skillfully craft friendships. These social tools are 'non-material artifacts' because making them requires the same cognitive achievement as making material artifacts: their shape and use isn't intuitively obvious on the surface of the material."*
>
> *Shirley Strum (2012, p.17)*

Ce second projet de « linguistique animale » s'intéresse à la pragmatique des conflits sociaux dans la société très complexe du babouin olive. Il se situe à la croisée de la linguistique évolutionnaire et de l'étude de la complexité sociale telle que portée par les sciences sociales. Sur le plan évolutionnaire, le mensonge et la tromperie sont supposés être écartés ou tenus à fréquence minimale par différents mécanismes de régulation sociale et de coût-bénéfice imposés aux tricheurs (Maynard-Smith & Harper, 1995; Scott-Phillips, 2008). Par ailleurs, la possibilité de mentir ou d'inventer est une caractéristique du langage humain déniée aux non-humains dans la proposition de Hockett (1960). Or, les coalitions formées par les primates pour soutenir un individu dans un conflit peuvent être manipulées à des fins stratégiques et s'appuient en partie sur des épisodes de faux-semblant (Strum, 2012). Dans cette perspective, attribuer une signification à un signal communicatif de manière flexible, c'est-à-dire en tenant compte du contexte, pourrait raisonnablement procurer des avantages adaptatifs aux individus les plus

compétents en la matière. Sur le plan de la complexité sociale, il est important de souligner que la signification d'un signal communicatif et sa manipulation se trouve 'en dehors' de l'individu ; en réalité, l'interprétation appartient au domaine social et dépasse largement les compétences et la volonté de l'individu (Servais, 2012). Cette approche de la pragmatique est cohérente avec la vision de la communication intentionnelle défendue dans cette recherche et converge avec la nécessité de prendre en compte le contexte pour capturer les mécanismes complexes régissant la vie sociale du babouin olive (Strum, 2012).

Deux grandes étapes sont envisagées dans ce projet. La première s'intéressera à la *signification* de chaque signal, évaluée avec la méthode de l'ASO (*Apparently satisfaying outcome*) (Hobaiter & Byrne, 2014, 2017) et permettra d'établir une signification lexicale répertoriée dans un 'dictionnaire'. La deuxième étape établira la *flexibilité fonctionnelle* des signaux communicatifs et permettra de quantifier l'usage des signaux de communication dans différents contextes de la vie sociale, puis examinera la *distorsion* de la signification lexicale en fonction du contexte. Différentes règles d'inférences sur les significations concurrentes pourront être examinées.

Ce projet pourra faire l'objet d'une thèse en collaboration avec Shirley Strum (Université de Californie, San Diego) et Fabio Montermini (Université Toulouse Jean Jaurès). D'autres collaborations peuvent être envisagées avec l'Institut Jean Nicod ou de jeunes chercheurs intéressés par cette thématique « montante ».

Objectif : Etudier l'émergence de structures phonologiques combinatoires et d'une flexibilité sémantique en passant du niveau 'lexical' au niveau 'pragmatique' dans la communication du babouin olive.

Méthode Observationnelle ; Application de formalismes linguistiques.

Collaborateurs principaux : Fabio Montermini (Université Toulouse Jean Jaurès) ; Shirley Strum (Université de Californie, San Diego).

Etat d'avancement : pilotes en cours ; perspectives à moyen et long termes.

4.2.2.1. Le dictionnaire du geste et de la voix chez le babouin

Cette première étude a pour objectif d'établir la signification 'lexicale' des signaux gestuels et vocaux émis par des babouins olive dans leur vie sociale naturelle. Elle s'appuie sur une méthode d'observation établie par Richard Byrne et ses collaborateurs (Cartmill & Byrne, 2010; Hobaiter & Byrne, 2014, 2017) visant à documenter les différentes conséquences de l'émission

d'un signal sur les receveurs qui semblent satisfaire le signaleur (ce dernier cesse alors de communiquer) ; i.e., méthode de l'ASO. Pour mener à bien cette analyse, il s'agira de réaliser des observations dans tous les contextes de la vie sociale et de noter pour chaque signal, les effets sur le receveur, et si le signaleur cesse de signaler ou persévère. Cette méthode appliquée aux gestes des chimpanzés a permis d'identifier 19 ASOs révélant les intentions communicatives du signaleurs (e.g., démarrer une interaction, créer un changement dans l'interaction, ou l'interrompre). Certains gestes avaient plus d'une signification, mais la plus grande part de flexibilité venait plutôt du fait que différents gestes étaient utilisés pour signifier la même chose (Hobaiter & Byrne, 2014). Il est important de souligner que le lexique généré par cette approche est indépendant de l'identité du signaleur et du receveur et est aveugle au contexte social. Cette étape est nécessaire afin d'investiguer la variation des interprétations d'un même signal faites par les receveurs en fonction du contexte.

4.2.2.2. La pragmatique du conflit social ou comment négocier de l'aide

La deuxième étude portera précisément sur la question de l'interprétation du signal faite par le receveur. En effet, dans une séquence complexe où un mâle babouin tente de pousser son amie femelle babouin à distance d'une troupe étrangère, les approches du mâle vers la femelle ne sont pas interprétées de la même manière par cette dernière que dans le contexte des interactions sexuelles ou d'épouillage. Pourtant, il s'agit du même comportement. Cette étude s'intéressera à ces variations contextuelles en se centrant sur les conflits sociaux et leurs contextes (très) variés. Dans une première analyse quantitative, il s'agira de dénombrer la prévalence de chaque signal dans tous les contextes observés. Dans une seconde analyse, alliant méthode qualitative et quantitative, il s'agira de documenter la variabilité des réponses données à chaque signal. Les contextes considérés seront (i) l'écologie (e.g., prédateur, autre espèce sauvage, point d'eau), (ii) la dynamique sociale de la troupe (e.g., changement de direction du mouvement, départ matinal, nouveau mâle immigrant, rencontre d'une troupe étrangère etc.), et (iii) les contextes sociaux immédiats (e.g., interactions entre mâles ou entre mâle et enfant, attention portée à un nouveau-né, consort).

La première étude pilote portera sur la variabilité observée (à l'échelle de la production des signaux et de leur interprétation) entre le jeu social et les conflits. Des juvéniles en train de jouer seront observés en suivi focal et si le jeu se transforme en conflit, les mêmes interactions seront notées dans le conflit, en documentant le contexte. Cette étude sera menée en collaboration avec

Shirley Strum et les observateurs du Projet Uaso Ngiro (Kenya). Il pourra, si les conditions le permettent donner lieu à un projet de thèse.

4.3. De la mathématisation au translationnel : perspectives à long terme

Au terme de cette présentation relativement exhaustive de mes projets de recherche sur la communication multimodale dans une perspective évolutive et comparée, je souhaite présenter quelques développements possibles auxquels ces recherches pourraient mener. Au cours de mes interactions scientifiques avec des collègues biologistes, psychologues, anthropologues, ou encore médecins, philosophes ou linguistes intéressés par l'évolution en général, j'ai pu constater que l'intérêt vif que nous portions aux processus évolutifs ne se traduisaient souvent pas dans les applications de la recherche, en particulier dans le domaine de la *santé*, ni dans celui de l'*éducation*, ni encore dans celui, plus controversé, des « *guerres culturelles* » relevant de conceptions opposées d'ordre moral et influençant directement les politiques publiques. Il me semble que deux faiblesses sont possiblement responsables de cette absence. L'intérêt pour l'évolution ne s'accompagne pas toujours des efforts nécessaires à sa théorisation, et en particulier, à sa mathématisation. Les scénarios évolutifs sont toujours hypothétiques mais travailler à la théorisation des hypothèses et à leurs probabilités statistiques permet de soulever des questions importantes responsables d'avancées scientifiques significatives. La deuxième faiblesse survient à l'autre extrémité de la chaine de la connaissance ; les évolutionnistes sont absents des études appliquées à la santé ou à l'éducation. A juste titre, les études formulées dans une perspective évolutionnaire sont presque absentes en Europe sur ces sujets. Pourtant, des approches mixtes permettant de faire dialoguer les sciences de l'évolution et les sciences sociales existent à la marge et semblent prometteuses (voir la proposition sur la dépression post-partum portée par Lauren Bader, section 4.3.2.1.). Ci-dessous, je présente très succinctement ces deux lignes d'investigation qui prolongent la réflexion et le questionnement proposés dans cette note de synthèse.

4.3.1. Mathématisation de l'évolution du langage

« La vraie grandeur de la science n'est pas tant de nommer et de classifier, d'enregistrer et de prévoir, que d'observer et de désirer connaitre les faits, où que cela paraisse mener. Plus grand est le scientifique, plus il est impressionné par son ignorance de la réalité et plus il conçoit que ses lois et étiquettes […] sont le produit de sa propre pensée. »

Alan Watts, (2005, p.164)

Pour progresser dans notre compréhension de l'évolution du langage, nous devons tester les scénarios évolutifs à l'aide de modèles computationnels. Cependant, ces derniers exigent qu'un certain type de données empiriques soient disponibles - de vrais « zéros » indiquant qu'un trait est absent chez une espèce par exemple – ce qui nécessite d'orienter la collecte de données chez les chercheurs travaillant avec l'approche comparative. Ainsi, j'envisage de travailler en étroite collaboration avec des biologistes de l'évolution (collaboration envisagée par exemple avec Valentin Thouzeau) afin de construire les premières matrices phylogénétiques de données langagières. Deux types de données linguistiques sont à ce jour assez bien documentés chez les primates : (1) la prise de parole à tour de rôle dans la conversation (*turn-taking*) (Levinson, 2016) et (2) l'intentionnalité de la communication (Fröhlich, Sievers, et al., 2019b; Townsend et al., 2017). Construire des matrices comparatives phylogénétiques nécessite de s'appuyer sur des conceptions théoriques concernant la nature des caractères. Par exemple, si l'intentionnalité est un caractère discret ou continu ou s'il est également probable de la gagner ou de la perdre au cours de l'évolution (Harmon, 2018). Pour répondre à ces questions, nous examinerons la littérature existante sur le sujet afin de compléter une première matrice comparative qui servira à modéliser des scénarios de continuité et de discontinuité des caractères au sein du phylum des primates. Cela permettra par voie de conséquence d'attirer l'attention des scientifiques sur les données manquantes et d'orienter la collecte de données à moyen et long termes.

Cette ligne de recherche s'apparente à une méta-analyse permettant de raffiner les concepts théoriques et de les formuler dans le langage matriciel d'une base de données permettant le calcul probabiliste des différents scénarios évolutifs des origines du langage. Dans la même lignée, une base de données partagée serait nécessaire pour identifier les critères pertinents dans l'évolution du langage et orienter les efforts d'échantillonnage à l'échelle mondiale.

4.3.2. L'économie de l'attention dans le domaine de la santé

« Penser, ce n'est pas unifier, rendre familière l'apparence sous le visage d'un grand principe. Penser, c'est réapprendre à voir, diriger sa conscience, faire de chaque image un lieu privilégié. Autrement dit, la phénoménologie se refuse à expliquer le monde, elle veut être seulement une description du vécu. »

Albert Camus, (1986, p. 65)

Bien que les travaux présentés jusqu'ici aient examiné la communication dans des perspectives évolutionnaires et comparatives, il est indéniable que les études sur les populations humaines génèrent des connaissances fondamentales susceptibles de générer des changements pratiques. En particulier, les travaux portant sur les ruptures attentionnelles, et plus généralement sur *l'attention* - au sens étroit de l'attention visuelle, jusqu'au sens plus large de l'intérêt porté à autrui – dans les interactions entre mère et enfant, entre parents et enfants et entre enfants peuvent donner lieu à des développements dans le domaine de la santé. Parmi ceux-ci, plusieurs projets ont déjà été discutés à différentes occasions avec des médecins ou des collègues impliqués dans des recherches plus appliquées. Les processus d'attention visuelle et le relais que peut prendre la modalité vocale est par exemple intéressante à appliquer à l'étude de *l'autisme*, du fait que les personnes atteintes de troubles du spectre de l'autisme présentent souvent des déficiences sur le plan de l'attention sociale (collaboration possible avec Lucie Bouvet, Université Toulouse Jean Jaurès). Un autre domaine de la prise en charge médicale concernée par l'attention parent-enfant est le domaine de *la prématurité* où les nouveau-nés présentent des réponses physiologiques et comportementales spécifiques aux stimulations tactiles ou visuelles des soignants et des parents (collaboration possible avec Jacques Sizun, CHU Toulouse). Du côté du parent, deux champs d'application attirent particulièrement mon attention du fait de leur prévalence dans la société et de leur importance dans le développement de l'enfant. L'un concerne la défaillance de l'attention maternelle dans les dépressions post-partum et ses impacts sur le développement socioaffectif et langagier du bébé (projet en collaboration avec Lauren Bader, IAST Toulouse, présenté ci-dessous). L'autre concerne l'utilisation 'thérapeutique' de l'attention dans son acception large pour soigner. Spécifiquement, j'aimerais étudier le rôle de l'attention dans les effets dits 'placebo' observés dans les tests cliniques de traitement médicaux. Dans les deux cas, il s'agira de développer une approche 'darwinienne' à ces altérations de l'attention maternelle, parentale, ou thérapeutique.

4.3.2.1. La dépression post-partum sous l'angle évolutionnaire

La dépression représente 40% des maladies mentales et la cause première ou seconde d'invalidité dans les pays émergents et en voie de développement (*low- and middle-income countries*) (Syme & Hagen, 2020). Les femmes en âge de procréer ont un risque accru de dépression, notamment d'épisode de dépression post-partum (DPP) et peu de recherches sont menées sur cette question en dehors des pays occidentaux. Dans les pays occidentaux, les symptômes de DPP sont associés à des modifications du comportement maternel et du développement socioaffectif et cognitif de l'enfant (e.g., Gravener et al., 2012). L'hypothèse médicale dominante est une hypothèse physiologique permettant d'expliquer la DPP par la diminution d'une monoamine dans le corps. Cela ne prend en compte ni l'environnement social, ni l'influence des croyances culturelles, ni les possibles fonctions adaptatives de la DPP. Or, les approches intégratives de la parentalité, articulant des approches ethnographiques qualitatives et des études comportementales quantitatives ont permis de mettre en évidence que les mères Gamo qui se sentaient stressées par les pleurs de leur bébé avaient globalement moins d'interactions directes avec leur enfant que les mères qui ne rapportaient pas ce stress, et que les croyances des mères à propos de la signification des pleurs de leur enfant étaient associées à l'état de santé de l'enfant (Bader et al., 2019; Bader & Fouts, 2018, 2019). Il est donc important de tenir compte des croyances culturelles dans la DPP. De même, les modèles évolutionnaires de la DPP proposent qu'elle ait une fonction de signalement qui serait adaptative. Plus précisément, le modèle suggère que la mère expérimente une souffrance psychique lorsqu'elle n'est pas parfaitement à-même de s'occuper de son enfant (Hagen, 2002, 2003). Les symptômes fonctionneraient comme des signalements adressés à l'entourage pour encourager un soutien additionnel.

