

THÈSE DE DOCTORAT DE L'UNIVERSITÉ PARIS VI - SORBONNE UNIVERSITÉS

ÉCOLE DOCTORALE N°158
CERVEAU-COGNITION-COMPORTEMENT

Spécialité : NEUROSCIENCES COGNITIVES

Présentée et soutenue publiquement par

Élodie CAUVET

Le 28 Juin 2012

Traitement des Structures Syntaxiques dans le Langage et dans la Musique

Thèse dirigée par Christophe PALLIER
préparée dans l'unité de Neuroimagerie Cognitive U992

Jury :

<i>Rapporteurs :</i>	Emmanuel BIGAND Julie FRANCK
<i>Examineur :</i>	Laurent COHEN
<i>Directeur :</i>	Christophe PALLIER
<i>Co-directrice :</i>	Anne CHRISTOPHE

A Dil et Clare,

Résumé

Les phrases ne sont pas de simples suites linéaires de mots. Elles possèdent une structure hiérarchique indispensable à la construction du sens des phrases. Durant mon travail de thèse, je me suis intéressée à cette structure selon différents points de vue : son acquisition par les jeunes enfants, ses représentations cérébrales ainsi que les liens entre les structures hiérarchiques impliquées dans le langage et la musique.

La première partie de ce travail examine dans quelle mesure les mots grammaticaux et la prosodie sont utilisés pour reconnaître les catégories syntaxiques (Nom et Verbe). Dans le cadre de la théorie de l'initialisation syntaxique, ce mécanisme est crucial pour construire le lexique. Nous présentons une expérience comportementale avec des enfants de 18 mois qui montre qu'à cet âge, ils exploitent les déterminants pour reconnaître les noms, et les pronoms pour retrouver les verbes. Trois expériences comportementales chez l'adulte montrent que les indices prosodiques et les mots grammaticaux sont utilisés, en temps réel, pour déduire la catégorie grammaticale de mots connus et pseudomots. En intégrant nos données comportementales à celles de la littérature, nous proposons un modèle s'appuyant sur la prosodie et les mots grammaticaux pour contraindre l'accès lexical. Ce modèle rend également compte de l'encodage de nouveaux mots dans le lexique.

Le deuxième volet de ce travail s'intéresse aux bases neurales du traitement de la structure syntaxique.

Dans une première expérience en imagerie par résonance magnétique fonctionnelle (IRMf), des locuteurs français ont entendu des structures coordonnées, par exemple « Le chien et la vache et le lapin ou la poule ». La variation de la taille des frontières prosodiques entre les éléments provoque un changement des structures hiérarchiques syntaxique et prosodique sous-jacentes. Une analyse classique des images fonctionnelles permet d'isoler les régions cérébrales impliquées dans la construction des représentations prosodique et syntaxique. Grâce à une méthode de classification sur ces mêmes images, nous montrons qu'elles sont labellissables selon leur structure.

D'autres domaines que le langage peuvent être représentés sous forme de structures hiérarchiques comme par exemple la musique. Ainsi, l'existence d'un réseau cérébral représentant la structure syntaxique en langage permet de supposer qu'une partie de ce réseau pourrait être recrutée par la musique. Nous avons donc construit des stimuli musicaux et linguistiques comparables dont la complexité variait paramétriquement. Des tests comportementaux sur ces stimuli ont été menés avec des individus français non musiciens mais aussi des personnes d'une tribu amazonnienne non-exposées à la musique classique occidentale. Les résultats révèlent que les différents niveaux de structure sont perceptibles indépendamment de la tonalité.

Ces mêmes stimuli ont été utilisés dans une expérience d'IRMf sur des sujets musiciens et non-musiciens français. Nous montrons que les aires inféro-frontale, temporale

et temporo-pariétale sont recrutées pour le traitement de la structure syntaxique linguistique. Un sous-ensemble de ces régions est impliqué dans le traitement de la structure musicale particulièrement chez les musiciens. Un réseau de l'hémisphère gauche constitué des parties orbitalis et triangularis du gyrus frontal inférieur, du pôle temporal et de la portion postérieure du sillon temporal supérieur est recruté pour le traitement de la structure syntaxique en langage et dans une moindre mesure en musique.

Table des matières

Introduction	1
1 Mots grammaticaux et prosodie dans l'accès au lexique et leur implication dans l'acquisition du langage	7
1.1 Assigner un sens aux mots : un problème difficile	8
1.2 Résoudre le problème de la multiplicité des sens	10
1.2.1 Indices Socio-pragmatiques	11
1.2.2 Initialisation syntaxique	11
1.2.3 Ce dont les enfants disposent effectivement et à partir de quand	14
1.2.4 Les catégories grammaticales et le squelette syntaxique	18
1.3 Traitement des mots de grammaticaux dans l'acquisition du langage . .	20
1.4 Expérience 1 : exploitation en temps réel des mots grammaticaux pour contraindre l'accès au lexique chez des enfants de 18 mois	23
1.5 Accès au lexique	44
1.6 Expérience 2 : exploitation des indices syntaxiques pour contraindre l'accès au lexique en temps réel chez l'adulte	47
1.7 Expérience 3 : Traitement syntaxique sans lexique chez l'adulte : utilisation des mots de fonction en Jabberwocky	51
1.8 Expérience 3 bis : Traitement syntaxique sans lexique : Utilisation des indices prosodiques. La Corrélation entre prosodie et syntaxe	61
1.9 Conclusion	65
2 Décoder les activations cérébrales dues aux structures syntaxiques et prosodiques	69
2.1 La Coordination : un nouvel outil pour étudier la prosodie et la syntaxe	69
2.2 Expérience 4 : Bases neurales du traitement des structures syntaxique et prosodique	72
3 Structures hiérarchiques en langage et en musique	101
3.1 Un peu de théorie	101
3.2 Les données de la littérature	107
3.3 La théorie de partage des ressources syntaxiques	108
3.4 Étudier la syntaxe en langage	113
4 Structure hiérarchique en langage et en musique : étude comportementale de validation des stimuli	119
4.1 Expérience 5 : Structure en Langage et en Musique chez des non-musiciens Français	120
4.2 Perception consciente et traitement automatique de la structure en musique	128
4.2.1 Expérience 6 : Perception consciente, notation des stimuli musicaux	129

4.2.2	Expérience 7 : Traitement automatique, discrimination des stimuli musicaux	138
4.2.3	Discussion générale	146
5	Réseaux cérébraux de la structure syntaxique : langage et musique	149
5.1	Etat de l'art en IRMf	149
5.2	Expérience 8 : Bases neurales du traitement syntaxique en langage et en musique chez les musiciens et non musiciens	154
5.3	Conclusion	203
6	Discussion générale	207
6.1	Mots grammaticaux et prosodie dans l'accès lexical	208
6.2	Mots grammaticaux et prosodie dans l'acquisition lexicale	210
6.3	Les réseaux neuronaux recrutés pour la construction de la structure syntaxique sont-ils spécifiques au langage ?	212
A	Stimuli utilisés dans le chapitre 1	221
A.1	Phrases de test utilisées dans l'expérience comportementale 1 sur des enfants de 18 mois	221
A.2	Phrases pour l'expérience comportementale 2	223
A.3	Phrases utilisées dans l'expérience comportementale 3 : influence des mots grammaticaux dans le calcul des catégories syntaxique	228
A.4	Stimuli utilisés dans l'expérience 3bis : influence de la prosodie dans le calcul de la catégorie syntaxique en Jabberwocky	233
B	Stimuli langagier pour l'expérience Muslang	237
C	Un tout petit peu de théorie de détection du signal.	239
	Bibliographie	241