Ce projet en cours d'élaboration (deux demandes de financement déposées) porté par Lauren Bader (IAST, Toulouse) pourrait mener sur un travail doctoral en co-encadrement. L'objectif sera de tester le modèle évolutionnaire de la dépression post-partum en tenant compte du modèle culturel de parentalité et des croyances individuelles relatives à l'éducation d'un enfant. Le deuxième objectif sera d'évaluer les modifications attentionnelles de la mère associées aux symptômes de DPP. Enfin, il s'agira d'évaluer les modifications du développement communicatif de l'enfant selon les symptômes de DPP de la mère.

4.3.3. Conclusion

J’ai rédigé cette note de synthèse en pensant aux étudiants qui seraient susceptibles de la lire en débutant sous ma direction un mémoire de master ou une thèse sur l’un de ces sujets. J’ai accepté de rédiger cette note de synthèse et de m’engager à obtenir l’habilitation à diriger des recherches à partir du moment où des étudiants désireux de s’engager dans la recherche scientifique se sont adressés à moi. Mon envie de le faire s’est encore accru quand de jeunes collègues en postdoctorat ou en poste ont manifesté leur désir de co-encadrer un travail de master ou une thèse avec moi. Les directions de recherche présentées ci-dessus sont des sources d’inspiration et parfois d’avancées dans le questionnement scientifique que je mène sur la communication et l’attention. Le dernier chapitre présentant mon programme de recherche doit être interprété comme des propositions, et le possible ancrage de sujets de master, de thèse ou de projets en collaboration. Aucun étudiant ni aucun collègue n’est tenu d’adhérer à ces lignes de recherches à la lettre, ni de s’emparer de ces idées pour travailler avec moi. Les projets laissés de côté laissent généralement place à de beaux projets qui existent.

Chapitre 5. Bibliographie

Ainsworth, M. (1963). The development of infant-mother interaction among the Ganda. *Determinants of infant behavior*, 67-112.

Ainsworth, M., Blehar, M. C., Waters, E., & Wall, S. N. (2015). *Patterns of attachment : A psychological study of the strange situation*. Psychology Press.

Aldridge, M. A., Braga, E. S., Walton, G. E., & Bower, T. G. R. (1999). The intermodal representation of speech in newborns. *Developmental Science*, *2*(1), 42-46.

Als, H., Tronick, E., Lester, B. M., & Brazelton, T. B. (1977). The Brazelton neonatal behavioral assessment scale (BNBAS). *Journal of abnormal child psychology*, *5*(3), 215-229.

Altmann, J. (1974). Observational study of behavior : Sampling methods. *Behaviour*, *49*(3-4), 227-266.

Apicella, C. L., & Barrett, H. C. (2016). Cross-cultural evolutionary psychology. *Current Opinion in Psychology*, *7*, 92-97.

Apicella, C. L., Feinberg, D. R., & Marlowe, F. W. (2007). Voice pitch predicts reproductive success in male hunter-gatherers. *Biology letters*, *3*(6), 682-684.

Arbib, M. A. (2005). From monkey-like action recognition to human language : An evolutionary framework for neurolinguistics. *Behavioral and Brain Sciences*, *28*(2), 105-124. https://doi.org/10.1017/S0140525X05000038

Arnold, K., & Zuberbühler, K. (2008). Meaningful call combinations in a non-human primate. *Current Biology*, *18*(5), R202-R203. https://doi.org/10.1016/j.cub.2008.01.040

Atance, C. M., & O'Neill, D. K. (2001). Episodic future thinking. *Trends in cognitive sciences*, *5*(12), 533-539.

Attanasi, G., Hopfensitz, A., Lorini, E., & Moisan, F. (2016). Social connectedness improves co-ordination on individually costly, efficient outcomes. *European Economic Review*, *90*, 86-106.

Aureli, T., Spinelli, M., Fasolo, M., Garito, M. C., Perucchini, P., & D'Odorico, L. (2017). The Pointing–Vocal Coupling Progression in the First Half of the Second Year of Life. *Infancy*, *22*(6), 801-818.

Austin, J. L. (1962). *Speech acts*. Oxford.

Aychet, J., Blois-Heulin, C., & Lemasson, A. (2021). Sequential and network analyses to describe multiple signal use in captive mangabeys. *Animal Behaviour, 182,* 203-226.

Aychet, J., Blois-Heulin, C., Palagi, E., & Lemasson, A. (2021). Facial displays in red-capped mangabeys (Cercocebus torquatus) : Repertoire, social context, and potential intentionality. *Journal of Comparative Psychology*, *135*(1), 98.

Aychet, J., Pezzino, P., Rossard, A., Bec, P., Blois-Heulin, C., & Lemasson, A. (2020). Red-capped mangabeys (Cercocebus torquatus) adapt their interspecific gestural communication to the recipient's behaviour. *Scientific reports*, *10*(1), 1-12.

Bader, L. R., & Fouts, H. N. (2018). Cultural models of infant emotions and needs among the Gamo people of Southern Ethiopia. *Infant Mental Health Journal, 39*(5), 497-510. https://doi.org/10.1002/imhj.21734

Bader, L. R., & Fouts, H. N. (2019). Parents' perceptions about infant emotions : A narrative cross-disciplinary systematic literature review. *Developmental Review, 51*, 1-30. https://doi.org/10.1016/j.dr.2018.11.003

Bader, L. R., Fouts, H. N., & Jaekel, J. (2019). Mothers' feelings about infants' negative emotions and mother-infant interactions among the Gamo of Southern Ethiopia. *Infant Behavior and Development, 54*, 22-36. https://doi.org/10.1016/j.infbeh.2018.09.005

Bakeman, R., & Adamson, L. B. (1984). Coordinating attention to people and objects in mother-infant and peer-infant interaction. *Child development*, 1278-1289.

Bard, K. A. (1992). Intentional Behavior and Intentional Communication in Young Free-Ranging Orangutans. *Child Development, 63*(5), 1186-1197. https://doi.org/10.1111/j.1467-8624.1992.tb01688.x

Bard, K. A. (2019). Parenting in nonhuman primates. In M. Bornstein (ed), *Handbook of parenting* (p. 78-122). Routledge.

Bard, K. A., Bakeman, R., Boysen, S. T., & Leavens, D. A. (2014). Emotional engagements predict and enhance social cognition in young chimpanzees. *Developmental science*, *17*(5), 682-696.

Bard, K. A., & Leavens, D. A. (2014). The importance of development for comparative primatology. *Annual Review of Anthropology*, *43*(1), 183.

Barragan-Jason, G., Atance, C., Kopp, L., & Hopfensitz, A. (2018). Two facets of patience in young children : Waiting with and without an explicit reward. *Journal of experimental child psychology*, *171*, 14-30.

Barragan-Jason, G., Cauchoix, M., Regnier, A., Bourjade, M., Hopfensitz, A., & Chaine, A. S. (2021). Schoolchildren cooperate more successfully with non-kin than with siblings. *Proceedings of the Royal Society B, 288*(1944), 20202951.

Bates, E., Camaioni, L., & Volterra, V. (1975). The acquisition of performatives prior to speech. *Merrill-Palmer Quarterly, 21*(3), 205-226.

Beebe, B., Messinger, D., Bahrick, L. E., Margolis, A., Buck, K. A., & Chen, H. (2016). A systems view of mother–infant face-to-face communication. *Developmental Psychology, 52*(4), 556.

Bell, M. A. (2020). Mother-child behavioral and physiological synchrony. In *Advances in child development and behavior* (Vol. 58, p. 163-188). Elsevier.

Bergman, T. J. (2013). Speech-like vocalized lip-smacking in geladas. *Current Biology*, *23*(7), R268-R269. https://doi.org/10.1016/j.cub.2013.02.038

Bergman, T. J., Beehner, J. C., Painter, M. C., & Gustison, M. L. (2019). The speech-like properties of nonhuman primate vocalizations. *Animal Behaviour, 151*, 229-237.

Bernicot, J. (1994). Speech acts in young children : Vygotsky's contribution. *European Journal of Psychology of Education, 9*(4), 311-319.

Berwick, R. C., Okanoya, K., Beckers, G. J., & Bolhuis, J. J. (2011). Songs to syntax : The linguistics of birdsong. *Trends in cognitive sciences, 15*(3), 113-121.

Beuker, K. T., Rommelse, N. N., Donders, R., & Buitelaar, J. K. (2013). Development of early communication skills in the first two years of life. *Infant Behavior and Development, 36*(1), 71-83.

Bianchi, S., Reyes, L. D., Hopkins, W. D., Taglialatela, J. P., & Sherwood, C. C. (2016). Neocortical grey matter distribution underlying voluntary, flexible vocalizations in chimpanzees. *Scientific reports, 6*(1), 1-7.

Bickerton, D. (2007). Language evolution : A brief guide for linguists. *Lingua, 117*(3), 510-526.

Bigelow, A. E., & Rochat, P. (2006). Two-month-old infants' sensitivity to social contingency in mother–infant and stranger–infant interaction. *Infancy, 9*(3), 313-325.

Böckler, A., van der Wel, R. P., & Welsh, T. N. (2014). Catching eyes : Effects of social and nonsocial cues on attention capture. *Psychological science, 25*(3), 720-727.

Boë, L.-J., Berthommier, F., Legou, T., Captier, G., Kemp, C., Sawallis, T. R., Becker, Y., Rey, A., & Fagot, J. (2017). Evidence of a vocalic proto-system in the baboon (Papio papio) suggests pre-hominin speech precursors. *PloS one, 12*(1), e0169321.

Bogin, B. (1997). Evolutionary hypotheses for human childhood. *American Journal of Physical Anthropology, 104*(25), 63-89.

Bornstein, M. H., Putnick, D. L., Park, Y., Suwalsky, J. T., & Haynes, O. M. (2017). Human infancy and parenting in global perspective : Specificity. *Proceedings of the Royal Society B: Biological Sciences, 284*(1869), 20172168.

Bottineau, D. (2018). Du *languaging* au sens linguistique. *Intellectica, 2017/2*(68), 19-68.

Bouchet, H., Koda, H., & Lemasson, A. (2017). Age-dependent change in attention paid to vocal exchange rules in Japanese macaques. *Animal Behaviour, 129*, 81-92.

Bourjade, M. (2007). *Sociogeny and manifestation of individual and collective behaviours in horses*. [PhD thesis]. Université de Strasbourg.

Bourjade, M. (2016). Le pointage controversé des singes : Éléments empiriques chez le babouin olive (Papio anubis). *Enfance, 4*, 375-404.

Bourjade, M. (2017). Social Attention. In A. Fuentes (ed), *The International Encyclopedia of Primatology*. John Wiley & Sons, Inc. https://doi.org/10.1002/9781119179313.wbprim0356

Bourjade, M. (2019). *Disruption of social attention and the tactical use of sensory modalities in primate infant signaling*. Protolang 6, Libon.

Bourjade, M., Calderone, B., & Montermini, F. (2019). *Phonotactic modeling of baboon gestural threat sequences*. Animal Linguistics: Take the leap!, Ecole Normale Supérieure, Paris.

Bourjade, M., Call, J., Pelé, M., Maumy, M., & Dufour, V. (2014). Bonobos and orangutans, but not chimpanzees, flexibly plan for the future in a token-exchange task. *Animal cognition*, *17*(6), 1329-1340.

Bourjade, M., Canteloup, C., Meguerditchian, A., Vauclair, J., & Gaunet, F. (2015). Training experience in gestures affects the display of social gaze in baboons' communication with a human. *Animal cognition*, *18*(1), 239-250.

Bourjade, M., Cochet, H., Molesti, S., & Guidetti, M. (2020). Is conceptual diversity an advantage for scientific inquiry? A case study on the concept of 'gesture' in comparative psychology. *Integrative Psychological and Behavioral Science*, 1-28.

Bourjade, M., Dafreville, M., Scola, C., & Jover, M. (soumis). *Avoidance of mutual speech structures mother-infant interactions at six months*. Université Toulouse Jean Jaurès

Bourjade, M., de Boyer des Roches, A., & Hausberger, M. (2009). Adult-young ratio, a major factor regulating social behaviour of young: A horse study. *PLoS ONE*, *4*(3), e4888. https://doi.org/10.1371/journal.pone.0004888

Bourjade, M., Gaunet, F., Maugard, A., & Meguerditchian, A. (2019). Manipulating social cues in baboon gesture learning : What does it tell us about the evolution of communication? *Animal cognition*, *22*(1), 113-125.

Bourjade, M., Meguerditchian, A., Maille, A., Gaunet, F., & Vauclair, J. (2014). Olive baboons, Papio anubis, adjust their visual and auditory intentional gestures to the visual attention of others. *Animal Behaviour*, *87*, 121-128. https://doi.org/10.1016/j.anbehav.2013.10.019

Bourjade, M., Meunier, H., Blois-Heulin, C., & Vauclair, J. (2013). Baboons' hand preference resists to spatial factors for a communicative gesture but not for a simple manipulative action. *Developmental Psychobiology*, *55*(6), 651-661. https://doi.org/10.1002/dev.21128

Bourjade, M., Moulinot, M., Henry, S., Richard-Yris, M.-A., & Hausberger, M. (2008). Could adults be used to improve social skills of young horses, Equus caballus? *Developmental Psychobiology*, *50*(4), 408-417. https://doi.org/10.1002/dev.20301

Bourjade, M., & Sueur, C. (2010). Shared or unshared consensus for collective movement? Towards methodological concerns. *Behavioural Processes*, *84*(3), 648-652.