Table des figures

1.1	Différents points de vue d'une même scène visuelle. A gauche le point de vue de l'adulte qui pointe le cactus. A droite le point de vue de l'enfant qui assiste au pointage et qui essaie de deviner quel objet est pointé; la tasse, la tarte ou le cactus pourraient aussi bien être désignés par le pointage.	9
1.2	Graphique de l'onde acoustique désignant le mot <i>chien</i>	10
1.3	Résultats de l'expérience (Gillette et al., 1999) étudiant l'impact des différents indices pour deviner le sens d'un mot.	13
1.4	Hierarchie prosodique selon (Nespor and Vogel, 1986).	16
1.5	Un modèle des premières étapes de traitement de la parole. Figure adaptée de (Christophe et al., 2008)	21
1.6	Schéma CHT	28
1.7	Reconnaissance de la cible en fonction du type de phrase entendu	32
1.8	Reconnaissance de la cible en fonction de la familiarisation des mots grammaticaux	33
1.9	Reconnaissance de la cible en temps réel	35
1.10	Modèle où les prédictions syntaxiques contraignent l'accès lexical. En haut, le contexte syntaxique d'apparition du mot balle est correct. En bas, le contexte est incorrect.	40
1.11	Modèle de traitement de la parole en temps réel : L'accès au lexique est effectué à partir des informations phonologiques seulement. La syntaxe jouerait un rôle second. En haut, la prédiction de ce modèle dans le cas où le mot balle apparaît dans un contexte syntaxique correct; en bas, dans un contexte incorrect.	42
1.12	Exemple d'un essai. Les couleurs ont été omises dans ce schéma. Figure issue de (Dahan et al., 2000). Parmi les quatre objets présentés, on trouve la cible, par exemple <i>bouton</i> , un compétiteur qui commence par la même syllabe <i>bouteille</i> et deux distracteurs.	45
1.13	Modèle représentant l'intégration des informations syntaxiques dans l'accès lexical.	46
1.14	Pourcentage moyen de fausse alarme pour les différentes catégories de compétiteur. Lorsque la cible à détecter était un nom, le graphe est représenté en gris foncé et lorsque c'était un verbe, en gris clair. La catégorie grammaticale du compétiteur est congruente si le compétiteur et la cible à détecter partagent la même catégorie, elle est incongruente si le compétiteur et la cible sont de catégories syntaxiques différentes.	49

1.15	Pourcentage moyen de réponse de type nom ou verbe donnée en fonction du type de cible entendue. Les sujets ont répondu dans 91% des cas lorsque la cible à détecter était un nom et que la phrase contenait effectivement cette cible dans une position de nom, c'est le taux de réussite. Cette même cible Nom a été détectée dans 13% des cas lorsqu'elle était présentée dans un contexte de verbe (du type <i>je le + cible</i>), c'est le taux de fausse alarme pour la cible Nom. De la même façon pour la cible verbe, les sujets l'ont détecté dans 96% des cas lorsqu'ils entendaient effectivement cette cible dans un contexte de verbe et ils l'ont repéré dans 15% des cas lorsqu'elle était présentée dans un contexte de Nom du type <i>Adverbe le + cible</i> . Les sujets repèrent bien mieux la cible dans les phrases où elle est présentée dans le contexte syntaxique correct.	57
1.16	Réponses en temps réel en fonction de la cible et de la phrase entendue	59
1.17	Appui bouton en fonction du type de prosodie entendue	63
1.18	Modèle représentant l'intégration des informations syntaxiques dans l'accès lexical.	68
2.1	Différents arbres possibles selon l'emplacement et la taille des frontières prosodiques.	71
2.2	Différentes structures possibles pour une coordination de 4 éléments avec leurs arbres syntaxiques associées.	73
2.3	Carte des régions cérébrales plus actives pour les phrases contenant 2 frontières de groupe intonational par rapport aux hrases en contenant une seule.	75
2.4	Stimuli présentés dans l'expérience de (Santi and Grodzinsky, 2010). Dessin expérimental croisant le mouvement et l'enchâssement central. . . .	77
2.5	Résultats de l'expérience de (Santi and Grodzinsky, 2010), adapté. Régions dont l'activité est réduite par la répétition d'aspects particuliers de la syntaxe : le mouvement et l'enchâssement.	77
2.6	Exemple d'image vue	80
2.7	Les 3 arbres syntaxiques utilisés	81
2.8	Les 6 conditions créées	81
2.9	Différentes situations visuelles rencontrées dans l'expérience	82
2.10	Résultat de la tâche comportementale	88
2.11	Analyse univariée : cartes statistique de sensibilité aux structures prosodiques et syntaxiques	89
2.12	Les matrices de confusion pour la classification selon les 3 niveaux de structures syntaxiques (en haut à gauche), selon les deux niveaux de prosodie (en haut à droite), et les classifications des arbres syntaxiques deux à deux (en bas)	91
2.13	Cartes résultant A - des 3 classifications syntaxiques et B - des classifications prosodiques	93
3.1	Organisation hiérarchique des actions requises pour faire du café	102

3.2	Exemple de représentation de la structure de groupe	103
3.3	Exemple de représentation de la structure métrique	103
3.4	Représentation des structures de groupe et métrique	104
3.5	Représentation de la time-span reduction	105
3.6	Représentation de la prolongational reduction	106
3.7	Schéma des relations fonctionnelles entre le traitement syntaxique de la musique et du langage : Modèle SSIRH	109
3.8	Stimuli et résultats de l'expérience de (Koelsch et al., 2005)	111
3.9	Exemple de stimuli utilisé dans l'expérience (Fedorenko et al., 2009)	111
3.10	Exemple de stimuli utilisés dans l'expérience (Devauchelle, 2008; Pallier et al., 2011)	114
3.11	Schéma de création des stimuli utilisés dans l'expérience de langage écrit (Devauchelle, 2008; Pallier et al., 2011)	115
3.12	Simulation du modèle simple d'activité neurale évoquée par le traitement des constituants de taille variable (Pallier et al., 2011)	116
3.13	Résultats obtenus dans l'expérience linguistique à l'écrit (Pallier et al., 2011)	117
4.1	Schéma des différents niveaux de structure des stimuli musicaux	121
4.2	Schéma rotatif de concaténation des différents fragments	122
4.3	Schéma des étapes de création des stimuli musicaux	123
4.4	Résultats de la tâche d'estimation du niveau de structure	126
4.5	Carte du territoire des Mundurucus	128
4.6	Photo de Mundurucus passant l'expérience dans un village.	130
4.7	Stimuli en langue Mundurukù utilisé pour l'entraînement	131
4.8	Icônes utilisées pour sélectionner la réponse.	132
4.9	Moyenne des notes données aux phrases par les Mundurucus et les français	133
4.10	Comparaison des 2 groupes de sujet sur la perception de la structure.	134
4.11	Effet de la tonalité sur la perception de la structure.	135
4.12	Moyenne des notes données par tous les participants aux seuls stimuli en Do majeur.	136
4.13	Moyennes des notes données selon les groupes et la tonalité des fragments	137
4.14	Moyenne des réponses correctes (en haut) et d'prime associés (en bas) en séparant français et mundurucus	140
4.15	Effet du groupe sur la discrimination.	141
4.16	Effet de la tonalité sur la discrimination.	142
4.17	Moyennes des d-prime calculés en séparant par groupes et par tonalité des fragments	144
4.18	Effet de la structure sur les temps de réaction	145
5.1	Régions sensibles à la taille des constituants en langage écrit (Pallier et al., 2011)	150
5.2	Voxels plus activés pour les phrases bien construites par rapport aux phrases mélangées.	151