Bourjade, M., Thierry, B., Call, J., & Dufour, V. (2012). Are monkeys able to plan for future exchange? *Animal Cognition*, *15*(5), 783-795. https://doi.org/10.1007/s10071-012-0502-1

Bourjade, M., Thierry, B., Maumy, M., & Petit, O. (2009). Decision-making in przewalski horses (Equus ferus przewalskii) is driven by the ecological contexts of collective movements. *Ethology*, *115*(4), 321-330. https://doi.org/10.1111/j.1439-0310.2009.01614.x

Bourjade, M., & Vautier, S. (2017). *Un arbre de décision pour évaluer l'état des connaissances dans le domaine du mesurage en psychologie.* VIème édition de MODEVAIA, Marcevol, France.

Bowlby, J. (1969). *Attachment and Loss, Vol. 1: Attachment. New York:* Basic books.

Bronfenbrenner, U. (1977). Toward an experimental ecology of human development. *American psychologist*, *32*(7), 513.

Bruner, J. S. (1975). The ontogenesis of speech acts. *Journal of child language*, *2*(01), 1-19.

Byrne, R. W., Cartmill, E., Genty, E., Graham, K. E., Hobaiter, C., & Tanner, J. (2017). Great ape gestures : Intentional communication with a rich set of innate signals. *Animal cognition*, *20*(4), 755-769.

Call, J., & Tomasello, M. (2007). *The gestural communication of apes and monkeys* (Vol. viii). Taylor & Francis Group/Lawrence Erlbaum Associates.

Callaghan, T., Moll, H., Rakoczy, H., Warneken, F., Liszkowski, U., Behne, T., Tomasello, M., & Collins, W. A. (2011). Early social cognition in three cultural contexts. *Monographs of the society for research in child development*, i-142.

Camaioni, L. (1997). The emergence of intentional communication in ontogeny, phylogeny, and pathology. *European Psychologist*, *2*(3), 216-225.

Camaioni, L., Perucchini, P., Bellagamba, F., & Colonnesi, C. (2004). The role of declarative pointing in developing a theory of mind. *Infancy*, *5*(3), 291-308. https://doi.org/10.1207/s15327078in0503_3

Cámara-Leret, R., & Bascompte, J. (2021). Language extinction triggers the loss of unique medicinal knowledge. *Proceedings of the National Academy of Sciences*, *118*(24).

Camus, A. (1986). Le Mythe de Sisyphe.[1942]. *Gallimard, Folio essais*.

Canteloup, C., Bovet, D., & Meunier, H. (2015). Do Tonkean macaques (Macaca tonkeana) tailor their gestural and visual signals to fit the attentional states of a human partner? *Animal cognition*, *18*(2), 451-461.

Cap, H. (2015). Behavior and evolution: Crossed glances. In *Handbook of Evolutionary Thinking in the Sciences* (p. 471-497). Springer. http://link.springer.com/chapter/10.1007/978-94-017-9014-7_22

Carpenter, M., & Call, J. (2013). How joint is the joint attention of apes and human infants. In: J. Metcalfe & H. S. Terrace (eds), *Agency and joint attention* (p. *49-61*). Oxford.

Carpenter, M., Nagell, K., Tomasello, M., Butterworth, G., & Moore, C. (1998). Social Cognition, Joint Attention, and Communicative Competence from 9 to 15 Months of Age. *Monographs of the Society for Research in Child Development*, *63*(4), 1-174. https://doi.org/10.2307/1166214

Carra, C., Lavelli, M., Keller, H., & Kärtner, J. (2013). Parenting infants : Socialization goals and behaviors of Italian mothers and immigrant mothers from West Africa. *Journal of Cross-Cultural Psychology*, *44*(8), 1304-1320.

Cartmill, E. A., & Byrne, R. W. (2007). Orangutans modify their gestural signaling according to their audience's comprehension. *Current Biology*, *17*(15), 1345-1348. https://doi.org/10.1016/j.cub.2007.06.069

Cartmill, E., & Byrne, R. (2010). Semantics of primate gestures : Intentional meanings of orangutan gestures. *Animal Cognition*, *13*(6), 793-804. https://doi.org/10.1007/s10071-010-0328-7

Casillas, M., Brown, P., & Levinson, S. C. (2020). Early language experience in a Tseltal Mayan village. *Child Development*, *91*(5), 1819-1835.

Casillas, M., Brown, P., & Levinson, S. C. (2021). Early language experience in a Papuan community. *Journal of Child Language*, *48*(4), 792-814.

Cerling, T. E., Wynn, J. G., Andanje, S. A., Bird, M. I., Korir, D. K., Levin, N. E., Mace, W., Macharia, A. N., Quade, J., & Remien, C. H. (2011). Woody cover and hominin environments in the past 6 million years. *Nature*, *476*(7358), 51-56.

Chance, M. R. (1967). Attention structure as the basis of primate rank orders. *Man*, *2*(4), 503-518.

Chapelain, A., Laurence, A., Vimond, M., Maille, A., Meunier, H., Fagard, J., Vauclair, J., & Blois-Heulin, C. (2012). Hand preference and its flexibility according to the position of the object : A study in cercopithecines examining spontaneous behaviour and an experimental task (the Bishop QHP task). *Animal Cognition*, *15*(5), 937-953. https://doi.org/10.1007/s10071-012-0520-z

Chapelain, A. S., Hogervorst, E., Mbonzo, P., & Hopkins, W. D. (2011). Hand preferences for bimanual coordination in 77 bonobos (Pan paniscus) : Replication and extension. *International Journal of Primatology*, *32*(2), 491-510.

Chapman, S. (2009). *Key ideas in linguistics and the philosophy of language*. Edinburgh University Press.

Cheney, D. L., & Seyfarth, R. M. (2018). Flexible usage and social function in primate vocalizations. *Proceedings of the National Academy of Sciences*, *115*(9), 1974-1979.

Chevallier, C., Kohls, G., Troiani, V., Brodkin, E. S., & Schultz, R. T. (2012). The social motivation theory of autism. *Trends in cognitive sciences*, *16*(4), 231-239.

Cochet, H., & Byrne, R. W. (2016). Communication in the second and third year of life : Relationships between nonverbal social skills and language. *Infant Behavior and Development*, *44*, 189-198.

Cochet, H., & Vauclair, J. (2010a). Pointing gesture in young children : Hand preference and language development. *Gesture*, *10*(2-3), 129-149. https://doi.org/10.1075/gest.10.2-3.02coc

Cochet, H., & Vauclair, J. (2010b). Pointing gestures produced by toddlers from 15 to 30 months : Different functions, hand shapes and laterality patterns. *Infant Behavior and Development*, *33*(4), 431-441. https://doi.org/10.1016/j.infbeh.2010.04.009

Colonnesi, C., Stams, G. J. J., Koster, I., & Noom, M. J. (2010). The relation between pointing and language development : A meta-analysis. *Developmental Review*, *30*(4), 352-366.

Conradt, L., & Roper, T. J. (2005). Consensus decision making in animals. *Trends in ecology & evolution*, *20*(8), 449-456.

Corballis, M. C. (2003). From mouth to hand : Gesture, speech, and the evolution of right-handedness. *Behavioral and Brain Sciences*, *26*(02), 199-208. https://doi.org/10.1017/S0140525X03000062

Corballis, M. C., Badzakova-Trajkov, G., & Häberling, I. S. (2012). Right hand, left brain : Genetic and evolutionary bases of cerebral asymmetries for language and manual action. *Wiley Interdisciplinary Reviews: Cognitive Science*, *3*(1), 1-17. https://doi.org/10.1002/wcs.158

Coudé, G., Ferrari, P. F., Rodà, F., Maranesi, M., Borelli, E., Veroni, V., Monti, F., Rozzi, S., & Fogassi, L. (2011). Neurons controlling voluntary vocalization in the macaque ventral premotor cortex. *PloS one*, *6*(11), e26822.

Courtin, C., Jobard, G., Vigneau, M., Beaucousin, V., Razafimandimby, A., Hervé, P.-Y., Mellet, E., Zago, L., Petit, L., & Mazoyer, B. (2011). A common neural system is activated in hearing non-signers to process French Sign language and spoken French. *Brain research bulletin*, *84*(1), 75-87.

Couzin, I. D., Krause, J., Franks, N. R., & Levin, S. A. (2005). Effective leadership and decision-making in animal groups on the move. *Nature*, *433*(7025), 513-516.

Cristia, A., Dupoux, E., Gurven, M., & Stieglitz, J. (2019). Child-directed speech is infrequent in a forager-farmer population : A time allocation study. *Child development*, *90*(3), 759-773.

Csibra, G. (2010). Recognizing communicative intentions in infancy. *Mind & Language*, *25*(2), 141-168.

Dafreville, M., Hobaiter, C., Guidetti, M., Sillam-Dussès, D., & Bourjade, M. (2021). Sensitivity to the communicative partner's attentional state : A developmental study on mother–infant dyads in wild chimpanzees (Pan troglodytes schweinfurthii). *American Journal of Primatology*, e23339.

Dall, S. R., Lotem, A., Winkler, D. W., Bednekoff, P. A., Laland, K. N., Coolen, I., Kendal, R., Danchin, É., Giraldeau, L.-A., & Valone, T. J. (2005). Defining the concept of public information. *Science*, 353-356.

Danchin, E., Giraldeau, L.-A., Valone, T. J., & Wagner, R. H. (2004). Public information : From nosy neighbors to cultural evolution. *Science*, *305*(5683), 487-491.

Danchin, E., & Pocheville, A. (2014). Inheritance is where physiology meets evolution. *The Journal of physiology*, *592*(11), 2307-2317.

Davidson, R. J., Shackman, A. J., & Maxwell, J. S. (2004). Asymmetries in face and brain related to emotion. *Trends in cognitive sciences*, *8*(9), 389-391.

De Waal, F. B. (1999). Anthropomorphism and anthropodenial : Consistency in our thinking about humans and other animals. *Philosophical Topics*, *27*(1), 255-280.

Dediu, D., & Levinson, S. C. (2013). On the antiquity of language : The reinterpretation of Neandertal linguistic capacities and its consequences. *Frontiers in psychology*, *4*, 397.

Dennett, D. (2009). Intentional systems theory. *The Oxford handbook of philosophy of mind*, 339-350.

Dennett, D. C. (1971). Intentional systems. *The Journal of Philosophy*, *68*(4), 87-106.

Deshpande, A., Gupta, S., & Sinha, A. (2018). Intentional communication between wild bonnet macaques and humans. *Scientific reports*, *8*(1), 1-12.

Dezecache, G., Zuberbühler, K., Davila-Ross, M., & Dahl, C. D. (2021). Flexibility in wild infant chimpanzee vocal behavior. *Journal of Language Evolution*, *6*(1), 37-53.

Dissanayake, E. (1992). *Homo Aestheticus: Where art comes from and why. Seattle and*. London: University of Washington Press.

D'Odorico, L., Cassibba, R., & Salerni, N. (1997). Temporal relationships between gaze and vocal behavior in prelinguistic and linguistic communication. *Journal of Psycholinguistic Research*, *26*(5), 539-556.

Donnellan, E., Bannard, C., McGillion, M. L., Slocombe, K. E., & Matthews, D. (2019). Infants' intentionally communicative vocalizations elicit responses from caregivers and are the best predictors of the transition to language : A longitudinal investigation of infants' vocalizations, gestures and word production. *Developmental science*, e12843.

Driver, J. (1996). Enhancement of selective listening by illusory mislocation of speech sounds due to lip-reading. *Nature*, *381*(6577), 66.

Dufour, V., Pelé, M., Neumann, M., Thierry, B., & Call, J. (2009). Calculated reciprocity after all : Computation behind token transfers in orang-utans. *Biology Letters*, *5*(2), 172-175. https://doi.org/10.1098/rsbl.2008.0644

Dufour, V., & Sterck, E. H. M. (2008). Chimpanzees fail to plan in an exchange task but succeed in a tool-using procedure. *Behavioural Processes*, *79*(1), 19-27. https://doi.org/10.1016/j.beproc.2008.04.003

Dunbar, R. (1998). *Grooming, gossip, and the evolution of language*. Harvard University Press.

Emery, N. J. (2000). The eyes have it : The neuroethology, function and evolution of social gaze. *Neuroscience & Biobehavioral Reviews*, *24*(6), 581-604. https://doi.org/10.1016/S0149-7634(00)00025-7

Emmorey, K., Mehta, S., & Grabowski, T. J. (2007). The neural correlates of sign versus word production. *NeuroImage*, *36*(1), 202-208. https://doi.org/10.1016/j.neuroimage.2007.02.040

Enard, W., Przeworski, M., Fisher, S. E., Lai, C. S. L., Wiebe, V., Kitano, T., Monaco, A. P., & Pääbo, S. (2002). Molecular evolution of FOXP2, a gene involved in speech and language. *Nature*, *418*(6900), 869-872. https://doi.org/10.1038/nature01025

Esteve-Gibert, N., & Guellaï, B. (2018). Prosody in the auditory and visual domains : A developmental perspective. *Frontiers in psychology*, *9*, 338.

Esteve-Gibert, N., & Prieto, P. (2013). Prosody signals the emergence of intentional communication in the first year of life : Evidence from Catalan-babbling infants. *Journal of child language*, *40*(5), 919-944.

Everett, D. (2005). Cultural constraints on grammar and cognition in Pirahã : Another look at the design features of human language. *Current anthropology*, *46*(4), 621-646.

Ey, E., Rahn, C., Hammerschmidt, K., & Fischer, J. (2009). Wild female olive baboons adapt their grunt vocalizations to environmental conditions. *Ethology*, *115*(5), 493-503.

Falk, D. (2004). Prelinguistic evolution in early hominins : Whence motherese? *Behavioral and Brain Sciences*, *27*(04), 491-503. https://doi.org/10.1017/S0140525X04000111

Faurie, C., & Raymond, M. (2004). Handedness frequency over more than ten thousand years. *Proceedings of the Royal Society B: Biological Sciences*, *271*(Suppl_3), S43-S45. https://doi.org/10.1098/rsbl.2003.0092

Fay, N., Arbib, M., & Garrod, S. (2013). How to bootstrap a human communication system. *Cognitive science*, *37*(7), 1356-1367.

Fay, N., Ellison, M., & Garrod, S. (2014a). Iconicity : From sign to system in human communication and language. *Pragmatics & Cognition*, *22*(2), 244-263.

Fay, N., Lister, C. J., Ellison, T. M., & Goldin-Meadow, S. (2014b). Creating a communication system from scratch : Gesture beats vocalization hands down. *Frontiers in Psychology*, *5*, 354.