5.3	Amplitude moyenne pour chaque type de stimulus pour la région d'intérêt du pôle temporal gauche (ATL) et du gyrus fronto-temporal inférieur gauche (IFG). Adapté de (Rogalsky et al., 2011).	151
5.4	Sections coronales montrant le signal BOLD, activations et désactivations, comparant la musique bien construite et la musique mélangée (Levitin and Menon, 2003)	153
5.5	Schéma des différents niveaux de structure	155
5.6	Carte statistique associée aux activations globales en langage et en musique	159
5.7	Carte d'activation cérébrale étudiant les différences d'activation entre le langage et la musique	160
5.8	Carte d'activation cérébrale séparant les activations pour le langage de celles de la musique	162
5.9	Cartes statistiques des différences entre musiciens et non-musiciens pour la musique	163
5.10	Carte statistique représentant les zones sensibles à la structure en langage	166
5.11	Régions sensibles à la structure en langage	166
5.12	Carte statistique représentant les zones sensibles à la structure en langage avec les décours temporels correspondant aux différentes conditions dans différentes régions	167
5.13	Zones sensibles à la structure en musique	169
5.14	Amplitude des réponses hémodynamiques en fonction de la taille des constituants syntaxiques	170
5.15	Interaction entre la structure et le domaine	171
5.16	Les 8 régions d'intérêt pour la réponse positive à la structure	172
5.17	Les région d'intérêt pour la réponse négative à la structure	173
5.18	Sensibilité à la structure en langage oral des voxels sélectionnés au niveau individuel par la méthode Fedorenko	176
5.19	Sensibilité à la structure en musique des voxels sélectionnés au niveau individuel par la méthode Fedorenko	178
5.20	Traitement de la structure en musique musiciens vs non-musiciens	180
5.21	Carte statistique étudiant l'impact de la compétence musicale sur le traitement de la structure en musique	181
5.22	Taille de l'effet de structure en musique des musiciens et des non-musiciens, approche par ROIs individuelles	183
5.23	Comparaison anatomique entre musiciens et non-musiciens : analyse VBM	187
5.24	Carte statistique représentant les zones sensibles à la structure en langage dans l'expérience en langue écrite (Devauchelle, 2008) à un seuil corrigé au niveau du voxel à $p < 0.05$ FDR.	190
5.25	Effet de grammaticalité. (Snijders et al., 2009)	190
5.26	Illustration des aires de Brodmann 44 et 45 : respectivement IFG opercularis et IFG triangularis.	193
5.27	Exemples de stimuli utilisés dans l'expérience (Makuuchi et al., 2009) : Variation de la structure hiérarchique (proposition avec enchaînement central) et de la mise en mémoire (distance sujet-verbe).	194

5.28	Superposition du réseau du mode par défaut et du réseau sensible négativement à la structure chez les musiciens	202
5.29	Carte des territoires de la zone inféro-frontale et sa connectivité. Exemple pour un sujet. En rouge, une région à cheval entre BA45 et BA47, en bleue l'IFG triangularis (BA45) et en vert l'IFG opercularis BA44 avec les fibres de matières blanches qui les connectent au cortex temporal et pariétal. Image provenant de (Friederici, 2009)	203
6.1	Modèle d'accès au lexique dans lequel les indices phonologiques et syntaxiques contraignent simultanément la recherche de l'item.	210
6.2	Régions cérébrales impliquées dans le traitement de la structure syntaxique en langage et en musique. 1. IFG orbitalis (BA47) ; 2. IFG triangularis (BA45) ; 3. Pôle temporal ; 4. Sillon temporal supérieur postérieur.	214

Liste des tableaux

1.1	Exemple de stimuli pour les 3 contextes (correct, incorrect et distracteur)	26
1.2	Analyse acoustique des stimuli	27
1.3	Fréquence des suites de mots entendues par les enfants (corpus CHILDES (Demuth and Tremblay, 2008)) et leur corrélation avec le pourcentage d'orientation de la tête vers la cible. La corrélation a été calculée pour tous les essais mais aussi seulement pour les essais utilisant des mots de fonction non-utilisés dans la phase d'entraînement.	39
2.1	Effets fixes dans un modèle d'analyse à effet mixte des scores et des temps de réponse	88
4.1	Exemple de phrases pour chaque niveau de structure	124
5.1	Exemple de phrases pour chaque niveau de structure	155
5.2	Différents cas de figure pour les différences d'activation entre le langage et la musique. Les couleurs des cellules sont celles visibles sur la figure 5.8 page 162	161
5.3	Table statistique des zones dont l'activité est corrélée avec la taille des constituants en langage	165
5.4	Régions dont l'activité augmente plus avec la taille des constituants en langage par rapport à la musique	170
5.5	Pression acoustique des stimuli des différents domaines exprimé par leur RMS : root mean square (ou valeur efficace/ moyenne quadratique) en Pascal.	185
5.6	Volume sonore en Sone	185
5.7	Table statistique de la VBM comparant la taille de la matière grise entre les musiciens et les non-musiciens.	186
C.1	Différents types de réponses possible	239

Introduction

Comment appréhender en tant qu'adulte l'ensemble des problèmes auxquels sont confrontés les bébés pour apprendre leur langue maternelle ? En tant que locuteur, l'ensemble des concepts et des règles qui régissent une langue nous semble très naturel. Pour entrevoir ce à quoi est confronté un bébé, vous pouvez vous imaginer téléporté soudainement à l'autre bout de la galaxie parmi des individus totalement différents des êtres humains. Pour communiquer avec eux, il vous faut d'abord trouver le signal qui porte du sens. Ils pourraient communiquer en se frottant les oreilles ou en tapant du pied. A l'endroit où vous vous trouvez, vous vous apercevez qu'ils émettent des sons avec leur bouche ; ils font preuve de communication orale. Mais comment comprendre de quoi ils parlent ? Le monde qui vous entoure est complètement nouveau, votre référentiel a changé et vous vous rendez compte que très peu de vos expériences humaines pourront vous servir sur cette nouvelle planète. En vous imaginant dans cette situation, vous n'avez qu'un petit aperçu de ce à quoi sont confrontés les très jeunes enfants tous les jours. Cet exemple ne fait qu'esquisser le problème puisqu'en tant que locuteur vous avez déjà acquis un ensemble de représentations qui vous permettent de lier un élément physique, un son, à un concept à savoir son sens. Vous possédez donc un avantage non négligeable par rapport au jeune enfant. Pourtant, à partir du monde qui l'entoure et peut-être de ses prédispositions génétiques, l'enfant va apprendre sa langue maternelle jusqu'à la maîtriser presque complètement à l'âge de 3 ans, laps de temps relativement rapide si l'on considère la complexité de la tâche. Plusieurs compétences vont être acquises plus ou moins simultanément : identification des sons pertinents de la langue maternelle, segmentation de la parole en mots, création du lexique interne, création d'une structure entre les mots.

Chaque langue utilise un répertoire de sons différents. L'enfant apprend à représenter les mots à l'aide des sons pertinents pour sa langue maternelle, appelés phonèmes. Cette spécialisation pour les sons de sa langue maternelle a lieu durant la première année de vie

(Werker and Tees, 1984)¹. Ainsi les petits français vont comprendre que les sons « ou » /u/ comme dans *roue* et « u » /y/ comme dans *rue* sont différents puisqu'ils permettent de distinguer 2 mots dans leur langue alors qu'au contraire les petits anglais perdront cette sensibilité. Très vite ils ne distingueront plus ces deux sons. Après l'âge d'un an, les anglais ne perçoivent plus cette différence. Un autre exemple concernant les consonnes est la différence entre les sons r /ʀ/ et l /l/ que les enfants français conservent a contrario de leurs homologues japonais pour qui les deux sons appartiendront à la même catégorie de phonèmes. Ils ne pourront plus distinguer ce son au sein des mots ; les mots *Hello* et *héros* ne sont pas distinguables.

A l'oral, contrairement à l'écrit où les mots sont séparés les uns des autres par des espaces, le flux de parole est continu. L'enfant doit extraire les mots du signal de parole. Dans cette tâche, il peut s'aider de l'intonation spécifique de la parole dirigée vers les enfants qui présente des caractéristiques particulières. La mélodie et le rythme de la parole - prosodie - sont exacerbés. En effet, un adulte qui s'adresse à un enfant a naturellement tendance à parler plus lentement, à marquer plus fortement l'intonation, à utiliser un vocabulaire plus simple. La quantité et la qualité de la variation de la prosodie varie selon les individus et selon les cultures, cependant elle existe toujours. Il est probable que ce type de discours rende plus saillants les indices qui aident l'enfant à segmenter et comprendre le discours.

Une fois les mots segmentés, l'enfant va assigner un sens à chacun des mots, il va construire son lexique interne. Puis, il va combiner les mots selon les règles propres à sa langue - la syntaxe - pour former des phrases et transmettre des informations. Plusieurs problèmes doivent donc être solutionnés dans l'apprentissage du langage : isoler les sons pertinents, repérer les mots, leur assigner un sens et les combiner. Leur résolution serait concomitante et non linéaire. En effet, les informations récoltées par ces différentes étapes serviraient aux autres processus. Par exemple, pour apprendre le sens des mots, la structure syntaxique serait une aide non négligeable et inversement

1. La première étude à avoir mis en évidence cette perte de sensibilité pour les phonèmes qui ne sont pas propres à la langue maternelle

le lexique serait utile à l'élaboration de la structure syntaxique. En effet, les objets, qu'ils soient concrets ou abstraits, sont le plus souvent représentés par des noms et les actions par des verbes (Pinker, 1984). Donc savoir qu'un mot inconnu est un nom ou un verbe, permet de restreindre le sens possible de ce nouveau mot. Ces catégories syntaxiques seraient donc particulièrement utiles au développement du lexique chez les enfants. Elles présentent aussi un avantage dans le traitement de la parole chez les adultes. Cette information permettrait d'optimiser l'accès au lexique, notamment pour résoudre des ambiguïtés liées à l'homophonie. Par exemple, le mot *porte* peut aussi bien être un nom *la porte* qu'un verbe *je porte*. L'utilisation des catégories syntaxiques restreindrait la recherche lexicale à une partie seulement du lexique et les locuteurs ne confondraient jamais ces deux concepts. Plusieurs indices sont combinés pour calculer ces catégories syntaxiques comme par exemple la prosodie et les morphèmes grammaticaux. Les mots ou morphèmes grammaticaux désignent tous les petits mots qui ne portent pas de sens directement et permettent de structurer les langues comme par exemple les déterminants, les pronoms, les conjugaisons etc. Dans la première partie de ce travail de thèse, je me suis particulièrement intéressée à ces éléments qui structurent les langues ainsi qu'aux informations prosodiques dans le calcul des catégories syntaxiques. J'ai étudié l'importance de ces indices dans l'accès au lexique.

Le plus souvent, l'interprétation d'une phrase ne pose pas de problème. Cependant dans une phrase comme « La femme voit l'homme avec des lunettes », plusieurs interprétations sont possibles : soit l'homme porte des lunettes, soit la femme voit avec des lunettes. Il est possible de construire deux structures syntaxiques différentes qui décrivent les deux interprétations possibles. Le constituant [à lunettes] peut ainsi être lié ou au constituant [*l'homme*] ou au verbe *regarde*. Selon le constituant auquel il est relié, le sens sera distinct. Ces deux interprétations reposent sur des structures syntaxiques différentes. La majeure partie du temps, des phrases de ce type ne nous posent aucun problème ; le contexte sémantique lève l'ambiguïté. La prosodie des phrases est une autre source d'information qui clarifie le sens. En effet, l'endroit et la taille des frontières entre

les mots fournit des informations acoustiques à l'interlocuteur qui lui permettent de décider à quel élément syntagme (ou constituant syntaxique) relier l'élément ambigu. Pour distinguer les deux structures syntaxiques, les adultes se servent des indices prosodiques et du contexte pragmatique. Les linguistes ont émis l'hypothèse que les constituants des phrases sont organisés selon une structure enchâssée. Plusieurs éléments composent la syntaxe (Pinker and Jackendoff, 2005) :

- un système va récupérer les mots de la phrase pour **construire des constituants syntaxiques de façon hiérarchique**. Par exemple, il est possible de décomposer la phrase « le lapin mange la carotte » de la façon suivante : [lapin] est un groupe nominal contenu dans un groupe de déterminant [Le lapin] ; [mange la carotte] est un groupe verbal et ce groupe verbal contient lui-même un groupe déterminant [la carotte] qui contient un groupe nominal [carotte]. Les groupes syntaxiques peuvent contenir d'autres groupes syntaxiques, c'est donc un système récursif qui contient des constituants enchâssés.
- le deuxième système va **ordonner les mots et les constituants** au sein d'une phrase. Par exemple en allemand, le verbe conjugué se trouve toujours en 2^{ème} position dans la phrase. Ce système place donc les constituants à l'endroit de la phrase qui convient selon les règles propres à chaque langue. La plupart des langues du monde n'ont pas un ordre des mots aussi strict qu'en français ou en anglais. Le plus souvent pour ordonner les groupes syntaxiques, les règles privilégiées pour l'ordre d'apparition des constituants syntaxiques dans la phrase correspondent à la topicalisation et au focus (IL faudra sans doute mieux expliquer cela ou enlever cette dernière phrase).
- le troisième système va **accorder** les adjectifs et verbes (en genre, en nombre et aux autres caractéristiques de classification) avec les noms auxquels ils sont syntaxiquement reliés. Les noms sont accordés avec leur déterminant. Par exemple dans la phrase « Les lapins mangent les belles carottes » le verbe *manger* est conjugué au pluriel puisque le sujet était au pluriel et l'adjectif *beau* a été mis au

féminin et au pluriel pour s'accorder avec *carotte* qui a été mis au pluriel pour s'accorder avec son déterminant.

- le quatrième système **marque les cas avec les déclinaisons**. Un nom va se voir adjoindre une terminaison selon le rôle grammatical qu'il joue dans une phrase. Ce rôle du nom peut être calculé par rapport au verbe, à une préposition ou à un autre nom. Par exemple en allemand ou en latin, si un nom est sujet du verbe, il portera une déclinaison du nominatif, s'il est complément d'objet direct de ce verbe, ce même nom sera mis à l'accusatif : « Das Kaninchen isst den Salat » (le lapin mange la salade), *das* est l'article neutre au nominatif, *den* est l'article masculin à l'accusatif.

Les structures enchâssées ne sont pas spécifiques au langage, on les retrouve dans plusieurs domaines comme la planification de l'action, les mathématiques, l'informatique ou encore la musique. L'organisation structurale de la musique a été décrite en utilisant les formalismes empruntés à la linguistique, notamment par Leonard Bernstein avec son cycle de conférence en 1973 à Harvard *Unanswered questions* ou encore *A generative Theory of tonal music* (Lerdahl and Jackendoff, 1983). Étant donné les similitudes entre les descriptions des structures musicales et linguistiques, les réseaux neuronaux recrutés dans le traitement de la structure syntaxique linguistique le sont-ils pour la structure musicale ?