Fay, N., Walker, B., Swoboda, N., Umata, I., Fukaya, T., Katagiri, Y., & Garrod, S. (2018). Universal principles of human communication: Preliminary evidence from a cross-cultural communication game. *Cognitive science*, *42*(7), 2397-2413.

Fedurek, P., Zuberbühler, K., & Semple, S. (2017). Trade-offs in the production of animal vocal sequences : Insights from the structure of wild chimpanzee pant hoots. *Frontiers in zoology*, *14*(1), 1-10.

Feldman, R. (2012). Bio-behavioral synchrony : A model for integrating biological and microsocial behavioral processes in the study of parenting. *Parenting*, *12*(2-3), 154-164.

Fernald, A., Marchman, V. A., & Weisleder, A. (2013). SES differences in language processing skill and vocabulary are evident at 18 months. *Developmental science*, *16*(2), 234-248.

Ferrari, P. F., Gallese, V., Rizzolatti, G., & Fogassi, L. (2003). Mirror neurons responding to the observation of ingestive and communicative mouth actions in the monkey ventral premotor cortex. *European journal of neuroscience*, *17*(8), 1703-1714.

Fischer, J. (2017). Primate vocal production and the riddle of language evolution. *Psychonomic Bulletin & Review*, *24*(1), 72-78.

Fischer, J., Wegdell, F., Trede, F., Dal Pesco, F., & Hammerschmidt, K. (2020). Vocal convergence in a multi-level primate society : Insights into the evolution of vocal learning. *Proceedings of the Royal Society B*, *287*(1941), 20202531.

Fitch, W. T. (2010). *The Evolution of Language*. Cambridge University Press.

Fitch, W. T., & Reby, D. (2001). The descended larynx is not uniquely human. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *268*(1477), 1669-1675. https://doi.org/10.1098/rspb.2001.1704

Florkiewicz, B., & Campbell, M. (2021). Chimpanzee facial gestures and the implications for the evolution of language. *PeerJ*, *9*, e12237.

Font, E., & Carazo, P. (2010). Animals in translation : Why there is meaning (but probably no message) in animal communication. *Animal Behaviour*, *80*(2), e1-e6. https://doi.org/10.1016/j.anbehav.2010.05.015

Forrester, G. S. (2008). A multidimensional approach to investigations of behaviour : Revealing structure in animal communication signals. *Animal Behaviour*, *76*(5), 1749-1760. https://doi.org/10.1016/j.anbehav.2008.05.026

Forrester, G. S., Quaresmini, C., Leavens, D. A., Mareschal, D., & Thomas, M. S. (2013). Human handedness : An inherited evolutionary trait. *Behavioural brain research*, *237*, 200-206.

Franco, F., & Butterworth, G. (1996). Pointing and social awareness : Declaring and requesting in the second year. *Journal of Child Language*, *23*(02), 307-336. https://doi.org/10.1017/S0305000900008813

French, V. (2019). Ancient history of parenting. In M. Bornstein (ed), *Handbook of parenting* (p. 287-319). Routledge.

Fröhlich, M., Lee, K., Mitra Setia, T., Schuppli, C., & van Schaik, C. P. (2019). The loud scratch : A newly identified gesture of Sumatran orangutan mothers in the wild. *Biology letters*, *15*(7), 20190209.

Fröhlich, M., Müller, G., Zeiträg, C., Wittig, R. M., & Pika, S. (2017). Gestural development of chimpanzees in the wild : The impact of interactional experience. *Animal Behaviour*, *134*, 271-282.

Fröhlich, M., Sievers, C., Townsend, S. W., Gruber, T., & van Schaik, C. P. (2019). Multimodal communication and language origins : Integrating gestures and vocalizations. *Biological Reviews*, *94*(5), 1809-1829.

Fröhlich, M., & van Schaik, C. P. (2018). The function of primate multimodal communication. *Animal cognition*, *21*(5), 619-629.

Fröhlich, M., Wittig, R. M., & Pika, S. (2016). Should I stay or should I go? Initiation of joint travel in mother–infant dyads of two chimpanzee communities in the wild. *Animal Cognition*, *19*(3), 483-500.

Fröhlich, M., Wittig, R. M., & Pika, S. (2019). The ontogeny of intentional communication in chimpanzees in the wild. *Developmental science*, *22*(1), e12716.

Fureix, C., Bourjade, M., Henry, S., Sankey, C., & Hausberger, M. (2012). Exploring aggression regulation in managed groups of horses Equus caballus. *Applied Animal Behaviour Science*, *138*(3), 216-228. https://doi.org/10.1016/j.applanim.2012.02.009

Gainotti, G. (2012). Unconscious processing of emotions and the right hemisphere. *Neuropsychologia*, *50*(2), 205-218.

Galantucci, B., & Garrod, S. (2011). Experimental semiotics : A review. *Frontiers in human neuroscience*, *5*. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3043271/

Gallese, V., Fadiga, L., Fogassi, L., & Rizzolatti, G. (1996). Action recognition in the premotor cortex. *Brain*, *119*(2), 593-609.

Gaunet, F., & Deputte, B. L. (2011). Functionally referential and intentional communication in the domestic dog : Effects of spatial and social contexts. *Animal Cognition*, *14*(6), 849-860. https://doi.org/10.1007/s10071-011-0418-1

Geary, D. C. (1995). Reflections of evolution and culture in children's cognition : Implications for mathematical development and instruction. *American psychologist*, *50*(1), 24.

Geary, D. C., & Bjorklund, D. F. (2000). Evolutionary developmental psychology. *Child development*, *71*(1), 57-65.

Genty, E., Breuer, T., Hobaiter, C., & Byrne, R. (2009). Gestural communication of the gorilla (Gorilla gorilla): Repertoire, intentionality and possible origins. *Animal Cognition*, *12*(3), 527-546. https://doi.org/10.1007/s10071-009-0213-4

Genty, E., & Byrne, R. (2010). Why do gorillas make sequences of gestures? *Animal Cognition*, *13*(2), 287-301. https://doi.org/10.1007/s10071-009-0266-4

Genty, E., & Zuberbühler, K. (2014). Spatial reference in a bonobo gesture. *Current Biology*, *24*(14), 1601-1605.

Ghazanfar, A. A. (2013). Multisensory vocal communication in primates and the evolution of rhythmic speech. *Behavioral Ecology and Sociobiology*, *67*(9), 1441-1448.

Ghazanfar, A. A., Takahashi, D. Y., Mathur, N., & Fitch, W. T. (2012). Cineradiography of monkey lip-smacking reveals putative precursors of speech dynamics. *Current Biology*, *22*(13), 1176-1182.

Girard-Buttoz, C., Zaccarella, E., Bortolato, T., Friederici, A. D., Wittig, R. M., & Crockford, C. (2021). Chimpanzees use numerous flexible vocal sequences with more than two vocal units : A step towards language? *bioRxiv*. https://doi.org/10.1101/2021.02.03.429517

Gogate, L. J., Walker-Andrews, A. S., & Bahrick, L. E. (2001). The intersensory origins of word-comprehension : An ecological–dynamic systems view. *Developmental science*, *4*(1), 1-18.

Goldin-Meadow, S. (2012). Homesign : Gesture to language. In R. Pfau, M. Steinbach, & B. Woll (eds), *Sign Language* (p. 601-625). De Gruyter Mouton.

Goldstein, M. H., King, A. P., & West, M. J. (2003). Social interaction shapes babbling : Testing parallels between birdsong and speech. *Proceedings of the National Academy of Sciences*, *100*(13), 8030-8035.

Goldstein, M. H., & Schwade, J. A. (2008). Social feedback to infants' babbling facilitates rapid phonological learning. *Psychological science*, *19*(5), 515-523.

Goldstein, M. H., Schwade, J. A., & Bornstein, M. H. (2009). The value of vocalizing : Five-month-old infants associate their own noncry vocalizations with responses from caregivers. *Child development*, *80*(3), 636-644.

Gottlieb, G. (1987). The developmental basis of evolutionary change. *Journal of Comparative Psychology*, *101*(3), 262.

Gottlieb, G. (2007). Probabilistic epigenesis. *Developmental science*, *10*(1), 1-11.

Graham, K. E., Furuichi, T., & Byrne, R. W. (2017). The gestural repertoire of the wild bonobo (Pan paniscus) : A mutually understood communication system. *Animal Cognition*, *20*(2), 171-177.

Graham, K. E., Wilke, C., Lahiff, N. J., & Slocombe, K. E. (2020). Scratching beneath the surface : Intentionality in great ape signal production. *Philosophical Transactions of the Royal Society B*, *375*(1789), 20180403.

Gratier, M., Devouche, E., Guellai, B., Infanti, R., Yilmaz, E., & Parlato-Oliveira, E. (2015). Early development of turn-taking in vocal interaction between mothers and infants. *Frontiers in Psychology*, *6*(1167), 10-3389.

Gratier, M., & Trevarthen, C. (2008). Musical narrative and motives for culture in mother-infant vocal interaction. *Journal of Consciousness Studies*, *15*(10-11), 122-158.

Gravener, J. A., Rogosch, F. A., Oshri, A., Narayan, A. J., Cicchetti, D., & Toth, S. L. (2012). The relations among maternal depressive disorder, maternal expressed emotion, and toddler behavior problems and attachment. *Journal of abnormal child psychology*, *40*(5), 803-813.

Greenfield, P. M. (1991). From hand to mouth. *Behavioral and Brain Sciences*, *14*(04), 577-595. https://doi.org/10.1017/S0140525X0007148X

Grice, H. P. (1957). Meaning. *The philosophical review*, *66*, 377-388.

Gros-Louis, J. & Wu, Z. (2012). Twelve-month-olds' vocal production during pointing in naturalistic interactions : Sensitivity to parents' attention and responses. *Infant behavior & development*, *35*, 773-778.

Grünloh, T., & Liszkowski, U. (2015). Prelinguistic vocalizations distinguish pointing acts. *Journal of child language*, *42*(6), 1312-1336.

Guidetti, M. (2002). The emergence of pragmatics : Forms and functions of conventional gestures in young French children. *First Language*, *22*(3), 265-285.

Guidetti, M. (2005). Yes or no? How young French children combine gestures and speech to agree and refuse. *Journal of Child Language*, *32*(4), 911-924.

Guidetti, M., Fibigerova, K., & Colletta, J.-M. (2014). Gestures and multimodal development. In M. Seyfeddinipur & M. Gullberg (eds), *From Gesture in conversation to visible action as utterance: Essays in honor of Adam Kendon* (p. 351-371). John Benjamins Publishing Company

Gupta, S., & Sinha, A. (2016). Not here, there ! Possible referential gesturing during allogrooming by wild bonnet macaques, Macaca radiata. *Animal cognition*, *19*(6), 1243-1248.

Gupta, S., & Sinha, A. (2019). Gestural communication of wild bonnet macaques in the Bandipur National Park, Southern India. *Behavioural processes*, *168*, 103956.

Gustison, M. L., & Bergman, T. J. (2017). Divergent acoustic properties of gelada and baboon vocalizations and their implications for the evolution of human speech. *Journal of language evolution*, *2*(1), 20-36.

Gustison, M. L., le Roux, A., & Bergman, T. J. (2012). Derived vocalizations of geladas (Theropithecus gelada) and the evolution of vocal complexity in primates. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1597), 1847-1859.

Gustison, M. L., Semple, S., Ferrer-i-Cancho, R., & Bergman, T. J. (2016). Gelada vocal sequences follow Menzerath's linguistic law. *Proceedings of the National Academy of Sciences*, *113*(19), E2750-E2758.

Hage, S. R., & Nieder, A. (2013). Single neurons in monkey prefrontal cortex encode volitional initiation of vocalizations. *Nature communications*, *4*(1), 1-11.

Hagen, E. H. (2002). Depression as bargaining : The case postpartum. *Evolution and Human Behavior*, *23*(5), 323-336.

Hagen, E. H. (2003). The bargaining model of depression. *Genetic and cultural evolution of cooperation*, 95-123.

Hamilton, W. D. (1964). The genetical theory of kin selection. *Journal of Theoretical. Biology*, *7*(7), 1-52.

Harding, C. G., & Golinkoff, R. M. (1979). The origins of intentional vocalizations in prelinguistic infants. *Child development*, 33-40.

Hardt, B., & Benedict, L. (2021). Can you hear me now? A review of signal transmission and experimental evidence for the acoustic adaptation hypothesis. *Bioacoustics*, *30*(6), 716-742.

Hare, B., & Tomasello, M. (2004). Chimpanzees are more skilful in competitive than in cooperative cognitive tasks. *Animal Behaviour*, *68*(3), 571-581. https://doi.org/10.1016/j.anbehav.2003.11.011

Harmon, L. (2018). *Phylogenetic comparative methods : Learning from trees*.

Hattori, Y., Kuroshima, H., & Fujita, K. (2010). Tufted capuchin monkeys (Cebus apella) show understanding of human attentional states when requesting food held by a human. *Animal Cognition*, *13*(1), 87-92. https://doi.org/10.1007/s10071-009-0248-6

Hauser, M. D., Chomsky, N., & Fitch, W. T. (2002). The Faculty of Language : What Is It, Who Has It, and How Did It Evolve? *Science*, *298*(5598), 1569-1579. https://doi.org/10.1126/science.298.5598.1569

Heesen, R., Hobaiter, C., Ferrer-i-Cancho, R., & Semple, S. (2019). Linguistic laws in chimpanzee gestural communication. *Proceedings of the Royal Society B*, *286*(1896), 20182900.

Henrich, J., Heine, S. J., & Norenzayan, A. (2010a). Most people are not WEIRD. *Nature*, *466*(7302), 29-29.

Henrich, J., Heine, S. J., & Norenzayan, A. (2010b). The weirdest people in the world? *Behavioral and brain sciences*, *33*(2-3), 61-83.

Hewes, G. W. (1973). Primate communication and the gestural origin of language. *Current Anthropology*, *14*(1/2), 5-24.

Higham, J. P., & Hebets, E. A. (2013). An introduction to multimodal communication. *Behavioral Ecology and Sociobiology*, *67*(9), 1381-1388.

Hilbrink, E. E., Gattis, M., & Levinson, S. C. (2015). Early developmental changes in the timing of turn-taking : A longitudinal study of mother–infant interaction. *Frontiers in psychology*, *6*, 1492.

Hinde, R. A. (1983). *Primate social relationships : An integrated approach*. Sinauer Associates.

Ho, M. W. (1998). Evolution. In G. Greenberg & M. H. Haraway (eds), *Comparative psychology : A handbook* (p. 107). Routledge, Taylor & Francis Group/Lawrence Erlbaum Associates.

Hobaiter, C., & Byrne, R. (2011a). The gestural repertoire of the wild chimpanzee. *Animal Cognition*, *14*(5), 745-767. https://doi.org/10.1007/s10071-011-0409-2

Hobaiter, C., & Byrne, R. W. (2011b). Serial gesturing by wild chimpanzees : Its nature and function for communication. *Animal cognition*, *14*(6), 827-838.

Hobaiter, C., & Byrne, R. W. (2014). The meanings of chimpanzee gestures. *Current Biology*, *24*(14), 1596-1600.

Hobaiter, C., & Byrne, R. W. (2017). What is a gesture? A meaning-based approach to defining gestural repertoires. *Neuroscience & Biobehavioral Reviews*, *82*, 3-12.

Hobaiter, C., Byrne, R. W., & Zuberbühler, K. (2017). Wild chimpanzees' use of single and combined vocal and gestural signals. *Behavioral Ecology and Sociobiology*, *71*(6), 96. https://doi.org/10.1007/s00265-017-2325-1

Hockett, C. F. (1960). The origin of speech. *Scientific American*, *203*(3), 88-97.

Hopkins, W. D., & Cantero, M. (2003). From hand to mouth in the evolution of language : The influence of vocal behavior on lateralized hand use in manual gestures by chimpanzees (Pan troglodytes). *Developmental Science*, *6*(1), 55-61.

Hopkins, W. D., Russell, J., Freeman, H., Buehler, N., Reynolds, E., & Schapiro, S. J. (2005). The distribution and development of handedness for manual gestures in captive chimpanzees (Pan troglodytes). *Psychological Science*, *16*(6), 487-493. https://doi.org/10.1111/j.0956-7976.2005.01561.x

Hopkins, W. D., Taglialatela, J. P., & Leavens, D. A. (2007). Chimpanzees differentially produce novel vocalizations to capture the attention of a human. *Animal Behaviour*, *73*(2), 281-286. https://doi.org/10.1016/j.anbehav.2006.08.004

Hopkins, W. D., & Vauclair, J. (2012). Evolution of behavioral and brain asymmetries in primates. In K. R. Gibson & M. Tallerman (eds), *Handbook of language evolution*, (p. 184-197). *Oxford University Press, Oxford*.

Hostetter, A. B., Cantero, M., & Hopkins, W. D. (2001). Differential use of vocal and gestural communication by chimpanzees (Pan troglodytes) in response to the attentional status of a human (Homo sapiens). *Journal of Comparative Psychology*, *115*(4), 337-343. https://doi.org/10.1037/0735-7036.115.4.337

Hostetter, A., Russell, J., Freeman, H., & Hopkins, W. (2007). Now you see me, now you don't : Evidence that chimpanzees understand the role of the eyes in attention. *Animal Cognition, 10*(1), 55-62. https://doi.org/10.1007/s10071-006-0031-x

Hughes, S. M., Dispenza, F., & Gallup Jr, G. G. (2004). Ratings of voice attractiveness predict sexual behavior and body configuration. *Evolution and Human Behavior*, *25*(5), 295-304.

Igualada, A., Bosch, L., & Prieto, P. (2015). Language development at 18 months is related to multimodal communicative strategies at 12 months. *Infant Behavior and Development, 39*, 42-52.

Iverson, J. M., & Goldin-Meadow, S. (2005). Gesture paves the way for language development. *Psychological Science, 16*(5), 367-371. https://doi.org/10.1111/j.0956-7976.2005.01542.x

Iverson, J., & Thelen, E. (1999). *Hand, mouth and brain. The dynamic emergence of speech and gesture*. *Journal of Consciousness studies, 6* (11-12), 19-40.

Jablonka, E., & Lamb, M. J. (2014). *Evolution in four dimensions, revised edition : Genetic, epigenetic, behavioral, and symbolic variation in the history of life*. MIT press.

James, W. (1884). What is an emotion? *Mind, IX*, 188-204.

Joffe, T. H. (1997). Social pressures have selected for an extended juvenile period in primates. *Journal of human evolution, 32*(6), 593-605.

Johnson, M. H. (2005). Subcortical face processing. *Nature Reviews Neuroscience*, *6*(10), 766-774.

Jouen, F., & Molina, M. (2007). *Naissance et connaissance : La cognition néonatale*. Editions Mardaga.

Jover, M., Cellier, M., & Scola, C. (2019). Infants' motor activity during a mother–infant interaction alternating silent and singing phases. *Journal of Motor Learning and Development, 7*(3), 426-436.

Jürgens, U. (2002). Neural pathways underlying vocal control. *Neuroscience & Biobehavioral Reviews, 26*(2), 235-258. https://doi.org/10.1016/S0149-7634(01)00068-9

Jürgens, U., & von Cramon, D. (1982). On the role of the anterior cingulate cortex in phonation : A case report. *Brain and Language*, *15*(2), 234-248.

Kaminski, J., Call, J., & Fischer, J. (2004). Word learning in a domestic dog : Evidence for "fast mapping". *Science, 304*(5677), 1682-1683.

Kaminski, J., Call, J., & Tomasello, M. (2004). Body orientation and face orientation : Two factors controlling apes' begging behavior from humans. *Animal Cognition, 7*(4), 216-223. https://doi.org/10.1007/s10071-004-0214-2

Kampe, K. K., Frith, C. D., & Frith, U. (2003). "Hey John" : Signals conveying communicative intention toward the self activate brain regions associated with "mentalizing," regardless of modality. *Journal of Neuroscience*, *23*(12), 5258-5263.

Katz, W. F., & Mehta, S. (2015). Visual feedback of tongue movement for novel speech sound learning. *Frontiers in human neuroscience, 9,* 612.

Keller, H. (2008). Culture and biology : The foundation of pathways of development. *Social and personality psychology Compass*, *2*(2), 668-681.

Keller, H., Bard, K., Morelli, G., Chaudhary, N., Vicedo, M., Rosabal-Coto, M., Scheidecker, G., Murray, M., & Gottlieb, A. (2018). The myth of universal sensitive responsiveness : Comment on Mesman et al.(2017). *Child Development*, *89*(5), 1921-1928.

Keller, H., Borke, J., Staufenbiel, T., Yovsi, R. D., Abels, M., Papaligoura, Z., Jensen, H., Lohaus, A., Chaudhary, N., & Lo, W. (2009). Distal and proximal parenting as alternative parenting strategies during infants' early months of life : A cross-cultural study. *International Journal of Behavioral Development*, *33*(5), 412-420.

Keller, H., Lamm, B., Abels, M., Yovsi, R., Borke, J., Jensen, H., Papaligoura, Z., Holub, C., Lo, W., & Tomiyama, A. J. (2006). Cultural models, socialization goals, and parenting ethnotheories : A multicultural analysis. *Journal of cross-cultural psychology*, *37*(2), 155-172.

Keller, H., Völker, S., & Yovsi, R. D. (2005). Conceptions of parenting in different cultural communities : The case of West African Nso and Northern German women. *Social development*, *14*(1), 158-180.

Keller, H., Yovsi, R., Borke, J., Kärtner, J., Jensen, H., & Papaligoura, Z. (2004). Developmental consequences of early parenting experiences : Self-recognition and self-regulation in three cultural communities. *Child development*, *75*(6), 1745-1760.

Kendon, A. (2011). Vocalisation, speech, gesture, and the language origins debate : An essay review on recent contributions. *Gesture*, *11*(3), 349-370.

Kershenbaum, A., Bowles, A. E., Freeberg, T. M., Jin, D. Z., Lameira, A. R., & Bohn, K. (2014). Animal vocal sequences : Not the Markov chains we thought they were. *Proceedings of the Royal Society B: Biological Sciences*, *281*(1792), 20141370.

Keysers, C., Kohler, E., Umiltà, M. A., Nanetti, L., Fogassi, L., & Gallese, V. (2003). Audiovisual mirror neurons and action recognition. *Experimental brain research*, *153*(4), 628-636.

Kik, A., Adamec, M., Aikhenvald, A. Y., Bajzekova, J., Baro, N., Bowern, C., Colwell, R. K., Drozd, P., Duda, P., & Ibalim, S. (2021). Language and ethnobiological skills decline precipitously in Papua New Guinea, the world's most linguistically diverse nation. *Proceedings of the National Academy of Sciences*, *118*(22).

Kirby, S. (1999). *Function, selection, and innateness : The emergence of language universals*. OUP Oxford.

Kita, S. (2003, juin 20). *Pointing : A foundational building block of human communication*. In S. Kita (ed), Pointing (p.8-25). New York Psychology Press.

Klein, J. T., Shepherd, S. V., & Platt, M. L. (2009). Social attention and the brain. *Current Biology*, *19*(20), R958-R962.

Knecht, S., Dräger, B., Deppe, M., Bobe, L., Lohmann, H., Flöel, A., Ringelstein, E.-B., & Henningsen, H. (2000). Handedness and hemispheric language dominance in healthy humans. *Brain, 123*(12), 2512-2518. https://doi.org/10.1093/brain/123.12.2512

Kobayashi, H., & Kohshima, S. (2001). Unique morphology of the human eye and its adaptive meaning : Comparative studies on external morphology of the primate eye. *Journal of Human Evolution, 40*(5), 419-435. https://doi.org/10.1006/jhev.2001.0468

Koda, H. (2004). Flexibility and context-sensitivity during the vocal exchange of coo calls in wild Japanese macaques (Macaca fuscata yakui). *Behaviour, 141*(10), 1279-1296.

Lai, C. S., Fisher, S. E., Hurst, J. A., Vargha-Khadem, F., & Monaco, A. P. (2001). A forkhead-domain gene is mutated in a severe speech and language disorder. *Nature, 413*(6855), 519-523.

Lamaury, A., Cochet, H., & Bourjade, M. (2019). Acquisition of joint attention by olive baboons gesturing toward humans. *Animal cognition*, *22*(4), 567-575.

Lameira, A. R. (2018). Origins of human consonants and vowels : Articulatory continuities with great apes. *Origins of Human Language: Continuities and Discontinuities with Nonhuman Primates*, 75-100.

Lameira, A. R., Hardus, M. E., Bartlett, A. M., Shumaker, R. W., Wich, S. A., & Menken, S. B. (2015). Speech-like rhythm in a voiced and voiceless orangutan call. *PloS one, 10*(1), e116136.

Lameira, A. R., Hardus, M. E., Mielke, A., Wich, S. A., & Shumaker, R. W. (2016). Vocal fold control beyond the species-specific repertoire in an orang-utan. *Scientific reports, 6*(1), 1-10.

Lameira, A. R., Hardus, M. E., & Wich, S. A. (2012). Orangutan instrumental gesture-calls : Reconciling acoustic and gestural speech evolution models. *Evolutionary Biology*, *39*(3), 415-418.

Lange, B. P. (2011). *Verbal proficiency as fitness indicator : Experimental and comparative research on the evolutionary psychology of language and verbal displays* [PhD Thesis]. University of Cassel.

Lavelli, M., & Fogel, A. (2013). Interdyad differences in early mother–infant face-to-face communication : Real-time dynamics and developmental pathways. *Developmental Psychology*, *49*(12), 2257.

Leavens, D. A. (2004). Manual deixis in apes and humans. *Interaction Studies, 5*(3), 387-408. https://doi.org/10.1075/is.5.3.05lea

Leavens, D. A., Bard, K. A., & Hopkins, W. D. (2017). The mismeasure of ape social cognition. *Animal Cognition*, 1-18. https://doi.org/10.1007/s10071-017-1119-1

Leavens, D. A., & Hopkins, W. D. (1998). Intentional communication by chimpanzees : A cross-sectional study of the use of referential gestures. *Developmental Psychology, 34*(5), 813-822. https://doi.org/10.1037/0012-1649.34.5.813

Leavens, D. A., Hopkins, W. D., & Bard, K. A. (1996). Indexical and referential pointing in chimpanzees (Pan troglodytes). *Journal of Comparative Psychology, 110*(4), 346-353. https://doi.org/10.1037/0735-7036.110.4.346

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*(4), 185-189. https://doi.org/10.1111/j.0963-7214.2005.00361.x

Leavens, D. A., Hostetter, A. B., Wesley, M. J., & Hopkins, W. D. (2004). Tactical use of unimodal and bimodal communication by chimpanzees, Pan troglodytes. *Animal Behaviour*, *67*(3), 467-476. https://doi.org/10.1016/j.anbehav.2003.04.007

Leavens, D. A., Russell, J. L., & Hopkins, W. D. (2010). Multimodal communication by captive chimpanzees (Pan troglodytes). *Animal Cognition*, *13*(1), 33-40. https://doi.org/10.1007/s10071-009-0242-z

Leavens, D. A., Sansone, J., Burfield, A., Lightfoot, S., O'Hara, S., & Todd, B. K. (2015). Putting the "Joy" in joint attention : Affective-gestural synchrony by parents who point for their babies. *Frontiers in Psychology*, *5*, 171-177.

Lecointre, G., & Le Guyader, H. (2001). *Classification phylogénétique du vivant*. Belin. http://afrilib.odinafrica.org/handle/0/24073

Lemasson. (2011). What can forest guenons « tell » us about the origin of language? In A. Vilain, C. Abry, J.-L. Schwartz, & J. Vauclair (Éds.), *Primate Communication and Human Language : Vocalisation, Gestures, Imitation and Deixis in Humans and Non-Humans* (p. 39-70). John Benjamins Publishing.

Lemasson, A., Ouattara, K., Petit, E. J., & Zuberbühler, K. (2011). Social learning of vocal structure in a nonhuman primate? *BMC Evolutionary Biology*, *11*(1), 1-7.

Lerner, R. M. (1998). Developmental contextualism. *Comparative psychology: A handbook*, *894*, 88.

Levinson, S. C. (2006). *On the human 'interactional engine'*. In N. Enfield, & S.C. Levinson (eds), Roots of human sociality. Culture, cognition and human interaction, pp. 39-69. Oxford Berg Publishers

Levinson, S. C. (2016). Turn-taking in human communication–origins and implications for language processing. *Trends in cognitive sciences*, *20*(1), 6-14.

Levinson, S. C., & Holler, J. (2014). The origin of human multi-modal communication. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1651), 20130302.