Le travail présenté dans cette thèse s'intéresse aux compétences syntaxiques (traitement de stimuli structurés) de l'être humain depuis son développement chez le jeune enfant jusqu'à son traitement cérébral ainsi que sa spécificité au langage. Comprendre les différentes étapes de mise en place de la syntaxe nous permettra de mieux l'étudier y compris chez l'adulte. Le premier chapitre s'attachera à comprendre dans quelle mesure les mots grammaticaux et la prosodie sont utilisés pour effectuer le calcul des catégories grammaticales (Nom et Verbe) des mots connus et inconnus. Nous proposons un modèle d'accès lexical rendant compte des données comportementales obtenues chez l'adulte et les bébés. Le deuxième chapitre de ce travail s'intéressera aux bases neu-

rales du traitement des structures hiérarchiques. Nous nous appliquerons à décoder les activités cérébrales dues au traitement des structures prosodique et syntaxique. Puis, nous présenterons un paradigme permettant de tester de façon comparable la structure hiérarchique en langage et en musique. Nous testerons la perception de la structure dans nos stimuli sur des français non musiciens mais aussi sur des individus d'une tribu amazonienne peu exposés à la musique classique occidentale dans le chapitre 4. Ensuite, nous utiliserons ces mêmes stimuli dans une expérience en IRMf pour tester si les mêmes réseaux cérébraux réagissent à la structure en constituants dans les deux domaines. Cette étude, présentée dans le dernier chapitre, comportera un groupe de musiciens et de non musiciens dans le but de tester l'influence de l'expertise sur le traitement fonctionnel de la structure, particulièrement en musique.

Mots grammaticaux et prosodie dans l'accès au lexique et leur implication dans l'acquisition du langage

Sommaire

1.1	Assigner un sens aux mots : un problème difficile	8
1.2	Résoudre le problème de la multiplicité des sens	10
1.2.1	Indices Socio-pragmatiques	11
1.2.2	Initialisation syntaxique	11
1.2.3	Ce dont les enfants disposent effectivement et à partir de quand . .	14
1.2.4	Les catégories grammaticales et le squelette syntaxique	18
1.3	Traitement des mots de grammaticaux dans l'acquisition du langage	20
1.4	Expérience 1 : exploitation en temps réel des mots grammati- caux pour contraindre l'accès au lexique chez des enfants de 18 mois	23
1.5	Accès au lexique	44
1.6	Expérience 2 : exploitation des indices syntaxiques pour contraindre l'accès au lexique en temps réel chez l'adulte	47
1.7	Expérience 3 : Traitement syntaxique sans lexique chez l'adulte : utilisation des mots de fonction en Jabberwocky	51

1.8	Expérience 3 bis : Traitement syntaxique sans lexique : Utilisation des indices prosodiques. La Corrélation entre prosodie et syntaxe	61
1.9	Conclusion	65

Cette première partie s'attache à mieux comprendre comment les enfants réussissent à assigner un sens aux mots et en quoi la structure syntaxique et la prosodie sont des aides précieuses dans cette tâche. Une première expérience comportementale testera avec des enfants de 18 mois le traitement des catégories syntaxiques nom et verbe pour contraindre l'accès lexical. Puis, trois expériences comportementales chez l'adulte étudieront l'influence des mots grammaticaux et de la prosodie dans le calcul de ces catégories. Enfin, nous intégrons nos données et celles de la littérature dans un modèle d'accès au lexique tenant compte simultanément des indices syntaxiques, prosodiques et phonémiques.

1.1 Assigner un sens aux mots : un problème difficile

Il n'est pas forcément évident d'imaginer la difficulté que représente l'association entre un mot et son sens puisque chaque enfant apprend sa ou ses langue(s) maternelle(s) sans effort par exposition passive au langage autour de lui. On imagine intuitivement que les enfants apprennent leurs premiers mots en combinant les objets pointés par les adultes avec les sons prononcés simultanément. La connaissance de premiers mots de vocabulaire a été montrée dès 6-9 mois ([Bergelson and Swingley, 2012](#)). Ensuite, ces premiers mots aident à la déduction de nouveaux mots. De cette façon, les petits humains associent la forme sonore du mot à la cible, le mot acquiert alors un sens. Cependant dans ce type de situations, qui ne représente qu'un petit pourcentage de ce à quoi il est confronté, l'enfant se heurte déjà à deux grandes difficultés. La première est d'identifier le concept, objet ou action, dont il est question. En effet, le point de vue de l'adulte qui pointe est différent de celui de l'enfant puisqu'ils se trouvent à des endroits différents



FIGURE 1.1 – Différents points de vue d’une même scène visuelle. A gauche le point de vue de l’adulte qui pointe le cactus. A droite le point de vue de l’enfant qui assiste au pointage et qui essaie de deviner quel objet est pointé ; la tasse, la tarte ou le cactus pourraient aussi bien être désignés par le pointage.

comme représenté sur les images 1.1.

En outre, on pourra objecter que ce type d’interaction utilisant le pointage, même s’il est fréquent, ne permet d’apprendre que des mots imageables, des événements qui se passent effectivement en face de l’enfant. Or, il existe des mots qui ne sont pas imageables comme par exemple *l’envie*, *regarder*, *voir* ou *penser*. De plus, il existe des cultures dans lesquelles les adultes ne s’adressent pas directement aux enfants ce qui ne les empêche pourtant pas d’apprendre à parler. Il faut donc qu’ils soient capables d’utiliser d’autres sources d’informations que le pointage seul. Restons néanmoins dans le cas simple, d’actions ou d’objets imageables.

Une fois que l’enfant a restreint la scène dont il est question parmi tout son environnement, la deuxième difficulté est de trouver le sens du mot. En effet, il existe une multiplicité de sens possibles pour la forme sonore du mot entendu. Prenons un exemple : « Regarde, le petit chien ». L’enfant essaie d’apprendre le mot chien. Rien dans la forme sonore du mot chien (représentée sur la figure 1.2 page suivante) ne permet

de déduire qu'il s'agit de l'animal à quatre pattes : la relation entre le son et le sens est arbitraire. La suite de sons /ʃj ɛ/ *chien* ne symbolise pas l'animal. De plus, les scènes visuelles sont souvent ambiguës : même si vous pointez du doigt vers un chien tout en disant « voilà un chien », l'enfant doit faire l'hypothèse que vous parlez de la catégorie d'animal à quatre pattes et non du fait qu'il court, qu'il est noir ou doux par exemple. Se pose donc le problème de la multiplicité des sens soulevé par le philosophe américain Willard Van Orman Quine en 1960 (Quine, 1960). Il est raisonnable de penser, qu'il existe des indices pour restreindre le sens possible d'un mot. Tout d'abord, il nous faut comprendre quels indices permettent à l'enfant d'associer la forme sonore d'un mot avec son sens.

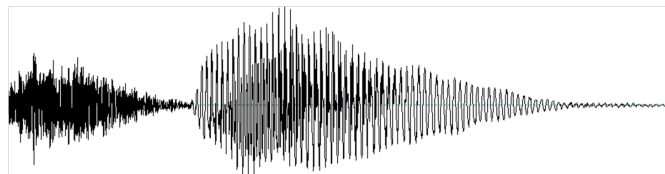


FIGURE 1.2 – Graphique de l'onde acoustique désignant le mot *chien*.