Lewis, L., & Krupenye, C. (2021). *Theory of Mind in Nonhuman Primates*. PsyArXiv. https://doi.org/10.31234/osf.io/c568f

Liebal, K., & Call, J. (2012). The origins of non-human primates' manual gestures. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1585), 118-128. https://doi.org/10.1098/rstb.2011.0044

Liebal, K., Call, J., & Tomasello, M. (2004). Chimpanzee gesture sequences. *Primates*, *64*, 377-396.

Liebal, K., Call, J., Tomasello, M., & Pika, S. (2004). To move or not to move : How apes adjust to the attentional state of others. *Interaction Studies*, *5*(2), 199-219. https://doi.org/10.1075/is.5.2.03lie

Liebal, K., & Oña, L. (2018). Mind the gap–moving beyond the dichotomy between intentional gestures and emotional facial and vocal signals of nonhuman primates. *Interaction Studies*, *19*(1-2), 121-135.

Liebal, K., Pika, S., & Tomasello, M. (2006). Gestural communication of orangutans (Pongo pygmaeus). *Gesture*, *6*(1), 1-38.

Liebal, K., Schneider, C., & Errson-Lembeck, M. (2019). How primates acquire their gestures : Evaluating current theories and evidence. *Animal cognition*, *22*(4), 473-486.

Lieberman, P. (2012). Vocal tract anatomy and the neural bases of talking. *Journal of Phonetics*, *40*(4), 608-622. https://doi.org/10.1016/j.wocn.2012.04.001

Liszkowski, U., Albrecht, K., Carpenter, M., & Tomasello, M. (2008). Infants' visual and auditory communication when a partner is or is not visually attending. *Infant Behavior and Development*, *31*(2), 157-167. https://doi.org/10.1016/j.infbeh.2007.10.011

Liszkowski, U., Brown, P., Callaghan, T., Takada, A., & Vos, C. de. (2012). A Prelinguistic Gestural Universal of Human Communication. *Cognitive Science*, *36*(4), 698-713. https://doi.org/10.1111/j.1551-6709.2011.01228.x

Liszkowski, U., Carpenter, M., Henning, A., Striano, T., & Tomasello, M. (2004). Twelve-month-olds point to share attention and interest. *Developmental science*, *7*(3), 297-307.

Liszkowski, U., Carpenter, M., & Tomasello, M. (2007). Reference and attitude in infant pointing. *Journal of Child Language*, *34*(1), 1-20. https://doi.org/10.1017/S0305000906007689

Little, E. E., Carver, L. J., & Legare, C. H. (2016). Cultural variation in triadic infant–caregiver object exploration. *Child Development*, *87*(4), 1130-1145.

Locke, J. L. (2006). Parental selection of vocal behavior. *Human Nature*, *17*(2), 155-168.

Locke, J. L. (2009). Evolutionary developmental linguistics : Naturalization of the faculty of language. *Language Sciences*, *31*(1), 33-59.

Locke, J. L. (2017). Emancipation of the voice : Vocal complexity as a fitness indicator. *Psychonomic bulletin & review*, *24*(1), 232-237.

Locke, J. L., & Bogin, B. (2006). Language and life history : A new perspective on the development and evolution of human language. *Behavioral and Brain Sciences*, *29*(3), 259-280.

MacArthur, R. H., & Wilson, E. O. (1967). *The theory of island biogeography*, *vol. 1 Princeton*. NJ: Princeton University Press.

MacNeilage, P. F. (1998). The frame/content theory of evolution of speech production. *Behavioral and brain sciences*, *21*(4), 499-511.

Maestripieri, D., Jovanovic, T., & Gouzoules, H. (2000). Crying and infant abuse in rhesus monkeys. *Child Development*, *71*(2), 301-309.

Maille, A., Engelhart, L., Bourjade, M., & Blois-Heulin, C. (2012). To beg, or not to beg? That is the question: mangabeys modify their production of requesting gestures in response to human's attentional states. *PLoS ONE*, *7*(7), e41197. https://doi.org/10.1371/journal.pone.0041197

Malson, L., & Itard, J. M. G. (1985). *Les enfants sauvages: Mythe et réalité, par Lucien Malson; suivi de Mémoire et rapport sur Victor de l'Aveyron, par Jean Itard*. Union générale d'éditions.

Margiotoudi, K., Marie, D., Claidière, N., Coulon, O., Roth, M., Nazarian, B., Lacoste, R., Hopkins, W. D., Molesti, S., & Fresnais, P. (2019). Handedness in monkeys reflects hemispheric specialization within the central sulcus. An in vivo MRI study in right-and left-handed olive baboons. *Cortex*, *118*, 203-211.

Martin, P., & Bateson, P. (2007). *Measuring Behaviour : An Introductory Guide*. Cambridge University Press.

Martinet, A. (1984). Double articulation as a criterion of linguisticity. *Language Sciences*, *6*(1), 31-38.

Mastin, J. D., & Vogt, P. (2011). Joint attention and vocabulary development : An observational study of Mozambican infants. *Poster Presented at the XIIth Meeting of the International Association for the Study of Child Language, Montréal, Canada*.

Mastin, J. D., & Vogt, P. (2016). Infant engagement and early vocabulary development : A naturalistic observation study of Mozambican infants from 1;1 to 2;1. *Journal of Child Language*, *43*(2), 235-264. https://doi.org/10.1017/S0305000915000148

Matsumoto, D., & Willingham, B. (2006). The thrill of victory and the agony of defeat : Spontaneous expressions of medal winners of the 2004 Athens Olympic Games. *Journal of personality and social psychology*, *91*(3), 568.

Matsumoto, D., & Willingham, B. (2009). Spontaneous facial expressions of emotion of congenitally and noncongenitally blind individuals. *Journal of personality and social psychology*, *96*(1), 1.

Matthews, D., Behne, T., Lieven, E., & Tomasello, M. (2012). Origins of the human pointing gesture : A training study. *Developmental Science*, *15*(6), 817-829. https://doi.org/10.1111/j.1467-7687.2012.01181.x

Maynard-Smith, J., & Harper, D. G. (1995). Animal signals : Models and terminology. *Journal of theoretical biology*, *177*(3), 305-311.

McCune, L. (2021). Laryngeal vocalizations : Grunts as a gateway to language? *Lingua*, 103161.

McCune, L., & Vihman, M. M. (2001). Early phonetic and lexical development. J*ournal of Speech, Language and Hearing Research*, *44*(3), 670-684.

McCune, L., Vihman, M. M., Roug-Hellichius, L., Delery, D. B., & Gogate, L. (1996). Grunt communication in human infants (Homo sapiens). *Journal of Comparative Psychology*, *110*(1), 27.

McCune, L., & Zlatev, J. (2015). Dynamic systems in semiotic development : The transition to reference. *Cognitive Development*, *36*, 161-170.

McNeill, D. (2012). *How language began : Gesture and speech in human evolution*. Cambridge University Press.

Meguerditchian, A., Cochet, H., & Vauclair, J. (2011). From gesture to language Ontogenetic and phylogenetic perspectives on gestural communication and its cerebral lateralization. In A.

Vilain, C. Abry, J.-L. Schwartz, & J. Vauclair (Éds.), *Primate Communication and Human Language : Vocalisation, Gestures, Imitation and Deixis in Humans and Non-Humans* (p. 91-120). John Benjamins Publishing.

Meguerditchian, A., Gardner, M. J., Schapiro, S. J., & Hopkins, W. D. (2012). The sound of one-hand clapping : Handedness and perisylvian neural correlates of a communicative gesture in chimpanzees. *Proceedings of the Royal Society B: Biological Sciences*, *279*(1735), 1959-1966. https://doi.org/10.1098/rspb.2011.2485

Meguerditchian, A., Molesti, S., & Vauclair, J. (2011). Right-handedness predominance in 162 baboons (Papio anubis) for gestural communication : Consistency across time and groups. *Behavioral Neuroscience*, *125*(4), 653-660. https://doi.org/10.1037/a0023823

Meguerditchian, A., Taglialatela, J. P., Leavens, D. A., & Hopkins, W. D. (2014). Why vocal production of atypical sounds in apes and its cerebral correlates have a lot to say about the origin of language. *Behavioral and Brain Sciences*, *37*(6), 565.

Meguerditchian, A., & Vauclair, J. (2006). Baboons communicate with their right hand. *Behavioural Brain Research*, *171*(1), 170-174. https://doi.org/10.1016/j.bbr.2006.03.018

Meguerditchian, A., & Vauclair, J. (2009). Contrast of hand preferences between communicative gestures and non-communicative actions in baboons : Implications for the origins of hemispheric specialization for language. *Brain and Language*, *108*(3), 167-174. https://doi.org/10.1016/j.bandl.2008.10.004

Meguerditchian, A., & Vauclair, J. (2014). Communicative signaling, lateralization and brain substrate in nonhuman primates : Toward a gestural or a multimodal origin of language? *Humana. Mente Journal of Philosophical Studies*, *7*(27), 135-160.

Mehr, S. A., & Krasnow, M. M. (2017). Parent-offspring conflict and the evolution of infant-directed song. *Evolution and Human Behavior*, *38*(5), 674-684.

Mehr, S. A., Singh, M., Knox, D., Ketter, D. M., Pickens-Jones, D., Atwood, S., Lucas, C., Jacoby, N., Egner, A. A., & Hopkins, E. J. et al. (2019). Universality and diversity in human song. *Science*, *366*(6468).

Mesman, J., Minter, T., Angnged, A., Cissé, I. A., Salali, G. D., & Migliano, A. B. (2018). Universality without uniformity : A culturally inclusive approach to sensitive responsiveness in infant caregiving. *Child Development*, *89*(3), 837-850.

Mesman, J., van IJzendoorn, M. H., & Bakermans-Kranenburg, M. J. (2009). The many faces of the Still-Face Paradigm : A review and meta-analysis. *Developmental review*, *29*(2), 120-162.

Meunier, H., Fizet, J., & Vauclair, J. (2013). Tonkean macaques communicate with their right hand. *Brain and language*, *126*(2), 181-187.

Meunier, H., Prieur, J., & Vauclair, J. (2013). Olive baboons communicate intentionally by pointing. *Animal Cognition*, *16*(2), 155-163. https://doi.org/10.1007/s10071-012-0558-y

Meunier, H., & Vauclair, J. (2007). Hand preferences on unimanual and bimanual tasks in white-faced capuchins (Cebus capucinus). *American Journal of Primatology: Official Journal of the American Society of Primatologists*, *69*(9), 1064-1069.

Meunier, H., Vauclair, J., & Fagard, J. (2012). Human infants and baboons show the same pattern of handedness for a communicative gesture. *PLoS ONE*, *7*(3), e33959. https://doi.org/10.1371/journal.pone.0033959

Micheletta, J., Engelhardt, A., Matthews, L. E. E., Agil, M., & Waller, B. M. (2013). Multicomponent and multimodal lipsmacking in crested macaques (Macaca nigra). *American Journal of Primatology*, *75*(7), 763-773.

Micheletti, A. J. (2020). Modelling cultural selection on biological fitness to integrate social transmission and adaptive explanations for human behaviour. *Evolutionary Human Sciences*, *2*.

Molesti, S., & Majolo, B. (2017). Evidence of direct reciprocity, but not of indirect and generalized reciprocity, in the grooming exchanges of wild Barbary macaques (Macaca sylvanus). *American journal of primatology*, *79*(9), e22679.

Molesti, S., Meguerditchian, A., & Bourjade, M. (2020). Gestural communication in olive baboons (Papio anubis) : Repertoire and intentionality. *Animal Cognition*, *23*(1), 19-40.

Moore, R. (2016). Meaning and ostension in great ape gestural communication. *Animal cognition*, *19*(1), 223-231.

Moore, R. (2017). Pragmatics-first approaches to the evolution of language. *Psychological Inquiry*, *28*(2-3), 206-210.

Moreno-Núñez, A., Rodríguez, C., & Miranda-Zapata, E. (2020). Getting away from the point : The emergence of ostensive gestures and their functions. *Journal of child language*, *47*(3), 556-578.

Mozzi, A., Forni, D., Clerici, M., Pozzoli, U., Mascheretti, S., Guerini, F. R., Riva, S., Bresolin, N., Cagliani, R., & Sironi, M. (2016). The evolutionary history of genes involved in spoken and written language : Beyond FOXP2. *Scientific reports*, *6*(1), 1-12.

Murdoch, I. (1962). *The idea of perfection*.

Murillo, E., & Casla, M. (2021). Multimodal representational gestures in the transition to multi-word productions. *Infancy*, *26*(1), 104-122.

Murillo, E., Montero, I., & Casla, M. (2021). On the multimodal path to language : The relationship between rhythmic movements and deictic gestures at the end of the first year. *Frontiers in Psychology*, *12*.

Murphy, S. L., Xu, J., Kochanek, K. D., Curtin, S. C., & Arias, E. (2013). National vital statistics reports. *National vital statistics reports*, *61*(4).

Nakata, T., & Trehub, S. E. (2004). Infants' responsiveness to maternal speech and singing. *Infant Behavior and Development*, *27*(4), 455-464.

Nawroth, C., Ebersbach, M., & von Borell, E. (2013). Are juvenile domestic pigs (Sus scrofa domestica) sensitive to the attentive states of humans?—The impact of impulsivity on choice behaviour. *Behavioural processes*, *96*, 53-58.

Nawroth, C., von Borell, E., & Langbein, J. (2015). 'Goats that stare at men' : Dwarf goats alter their behaviour in response to human head orientation, but do not spontaneously use head direction as a cue in a food-related context. *Animal cognition*, *18*(1), 65-73.

Nettle, D. (2009). Ecological influences on human behavioural diversity : A review of recent findings. *Trends in ecology & evolution*, *24*(11), 618-624.

Nettle, D. (2020). Selection, adaptation, inheritance and design in human culture : The view from the Price equation. *Philosophical Transactions of the Royal Society B*, *375*(1797), 20190358.

Nielsen, M., Haun, D., Kärtner, J., & Legare, C. H. (2017). The persistent sampling bias in developmental psychology : A call to action. *Journal of Experimental Child Psychology*, *162*, 31-38.

Nishida, T., Kano, T., Goodall, J., McGrew, W. C., & Nakamura, M. (1999). Ethogram and ethnography of Mahale chimpanzees. *Anthropological Science*, *107*(2), 141-188.

Northrup, J. B., & Iverson, J. M. (2020). Multimodal coordination of vocal and gaze behavior in mother–infant dyads across the first year of life. *Infancy*, *25*(6), 952-972.

Oller, D. K., Buder, E. H., Ramsdell, H. L., Warlaumont, A. S., Chorna, L., & Bakeman, R. (2013). Functional flexibility of infant vocalization and the emergence of language. *Proceedings of the National Academy of Sciences*, *110*(16), 6318-6323.

Oller, D. K., Griebel, U., & Warlaumont, A. S. (2016). Vocal development as a guide to modeling the evolution of language. *Topics in cognitive science*, *8*(2), 382-392.

Oller, D. K., Ramsay, G., Bene, E., Long, H. L., & Griebel, U. (2021). Protophones, the precursors to speech, dominate the human infant vocal landscape. *Philosophical Transactions of the Royal Society B*, *376*(1836), 20200255.

Orzechowski, S., Wacewicz, S., & Żywiczyński, P. (2016). The Problem of "Modality Transition" in Gestural Primacy Hypothesis in Language Evolution : Towards Multimodal Hypotheses. *Studia Semiotyczne—English Supplement Volume XXVIII*, 112.

Ouattara, K., Lemasson, A., & Zuberbühler, K. (2009). Campbell's monkeys concatenate vocalizations into context-specific call sequences. *Proceedings of the National Academy of Sciences*, *106*(51), 22026-22031. https://doi.org/10.1073/pnas.0908118106

Partan, S., & Marler, P. (1999). Communication goes multimodal. *Science*, *283*(5406), 1272-1273.

Patterson, M. L., & Werker, J. F. (1999). Matching phonetic information in lips and voice is robust in 4.5-month-old infants. *Infant Behavior and Development*, *22*(2), 237-247.

Patterson, M. L., & Werker, J. F. (2003). Two-month-old infants match phonetic information in lips and voice. *Developmental Science*, *6*(2), 191-196.

Penn, D. C., & Povinelli, D. J. (2013). The Comparative Delusion : The "Behavioristic/Mentalistic" Dichotomy in Comparative Theory. In J. Metcalfe & H. S. Terrace (eds), *Agency and joint attention*, (p. 62-81). Oxford University Press

Pepperberg, I. M. (2002). Cognitive and communicative abilities of grey parrots. *Current Directions in Psychological Science*, *11*(3), 83-87.

Perani, D., Saccuman, M. C., Scifo, P., Anwander, A., Spada, D., Baldoli, C., Poloniato, A., Lohmann, G., & Friederici, A. D. (2011). Neural language networks at birth. *Proceedings of the National Academy of Sciences*, *108*(38), 16056-16061.

Perea-García, J. O., Kret, M. E., Monteiro, A., & Hobaiter, C. (2019). Scleral pigmentation leads to conspicuous, not cryptic, eye morphology in chimpanzees. *Proceedings of the National Academy of Sciences*, *116*(39), 19248-19250.

Perelman, P., Johnson, W. E., Roos, C., Seuánez, H. N., Horvath, J. E., Moreira, M. A., Kessing, B., Pontius, J., Roelke, M., & Rumpler, Y. (2011). A molecular phylogeny of living primates. *PLoS genetics*, *7*(3), e1001342.

Piaget, J. (1936). La naissance de l'intelligence chez l'enfant. *Neuchâtel-Paris, Delachaux-Niestlé*.

Pika, S. (2008). Gestures of apes and pre-linguistic human children : Similar or different? *First Language*, *28*(2), 116-140. https://doi.org/10.1177/0142723707080966

Pika, S., & Fröhlich, M. (2019). Gestural acquisition in great apes : The social negotiation hypothesis. *Animal Cognition*, *22*(4), 551-565.

Pika, S., Liebal, K., Call, J., & Tomasello, M. (2005). Gestural communication of apes. *Gesture*, *5*(1-2), 41-56. https://doi.org/10.1075/gest.5.1.05pik

Pika, S., & Mitani, J. (2006). Referential gestural communication in wild chimpanzees (Pan troglodytes). *Current biology: CB*, *16*(6), R191-192. https://doi.org/10.1016/j.cub.2006.02.037

Pinker, S. (1986). *The language instinct*. New York: HarperCollins.

Pougnault, L., Levréro, F., Mulot, B., & Lemasson, A. (2020). Breaking conversational rules matters to captive gorillas : A playback experiment. *Scientific Reports*, *10*(1), 6947. https://doi.org/10.1038/s41598-020-63923-7

Povinelli, D. J., Eddy, T. J., Hobson, R. P., & Tomasello, M. (1996). What Young chimpanzees know about seeing. *Monographs of the Society for Research in Child Development*, *61*(3), 1-152. https://doi.org/10.2307/1166159

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and brain sciences*, *1*(04), 515-526.

Prieur, J., Barbu, S., Blois-Heulin, C., & Lemasson, A. (2020). The origins of gestures and language : History, current advances and proposed theories. *Biological Reviews*, *95*(3), 531-554.

Proops, L., & McComb, K. (2010). Attributing attention : The use of human-given cues by domestic horses (Equus caballus). *Animal cognition*, *13*(2), 197-205.

Pusey, A. E. (1983). Mother-offspring relationships in chimpanzees after weaning. *Animal Behaviour*, *31*(2), 363-377. https://doi.org/10.1016/S0003-3472(83)80055-4

Raymond, M., & Pontier, D. (2004). Is there geographical variation in human handedness? *Laterality*, *9*(1), 35-51. pbh.

Rendall, D., Owren, M. J., & Ryan, M. J. (2009). What do animal signals mean? *Animal Behaviour*, *78*(2), 233-240. https://doi.org/10.1016/j.anbehav.2009.06.007

Rendell, L., & Whitehead, H. (2001). Culture in whales and dolphins. *Behavioral and brain sciences*, *24*(2), 309-324.

Reynold Losin, E. A., Russell, J. L., Freeman, H., Meguerditchian, A., & Hopkins, W. D. (2008). Left Hemisphere Specialization for Oro-Facial Movements of Learned Vocal Signals by Captive Chimpanzees. *PLoS ONE*, *3*(6), e2529. https://doi.org/10.1371/journal.pone.0002529

Rissman, L., Horton, L., Flaherty, M., Senghas, A., Coppola, M., Brentari, D., & Goldin-Meadow, S. (2020). The communicative importance of agent-backgrounding : Evidence from homesign and Nicaraguan Sign Language. *Cognition*, *203*, 104332.

Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. *Annual Reviews of Neuroscience*, *27*, 169-192.

Roberts, A. I., Vick, S.-J., & Buchanan-Smith, H. M. (2013). Communicative intentions in wild chimpanzees : Persistence and elaboration in gestural signalling. *Animal cognition*, *16*(2), 187-196.

Rochat, P., Querido, J. G., & Striano, T. (1999). Emerging sensitivity to the timing and structure of protoconversation in early infancy. *Developmental psychology*, *35*(4), 950.

Rodrigues, E. D., Santos, A. J., Veppo, F., Pereira, J., & Hobaiter, C. (2021). Connecting primate gesture to the evolutionary roots of language : A systematic review. *American Journal of Primatology*, *83*(9), e23313. https://doi.org/10.1002/ajp.23313

Rodríguez, C., Moreno-Núñez, A., Basilio, M., & Sosa, N. (2015). Ostensive gestures come first : Their role in the beginning of shared reference. *Cognitive Development*, *36*, 142-149.

Rosa, H. (2019). *Resonance : A sociology of our relationship to the world*. John Wiley & Sons.

Rosenberg, K., & Trevathan, W. (1995). Bipedalism and human birth : The obstetrical dilemma revisited. *Evolutionary Anthropology: Issues, News, and Reviews*, *4*(5), 161-168.

Rossano, F. (2013). Sequence organization and timing of bonobo mother-infant interactions. *Interaction studies*, *14*(2), 160-189.

Rowe, C. (1999). Receiver psychology and the evolution of multicomponent signals. *Animal behaviour*, *58*(5), 921-931.

Russell, J. L., McIntyre, J. M., Hopkins, W. D., & Taglialatela, J. P. (2013). Vocal learning of a communicative signal in captive chimpanzees, Pan troglodytes. *Brain and Language*, *127*(3), 520-525.

Russon, A., & Andrews, K. (2011). Orangutan pantomime : Elaborating the message. *Biology Letters*, *7*(4), 627-630. https://doi.org/10.1098/rsbl.2010.0564

Salomo, D., & Liszkowski, U. (2013). Sociocultural Settings Influence the Emergence of Prelinguistic Deictic Gestures. *Child Development*, *84*(4), 1296-1307. https://doi.org/10.1111/cdev.12026

Sandler, W., & Lillo-Martin, D. (2006). *Sign language and linguistic universals*. Cambridge University Press.

Savage-Rumbaugh, S., McDonald, K., Sevcik, R. A., Hopkins, W. D., & Rubert, E. (1986). Spontaneous symbol acquisition and communicative use by pygmy chimpanzees (Pan paniscus). *Journal of Experimental Psychology: General*, *115*(3), 211.

Scaife, M., & Bruner, S. J. (1975). The capacity for joint visual attention in the infant. *Nature*, *253*(5489), 265-266. https://doi.org/10.1038/253265a0

Scarantino, A., & Clay, Z. (2015). Contextually variable signals can be functionally referential. *Animal Behaviour*, *100*, e1-e8.

Schacter, D. L., Addis, D. R., & Buckner, R. L. (2008). Episodic Simulation of Future Events. *Annals of the New York Academy of Sciences*, *1124*(1), 39-60. https://doi.org/10.1196/annals.1440.001

Schel, A. M., Townsend, S. W., Machanda, Z., Zuberbühler, K., & Slocombe, K. E. (2013). Chimpanzee alarm call production meets key criteria for intentionality. *PloS one*, *8*(10), e76674.

Schino, G. (2007). Grooming and agonistic support : A meta-analysis of primate reciprocal altruism. *Behavioral Ecology*, *18*(1), 115-120.

Schino, G., & Aureli, F. (2008). Trade-offs in primate grooming reciprocation : Testing behavioural flexibility and correlated evolution. *Biological Journal of the Linnean Society*, *95*(3), 439-446.

Schino, G., di Sorrentino, E. P., & Tiddi, B. (2007). Grooming and coalitions in Japanese macaques (Macaca fuscata) : Partner choice and the time frame reciprocation. *Journal of Comparative Psychology*, *121*(2), 181.

Schlenker, P., Chemla, E., Schel, A. M., Fuller, J., Gautier, J.-P., Kuhn, J., Veselinović, D., Arnold, K., Cäsar, C., Keenan, S., & others. (2016). Formal monkey linguistics. *Theoretical Linguistics*, *42*(1-2), 1-90.

Schlenker, P., Chemla, E., & Zuberbühler, K. (2017). Semantics and pragmatics of monkey communication. In *Oxford Research Encyclopedia of Linguistics*.

Scola, C., Bourjade, M., & Jover, M. (2015). Social interaction is associated with changes in infants' motor activity. *Socioaffective Neuroscience & Psychology*, *5*(1), 28256. https://doi.org/10.3402/snp.v5.28256

Scott, N. M., & Pika, S. (2012). A call for conformity Gesture studies in human and non-human primates. In S. Pika, and K. Liebal (eds.) *Developments in Primate Gesture Research*, Vol. xiii, (p. 147-164). John Benjamins Publishing Group

Scott-Phillips, T. C. (2008). On the correct application of animal signalling theory to human communication. In *The evolution of language* (p. 275-282). World Scientific.

Scott-Phillips, T. C. (2015a). Meaning in animal and human communication. *Animal cognition*, *18*(3), 801-805.

Scott-Phillips, T. C. (2015b). Nonhuman primate communication, pragmatics, and the origins of language. *Current Anthropology*, *56*(1), 56-80.

Searle, J. R. (1962). Meaning and speech acts. *The philosophical review*, *71*(4), 423-432.

Semple, S., Hsu, M. J., & Agoramoorthy, G. (2010). Efficiency of coding in macaque vocal communication. *Biology Letters*, *6*(4), 469-471. https://doi.org/10.1098/rsbl.2009.1062

Servais, V. (2012). Faut-il faire la sociologie des singes? *SociologieS*. http://sociologies.revues.org/4054

Seyfarth, R. M., & Cheney, D. L. (1986). Vocal development in vervet monkeys. *Animal Behaviour*, *34*(6), 1640-1658.

Seyfarth, R. M., Cheney, D. L., & Marler, P. (1980). Vervet monkey alarm calls : Semantic communication in a free-ranging primate. *Animal Behaviour*, *28*(4), 1070-1094.

Shannon, C. E., & Weaver, W. (1949). *The mathematical theory of communication, by CE Shannon (and recent contributions to the mathematical theory of communication), W. Weaver*. University of illinois Press.

Shapiro-Mendoza, C. K., Tomashek, K. M., Anderson, R. N., & Wingo, J. (2006). Recent national trends in sudden, unexpected infant deaths : More evidence supporting a change in classification or reporting. *American journal of epidemiology*, *163*(8), 762-769.

Shepherd, S. V. (2010). Following gaze : Gaze-following behavior as a window into social cognition. *Frontiers in integrative neuroscience*, *4*, 5.

Shepherd, S. V., Deaner, R. O., & Platt, M. L. (2006). Social status gates social attention in monkeys. *Trends in Cognitive Sciences*, *4*, 138-147.

Sherwood, C. C., Holloway, R. L., Gannon, P. J., Semendeferi, K., Erwin, J. M., Zilles, K., & Hof, P. R. (2003). Neuroanatomical basis of facial expression in monkeys, apes, and humans. *Annals of the New York Academy of Sciences*, *1000*(1), 99-103.

Shettleworth, S. J. (2010). Clever animals and killjoy explanations in comparative psychology. *Trends in cognitive sciences*, *14*(11), 477-481.

Shoup-Knox, M. L., Ostrander, G. M., Reimann, G. E., & Pipitone, R. N. (2019). Fertility-dependent acoustic variation in women's voices previously shown to affect listener physiology and perception. *Evolutionary Psychology*, *17*(2), 1474704919843103.

Silk, J. B., Kaldor, E., & Boyd, R. (2000). Cheap talk when interests conflict. *Animal behaviour*, *59*(2), 423-432.

Silk, J. B., Roberts, E. R., Städele, V., & Strum, S. C. (2018). To grunt or not to grunt : Factors governing call production in female olive baboons, Papio anubis. *PloS one*, *13*(11), e0204601.