1.2 Résoudre le problème de la multiplicité des sens

De nombreuses expériences utilisant des approches diverses ont testé différents types d'indice. Historiquement, deux hypothèses s'opposent, même si les divergences tendent à s'amoinrir avec les expériences les plus récentes. Une première hypothèse postule que les enfants utilisent leurs capacités sociales et particulièrement la socio-pragmatique des situations pour développer le langage (Tomasello, 1992; Tomasello and Akhtar, 1995). Cette approche constructiviste postule que les facultés d'analyses du langage comme la syntaxe ne seront apprises que tardivement à partir de régularités observées dans la langue. La deuxième hypothèse postule que les enfants acquerraient rapidement des compétences syntaxiques qui leur serviraient à contraindre le sens des mots (Landau and Gleitman, 1985; Gleitman, 1990). Une série d'expériences est présentée ci-dessous confirmant l'utilisation d'indices socio-pragmatiques et morphosyntaxiques.

1.2.1 Indices Socio-pragmatiques

Nous savons que les enfants utilisent une première série d'indices, liés au contexte social et pragmatique qui se crée dans une situation d'interaction, de dialogue. Un de ces indices correspond à l'attention conjointe. Il s'agit de la capacité à analyser la direction du regard du locuteur pour deviner ce dont il parle. Entre 10 et 12 mois, les enfants sont ainsi capables d'analyser les indices présents dans le contexte pour apprendre des mots nouveaux.

Par exemple, Baldwin et ses collègues (Baldwin et al., 1996) ont comparé deux situations d'apprentissage. Dans la première, une expérimentatrice, assise à côté de l'enfant, jouait avec lui avec des objets nouveaux, par exemple des dinosaures de différentes sortes, et disait en même temps : « oh, un *modi* ! tu as vu, c'est un *modi* t un *modi* !, etc. », tout en regardant alternativement le jouet et l'enfant, donc en lui montrant qu'elle parlait du dinosaure. Dans la seconde situation, le téléphone sonnait dès que l'expérimentatrice s'asseyait ; elle se levait pour répondre, puis, parlant au téléphone et tournant le dos à l'enfant, elle énonçait exactement la même série de phrases. Dans les deux cas, l'attention de l'enfant était tournée vers le nouveau jouet. Dans la phase de test, les expérimentateurs ont montré que seuls les enfants de la première situation avaient appris que *modi* désignait le dinosaure. Ceux de la situation du téléphone n'avaient rien appris du tout ; ce qui est normal, puisqu'ils n'avaient aucune raison de penser que l'expérimentatrice au téléphone parlait de l'objet auquel ils prêtaient attention. Les jeunes enfants peuvent donc identifier visuellement les situations les plus susceptibles de leur apprendre de nouveaux mots.

1.2.2 Initialisation syntaxique

Outre l'utilisation de ces indices dits socio-pragmatiques, Lila Gleitman, Barbara Landau et leurs collègues (Gleitman, 1990) ont suggéré que les très jeunes enfants pourraient exploiter l'ensemble des structures syntaxiques dans lesquelles un mot nouveau se produit pour déduire en partie son sens. Cette hypothèse d'initialisation syntaxique,

a été développée pour expliquer les étonnantes compétences linguistiques d'une petite fille aveugle (Landau and Gleitman, 1985). En effet, aucun retard de langage n'a été observé ; le vocabulaire de cette petite fille était normal dès l'âge de deux ans. De plus, elle connaissait même la différence subtile entre les verbes *voir* et *regarder*, alors que, ne voyant pas, elle était privée d'une grande partie des informations concernant la scène relative à l'apparition de ces mots. En outre, aucun des parents n'avait donné de définition explicite de ces deux verbes. Grâce aux enregistrements audio de ce que cette petite fille avait entendu au cours de son apprentissage du langage, les auteurs ont mis en évidence que le seul indice qui différait, était les structures syntaxiques dans lesquelles apparaissaient les deux mots. Certaines étaient compatibles avec un verbe de perception passive, *voir*, mais pas avec un verbe de perception active, *regarder*, et vice versa. Par exemple, il est possible de dire « je vois que le chien gratte la terre », mais pas « *je regarde que le chien gratte la terre ». Plus généralement, un verbe de transfert tel *donner* ou *prendre* peut avoir jusqu'à trois groupes nominaux comme arguments *quelqu'un donne quelque chose à quelqu'un d'autre*, tandis qu'un verbe de pensée, tel *penser* ou *croire*, prend une proposition entière comme complément *je pense que*. Ces relations entre la structure syntaxique et le sens du verbe se retrouvent dans toutes les langues du monde et elles semblent symboliser le sens de ce verbe (). Ainsi, l'hypothèse d'initialisation syntaxique repose sur l'idée que les très jeunes enfants exploiteraient le contexte syntaxique dans lequel ils entendent des mots nouveaux pour leur donner un sens. Ce qui explique pourquoi, bien qu'elle soit privée des informations visuelles, la petite fille aveugle avait un vocabulaire normal : la structure syntaxique des mots favorisait son apprentissage de mots nouveaux.

Cette hypothèse a d'abord été validée sur des adultes. Dans une série d'expériences (Gillette et al., 1999), des sujets devaient deviner le sens d'un mot-mystère, défini par sa catégorie grammaticale à savoir un verbe, dans différentes conditions fournissant des types d'indices distincts :

- de courts films muets représentant des situations d'interaction entre une mère et

- son enfant âgé de 18 à 24 mois avec un bip présenté au moment où le mot était prononcé ;
- une liste écrite des mots de contenu qui étaient prononcés en co-occurrence avec le verbe ;
 - un contexte syntaxique écrit dans lequel le verbe apparaissait, c'est-à-dire en présentant les phrases contenant le verbe mais dans lesquelles les mots de contenu avaient été remplacés par des pseudo-mots ;
 - des phrases écrites contenant le verbe ;
 - une combinaison des éléments précédents.

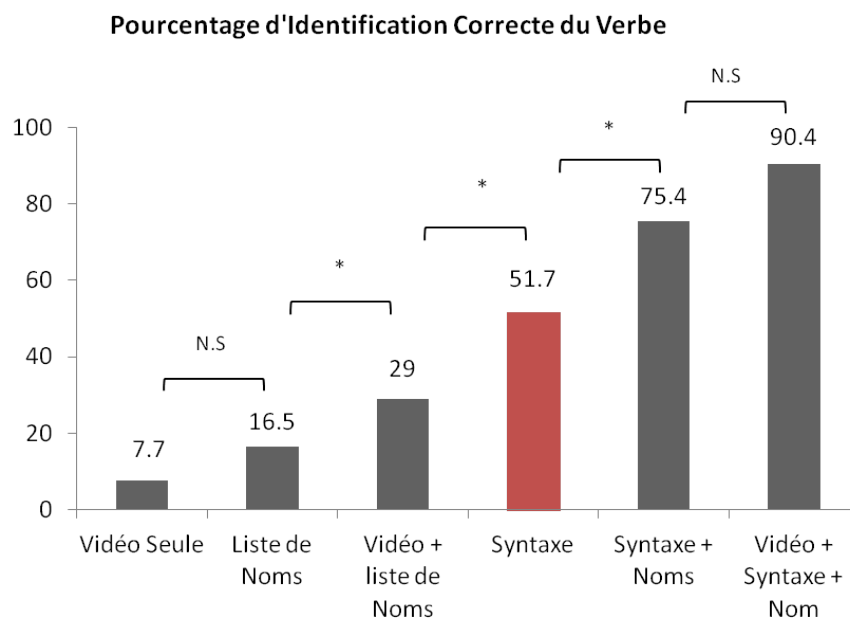


FIGURE 1.3 – Résultats de l'expérience (Gillette et al., 1999) étudiant l'impact des différents indices pour deviner le sens d'un mot. Adapté depuis l'article.

Les sujets ne participaient qu'à une seule condition et avaient accès à plusieurs situations pour un même mot. La figure 1.3 présente les résultats de cette expérience. Les auteurs montrent que les informations visuelles seules ne conduisent pas à une bonne déduction du sens puisque moins de 10% des mots sont découverts. Par contre, les informations syntaxiques seules, sans informations lexicales, ni visuelles, entraînent un bon taux d'identification du mot, plus de 50% des cas. En ajoutant le lexique, la tâche

devient encore plus aisée puisque 75% des mots sont identifiés. La syntaxe est donc à un bon indice pour déduire le sens des mots, au moins des verbes. Pour comprendre en quoi consiste la syntaxe dans cette expérience, il faut souligner que les adultes disposaient de plusieurs indices grammaticaux :

- la catégorie syntaxique de la cible : ils savaient qu'ils cherchaient le sens d'un verbe, et ont donc pu chercher le nombre et le type d'argument que ce verbe pouvait prendre au sein des structures syntaxiques dans lesquelles apparaissaient le verbe ;
- l'ordre des mots ;
- la morphologie verbale et les mots grammaticaux : il s'agit de tous les petits mots qui ne portent pas directement de sens, on les appelle mot de la classe fermé.

L'accès au cadre syntaxique dans lequel apparaît le mot-cible est crucial dans cette expérience pour la découverte de son sens, bien plus que l'accès aux informations visuelles. Découvrir le sens des mots serait donc plus facile si les enfants disposaient déjà d'informations syntaxiques. Cette initialisation syntaxique (syntactic bootstrapping) soulève cependant des questions cruciales :

- Comment initialiser cet apprentissage syntaxique si pratique pour découvrir le sens des mots si on ne dispose pas d'informations lexicales ?
- Quelles sont les informations nécessaires à la construction d'une ébauche de syntaxe ?
- Ces informations sont-elles disponibles et utilisables par de très jeunes enfants ?

1.2.3 Ce dont les enfants disposent effectivement et à partir de quand

Beaucoup d'adultes s'adressent aux jeunes enfants de façon particulière : l'intonation est gaie, le rythme de parole ralenti, le vocabulaire et la structure des phrases se simplifient. On sait que cette parole dirigée vers les enfants attire leur attention (Fernald, 1985; Cooper et al., 1997), notamment à cause de son contenu affectif. De plus, son débit lent et sa mélodie exagérée rendent plus saillantes les caractéristiques proso-

diques de la langue maternelle, et peuvent aider les enfants, par exemple, à découper les phrases en mots (Nelson et al., 1989). En effet, l’une des sources d’information que les enfants utilisent pour découvrir les mots dans les phrases est l’intonation et le rythme de la parole, c’est-à-dire la prosodie. Lorsqu’on parle, aux enfants ou à d’autres adultes, on ne produit pas une suite de syllabes toutes équivalentes sur un ton monocorde. Au contraire, les mots sont regroupés dans de petites unités d’intonation. Par exemple, une phrase comme *le petit garçon a mangé une pomme* sera produite selon deux unités, [*le petit garçon*] et [*a mangé une pomme*].

La prosodie ou mélodie des phrases désigne la variation de rythme et d’intonation produite par les locuteurs qui permettent à son auditeur de segmenter la phrase en unité de sens. La prosodie est organisée de façon hiérarchique c’est-à-dire qu’il existe des unités prosodiques de tailles différentes qui s’imbriquent les unes dans les autres comme des poupées russes. Le plus haut niveau sera donc l’énoncé qui contiendra plusieurs groupes intonationnels qui contiendront eux-mêmes des groupes phonologiques etc. La figure 1.4 page suivante présente un schéma de la hiérarchie prosodique selon la théorie de (Nespor and Vogel, 1986).

Voilà un exemple de structure prosodique contenant deux groupes intonationnels (GI : délimités par les crochets bleus) et 5 groupes phonologiques (GP : délimités par les crochets rouges).

Exemple de Structure prosodique :

« [[Le petit chien]_{GP} [a mangé]_{GP} [le gros os]_{GP}]_{GI} [[que lui a donné]_{GP} [son maître]_{GP}]_{GI}. »

Nous nous intéresserons particulièrement ici aux groupes phonologiques et intonationnels puisque ce sont ceux dont les frontières corréleront avec celles des groupes syntaxiques. Prenons un exemple de phrase plus simple :

Structure prosodique

« [[Le petit chien]_{GP} [a mangé]_{GP} [un gros os]_{GP}]_{GI} »

Structure syntaxique (GN : groupe nominal, GV : groupe verbal)

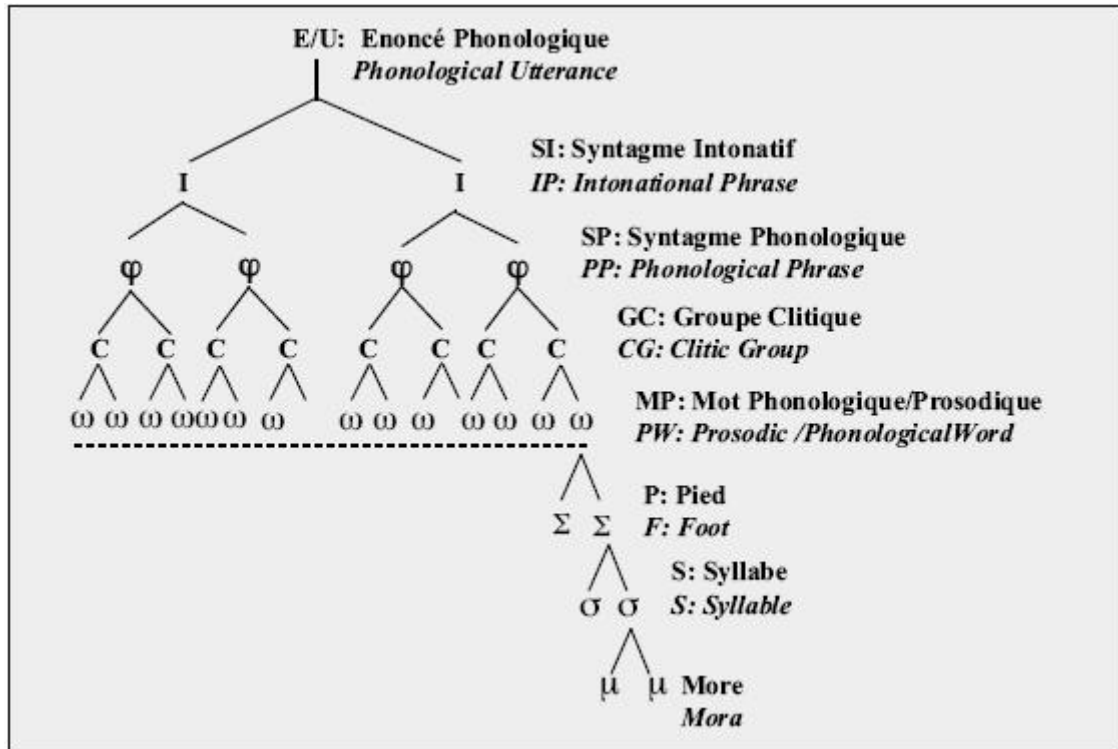


FIGURE 1.4 – Hiérarchie prosodique selon (Nespor and Vogel, 1986).