Simenel, R. (2018). Communiquer avec la nature pour apprendre sa culture : Le rôle de l'iconicité sonore animale dans la communication orale de l'enfant (Maroc/Pérou). *Anthropologica*, *60*(2), 480-493.

Slocombe, K. E., Townsend, S. W., & Zuberbühler, K. (2009). Wild chimpanzees (Pan troglodytes schweinfurthii) distinguish between different scream types : Evidence from a playback study. *Animal cognition*, *12*(3), 441-449.

Slocombe, K. E., Waller, B. M., & Liebal, K. (2011). The language void : The need for multimodality in primate communication research. *Animal Behaviour*, *81*(5), 919-924. https://doi.org/10.1016/j.anbehav.2011.02.002

Slocombe, K. E., & Zuberbühler, K. (2007). Chimpanzees modify recruitment screams as a function of audience composition. *Proceedings of the National Academy of Sciences*, *104*(43), 17228-17233. https://doi.org/10.1073/pnas.0706741104

Smith, L. B., & Thelen, E. (2003). Development as a dynamic system. *Trends in cognitive sciences*, *7*(8), 343-348.

Smuts, B. B., Cheney, D. L., Seyfarth, R. M., & Wrangham, R. W. (2008). *Primate societies*. University of Chicago Press.

Snowdon, C. T., & Hausberger, M. (1997). *Social Influences on Vocal Development*. Cambridge University Press.

Sober, E. (2005). Comparative psychology meets evolutionary biology. In L. Daston & Mitman (Éds.), *Thinking with animals : New perspectives on anthropomorphism* (Columbia University Press, p. 85-99).

Sperber, D., & Wilson, D. (2002). Pragmatics, modularity and mind-reading. *Mind & Language*, *17*(1-2), 3-23.

Spreng, R. N., & Grady, C. L. (2010). Patterns of brain activity supporting autobiographical memory, prospection, and theory of mind, and their relationship to the default mode network. *Journal of cognitive neuroscience*, *22*(6), 1112-1123.

Stamps, J. (2003). Behavioural processes affecting development : Tinbergen's fourth question comes of age. *Animal Behaviour*, *66*(1), 1-13.

Strum, S. C. (2012). Darwin's monkey : Why baboons can't become human. *American journal of physical anthropology*, *149*(S55), 3-23.

Strum, S. S., & Latour, B. (1987). Redefining the social link : From baboons to humans. *Social Science Information*, *26*(4), 783-802. https://doi.org/10.1177/053901887026004004

Suddendorf, T. (1997). The rise of the metamind. In *The descent of mind*. Oxford University Press.

Suddendorf, T., & Corballis, M. C. (2007). The evolution of foresight : What is mental time travel, and is it unique to humans? *Behavioral and Brain Sciences*, *30*(03), 299-313. https://doi.org/10.1017/S0140525X07001975

Syme, K. L., & Hagen, E. H. (2020). Mental health is biological health : Why tackling "diseases of the mind" is an imperative for biological anthropology in the 21st century. *American journal of physical anthropology*, *171*, 87-117.

Taglialatela, J. P., Cantalupo, C., & Hopkins, W. D. (2006). Gesture handedness predicts asymmetry in the chimpanzee inferior frontal gyrus. *NeuroReport*, *17*(9), 923.

Taglialatela, J. P., Russell, J. L., Schaeffer, J. A., & Hopkins, W. D. (2008). Communicative Signaling Activates 'Broca's' Homolog in Chimpanzees. *Current Biology*, *18*(5), 343-348. https://doi.org/10.1016/j.cub.2008.01.049

Taglialatela, J. P., Russell, J. L., Schaeffer, J. A., & Hopkins, W. D. (2011). Chimpanzee vocal signaling points to a multimodal origin of human language. *PloS one*, *6*(4), e18852.

Talmi, D., & Frith, C. D. (2011). Neuroscience, free will, and responsibility. *Conscious will and responsibility: A tribute to Benjamin Libet*, 124-133.

Taylor, D., Dezecache, G., & Davila-Ross, M. (2021). Filling in the gaps : Acoustic gradation increases in the vocal ontogeny of chimpanzees (Pan troglodytes). *American Journal of Primatology*, *83*(5), e23249.

Tempelmann, S., Kaminski, J., & Liebal, K. (2011). Focus on the essential : All great apes know when others are being attentive. *Animal Cognition*, *14*(3), 433-439. https://doi.org/10.1007/s10071-011-0378-5

Terrace, H. S., Petitto, L. A., Sanders, R. J., & Bever, T. G. (1979). Can an ape create a sentence? *Science*, *206*(4421), 891-902.

Theall, L. A., & Povinelli, D. J. (1999). Do chimpanzees tailor their gestural signals to fit the attentional states of others? *Animal Cognition*, *2*(4), 207-214. https://doi.org/10.1007/s100710050041

Thelen, E. (1981). Rhythmical behavior in infancy : An ethological perspective. *Developmental psychology*, *17*(3), 237.

Thomas, C. (1998). *Comment supporter sa liberté*. Payot.

Thornhill, R., & Fincher, C. L. (2013). The comparative method in cross-cultural and cross-species research. *Evolutionary Biology*, *40*(4), 480-493.

Tinbergen, N. (1963). On aims and methods of ethology. *Zeitschrift für tierpsychologie*, *20*(4), 410-433.

Tomasello, M. (1999). Social cognition before the revolution. *Early social cognition: Understanding others in the first months of life*, 301-314.

Tomasello, M. (2006). Why don't apes point? In N. J. Enfield & S. C. Levinson (Éds.), *Roots of human sociality : Culture, cognition and interaction* (p. 506-524). Oxford & New-York.

Tomasello, M. (2008). *Origins of Human Communication*. MIT Press.

Tomasello, M. (2014). *A natural history of human thinking*. Harvard University Press.

Tomasello, M., & Call, J. (2019). Thirty years of great ape gestures. *Animal cognition*, *22*(4), 461-469.

Tomasello, M., Call, J., Nagell, K., Olguin, R., & Carpenter, M. (1994). The learning and use of gestural signals by young chimpanzees : A trans-generational study. *Primates*, *35*(2), 137-154. https://doi.org/10.1007/BF02382050

Tomasello, M., & Carpenter, M. (2007). Shared intentionality. *Developmental Science*, *10*(1), 121-125. https://doi.org/10.1111/j.1467-7687.2007.00573.x

Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Understanding and sharing intentions : The origins of cultural cognition. *Behavioral and Brain Sciences, 28*(05), 675-691. https://doi.org/10.1017/S0140525X05000129

Tomasello, M., Carpenter, M., & Liszkowski, U. (2007). A new look at infant pointing. *Child development, 78*(3), 705-722.

Tomasello, M., George, B. L., Kruger, A. C., Jeffrey, M., Farrar, & Evans, A. (1985). The development of gestural communication in young chimpanzees. *Journal of Human Evolution, 14*(2), 175-186. https://doi.org/10.1016/S0047-2484(85)80005-1

Tomasello, M., Hare, B., Lehmann, H., & Call, J. (2007). Reliance on head versus eyes in the gaze following of great apes and human infants : The cooperative eye hypothesis. *Journal of Human Evolution, 52*(3), 314-320. https://doi.org/10.1016/j.jhevol.2006.10.001

Townsend, S. W., Koski, S. E., Byrne, R. W., Slocombe, K. E., Bickel, B., Boeckle, M., Braga Goncalves, I., Burkart, J. M., Flower, T., Gaunet, F., Glock, H. J., Gruber, T., Jansen, D. A. W. A. M., Liebal, K., Linke, A., Miklósi, Á., Moore, R., van Schaik, C. P., Stoll, S., … Manser, M. B. (2017). Exorcising Grice's ghost : An empirical approach to studying intentional communication in animals. *Biological Reviews, 92*(3), 1427-1433. https://doi.org/10.1111/brv.12289

Trevarthen, C. (1999). Musicality and the intrinsic motive pulse : Evidence from human psychobiology and infant communication. *Musicae scientiae, 3*(1_suppl), 155-215.

Trevarthen, C. (2011). What is it like to be a person who knows nothing? Defining the active intersubjective mind of a newborn human being. *Infant and Child Development, 20*(1), 119-135.

Trevarthen, C., & Aitken, K. J. (2001). Infant intersubjectivity : Research, theory, and clinical applications. *The Journal of Child Psychology and Psychiatry and Allied Disciplines, 42*(1), 3-48.

Trivers, R. L. (1974). Parent-offspring conflict. *Integrative and Comparative Biology, 14*(1), 249-264.

Tronick, E., Als, H., Adamson, L., Wise, S., & Brazelton, T. B. (1978). The infant's response to entrapment between contradictory messages in face-to-face interaction. *Journal of the American Academy of Child psychiatry, 17*(1), 1-13.

Tronick, E. Z., & Cohn, J. F. (1989). Infant-mother face-to-face interaction : Age and gender differences in coordination and the occurrence of miscoordination. *Child development*, 85-92.

Uddin, L. Q., Iacoboni, M., Lange, C., & Keenan, J. P. (2007). The self and social cognition : The role of cortical midline structures and mirror neurons. *Trends in cognitive sciences, 11*(4), 153-157.

Uomini, N. T. (2009). The prehistory of handedness : Archaeological data and comparative ethology. *Journal of Human Evolution, 57*(4), 411-419. https://doi.org/10.1016/j.jhevol.2009.02.012

van der Wel, R. P., Welsh, T., & Böckler, A. (2018). Talking heads or talking eyes? Effects of head orientation and sudden onset gaze cues on attention capture. *Attention, Perception, & Psychophysics, 80*(1), 1-6.

van Duijn, M. J. (2016). *The lazy mindreader : A humanities perspective on mindreading and multiple-order intentionality* [PhD Thesis]. Leiden University.

Vargha-Khadem, F., Watkins, K., Alcock, K., Fletcher, P., & Passingham, R. (1995). Praxic and nonverbal cognitive deficits in a large family with a genetically transmitted speech and language disorder. *Proceedings of the National Academy of Sciences*, *92*(3), 930-933.

Vauclair, J. (2004). Lateralization of communicative signals in nonhuman primates and the hypothesis of the gestural origin of language. *Interaction Studies*, *5*(3), 365-386. https://doi.org/10.1075/is.5.3.04vau

Vigliocco, G., Perniss, P., & Vinson, D. (2014). Language as a multimodal phenomenon : Implications for language learning, processing and evolution. *Philosophical Transactions of the Royal Society, B, 369*, 20130292. http://doi.org/10.1098/rstb.2013.0292

Vygotsky, L. S. (1962). *Thought and language (E. Hanfmann & G. Vakar, trans.)*. Cambridge, ma: mit Press.

Vygotsky, L. S. (1978). *Mind and society : The development of higher mental processes*. Cambridge, MA: Harvard University Press.

Wacewicz, S., & Żywiczyński, P. (2015). Language evolution : Why Hockett's design features are a non-starter. *Biosemiotics*, *8*(1), 29-46.

Wacewicz, S., & Zywiczynski, P. (2017). The multimodal origins of linguistic communication. *Language & Communication, 54*, 1-8.

Waddington, C. H. (2014). *The strategy of the genes*. Routledge.

Wagner, R. H., & Danchin, É. (2010). A taxonomy of biological information. *Oikos*, *119*(2), 203-209.

Waller, B. M., Whitehouse, J., & Micheletta, J. (2017). Rethinking primate facial expression : A predictive framework. *Neuroscience & Biobehavioral Reviews*, *82*, 13-21.

Wallez, C., Schaeffer, J., Meguerditchian, A., Vauclair, J., Schapiro, S. J., & Hopkins, W. D. (2012). Contrast of hemispheric lateralization for oro-facial movements between learned attention-getting sounds and species-typical vocalizations in chimpanzees : Extension in a second colony. *Brain and language*, *123*(1), 75-79.

Wallez, C., & Vauclair, J. (2011). Right hemisphere dominance for emotion processing in baboons. *Brain and Cognition*, *75*(2), 164-169.

Wallez, C., & Vauclair, J. (2013). Human (Homo sapiens) and baboon (Papio papio) chimeric face processing : Right-hemisphere involvement. *Journal of Comparative Psychology*, *127*(3), 237.

Wallez, C., Vauclair, J., & Bourjade, M. (2019). Baboon (Papio anubis) chimeric face processing by human (Homo sapiens) judges : Influence of stimuli complexity on the perception of oro-facial asymmetries. *Journal of Comparative Psychology*, *133*(1), 36.

Wallon, H. (1936). Les origines du caractère chez l'enfant. *The Journal of Nervous and Mental Disease*, *84*(3), 353.

Watts, A. (2005). Éloge de l'insécurité. *Petite bibliothèque Payot*.

West, M. J., King, A. P., & White, D. J. (2003). The case for developmental ecology. *Animal Behaviour*, *66*(4), 617-622.

Wheeler, B. C., & Fischer, J. (2015). The blurred boundaries of functional reference : A response to Scarantino & Clay. *Animal Behaviour*, *100*, e9-e13.

Whiten, A. (1996). When does smart behaviour-reading become mind-reading. *Theories of theories of mind*, 277-292.

Wilcox, S., & Xavier, A. N. (2013). A framework for unifying spoken language, signed language, and gesture. *Todas as Letras-Revista de Língua e Literatura*, *15*(1).

Wilke, C., Kavanagh, E., Donnellan, E., Waller, B. M., Machanda, Z. P., & Slocombe, K. E. (2017). Production of and responses to unimodal and multimodal signals in wild chimpanzees, Pan troglodytes schweinfurthii. *Animal Behaviour*, *123*, 305-316.

Wu, Z., & Gros-Louis, J. (2014). Infants' prelinguistic communicative acts and maternal responses : Relations to linguistic development. *First Language*, *34*(1), 72-90.

Wu, Z., & Gros-Louis, J. (2017). The value of vocalizing : 10-Month-Olds' vocal usage relates to language outcomes at 15 months. *Infancy*, *22*(1), 23-41.

Zuberbühler, K., Arnold, K., & Slocombe, K. E. (2011). Living links to human language. In A. Vilain, C. Abry, J.-L. Schwartz, & J. Vauclair (Éds.), *Primate Communication and Human Language : Vocalisation, Gestures, Imitation and Deixis in Humans and Non-Humans* (p. 13-38). John Benjamins Publishing.

Zywiczynski, P., Gontier, N., & Wacewicz, S. (2017). The evolution of (proto-) language : Focus on mechanisms. *Language Sciences*, *63*, 1-11.