« [Le petit chien]_{GN} [a mangé [un gros os]_{GN}]_{GV} »

La corrélation entre frontières prosodiques et syntaxiques, même si ce n'est pas une union parfaite, est déjà suffisante pour repérer les unités de sens. Ainsi, le découpage en unités prosodiques réalisés à partir des informations acoustiques a été impliqué dans la reconnaissance de mot chez les enfants ((Gout et al., 2004) à 10 mois et (Millotte et al., 2010) à 16 mois) et chez les adultes (Millotte et al., 2008).

Dès l'âge de 9 mois, les enfants repèrent les frontières prosodiques majeurs des phrases (frontière de groupe intonational) (Juszyk et al., 1992), mais aussi les frontières plus petites de « groupe phonologique » (Gerken et al., 1994). D'autres expériences utilisant la technique de conditionnement de l'orientation du regard (Gout et al., 2004; Millotte et al., 2010) révèlent que les enfants dès 10 mois perçoivent ces unités prosodiques et les exploitent pour découvrir les frontières de mots. Dans l'expérience de Millotte, des enfants francophones de 16 mois ont été entraînés à tourner la tête

vers la gauche lorsqu'ils entendaient un mot spécifique, par exemple *balcon*. En effet, chaque fois qu'ils entendaient ce mot et s'ils regardaient sur leur gauche, là où se trouvait les haut-parleurs, alors ils étaient récompensés par un petit nounours qui s'éclairait et s'animait. S'ils ne regardaient pas en direction des haut-parleurs, rien ne se passait et l'expérience continuait. Grâce à ce paradigme de conditionnement de l'orientation du regard, les enfants ont été amenés à tourner la tête sur un mot cible bisyllabique *balcon* ou monosyllabique *bal* selon le groupe d'enfant. Le comportement des enfants a ensuite été testé en réponse à des phrases de deux types :

1. [Le joli balcon] [cloisonnait la terrasse] : contenant l'item bisyllabique ;
2. [Le dernier bal] [conclura la saison] : contenant l'item monosyllabique suivi d'un mot commençant par la deuxième syllabe du mot bisyllabique.

Les enfants entraînés sur *balcon* tournaient plus la tête sur les phrases de type (1) tandis que ceux entraînés sur *bal* tournaient plus la tête pour les phrases (2). Les enfants sont donc capables d'utiliser les frontières prosodiques pour segmenter les phrases en unités de sens et discriminer deux items phonologiquement identiques.

Pour aller plus loin et segmenter les unités prosodiques comme *le joli balcon* en mots, les enfants utilisent plusieurs stratégies complémentaires. L'une d'elles consiste à identifier les mots, ou morphèmes grammaticaux, à savoir les articles, auxiliaires, terminaisons verbales, etc. Les mots grammaticaux, par exemple les déterminants, les pronoms, les auxiliaires etc., sont dits de classe fermée et servent à structurer les phrases. Ils sont opposés aux mots de contenu ou de classe ouverte qui correspondent à tous les mots qui contiennent un sens comme les noms, les verbes, les adjectifs etc. Plusieurs expériences ont montré que les enfants reconnaissaient déjà les mots grammaticaux de leur langue maternelle avant leur premier anniversaire (Morgan et al., 1996; Shi et al., 1998; Höhle and Weissenborn, 2003; Shi and Lepage, 2008). Ces morphèmes grammaticaux possèdent des propriétés phonologiques communes : ils ne sont pas accentués, contrairement aux mots de contenu, la durée de leur voyelle est plus courte que celle des mots de contenu et ils sont plus courts (Shi et al., 1998). De plus, ils sont extrêmement fréquents et ap-

paraissent en bordure d'unité prosodique. Ces propriétés, conservées dans les différentes langues du globe¹, les rendent particulièrement saillants et facilement reconnaissables. A 17 mois, les enfants sont capables d'utiliser la fréquence d'apparition des mots pour déduire les mots grammaticaux dans une langue inconnue et d'attribuer un sens aux mots qui ne sont pas susceptibles d'être des mots grammaticaux c'est-à-dire les mots les moins fréquents (Hochmann et al., 2010).

De plus, (Gerken and McIntosh, 1993; Kedar et al., 2006) ont montré que les enfants de 24 et 28 mois sont meilleurs pour montrer une image se référant à un mot cible lorsque celui-ci est précédé d'un vrai article plutôt qu'un mot grammatical inventé. Avec un paradigme de mesure de mouvement oculaire, (Zangl and Fernald, 2007) ont montré également une préférence des enfants de 18 mois pour les mots précédés par un déterminant correct par rapport à un mot grammatical inventé. Les mots grammaticaux sont donc utiles aux enfants pour reconnaître les mots connus. Ils pourraient également servir à la catégorisation syntaxique des nouveaux mots. Les objets, qu'ils soient concrets ou abstraits, sont le plus souvent représentés par des noms et les actions par des verbes (Pinker, 1984). Donc, savoir qu'après un article vient un nom et après un pronom vient un verbe serait intéressant pour assigner une catégorie syntaxique à un mot inconnu et pouvoir restreindre le sens de ce mot à un objet ou une action.

1.2.4 Les catégories grammaticales et le squelette syntaxique

En 2007, Bernal et collègues ont montré que les enfants se servaient du contexte grammatical d'un mot pour le catégoriser (Bernal et al., 2007). Elles ont appris un mot inventé à des enfants de 23 mois soit dans une forme verbale « daser » en utilisant des phrases comme « Regarde : elle dase ! », soit dans une forme nominale « une dase » avec des phrases comme « Regarde : une dase ! » en associant ces phrases à un objet animé sur un écran. Deux scènes contenant l'objet montré précédemment ont été présentées ensuite aux enfants : une dans laquelle l'objet effectuait l'action familière et l'autre dans

1. comme l'a montré une étude cross-linguistique turc et mandarin (Shi et al., 1998) dont les propriétés acoustiques sont très différentes

laquelle l'objet réalisait une nouvelle action. Les enfants devaient ensuite montrer celle qui *dasait*. Les enfants entraînés avec le pronom devant le non-mot ont significativement plus pointé vers l'action familière tandis que les enfants entraînés avec l'article suivi du non-mot ont préférentiellement montré l'action nouvelle (préférence pour la nouveauté). Les enfants de 23 mois déduisent un sens différent pour un même mot selon le contexte syntaxique dans lequel il est employé. Les mots grammaticaux sont déjà analysés avant deux ans pour catégoriser les mots inconnus puisque les enfants savent qu'après un déterminant doit se trouver un nom, tandis qu'un pronom précède un verbe. Le nombre de sens possibles de ce nouveau mot est donc restreint à un objet ou une action selon le type de mot grammatical qui le précède, respectivement un déterminant ou un pronom.

Pour mieux comprendre le processus que pourrait effectuer les enfants en entendant une phrase, il est possible de représenter le « squelette syntaxique » (Christophe et al., 1997; Christophe et al., 2008) de la phrase en utilisant les indices morphosyntaxiques et prosodiques dont disposent les enfants. Prenons un exemple, « Le petit chien a mangé un os ». La prosodie² permet de découper la phrases en trois unités de sens ou groupes phonologiques : [le petit chien] [a mangé] [un os]. Les frontières prosodiques qui existent entre chaque groupe corrélerent avec d'autres indices comme la morphologie verbale et les mots grammaticaux qui permettent de spécifier le contenu de chacun de ces groupes de la façon suivante : [le petit chien]_{GN} [a mangé]_{GV} [un gros os]_{GN} (GN : groupe nominal ; GV : groupe verbal). Ce découpage de la phrase ne correspond pas exactement avec sa structure syntaxique mais elle s'en rapproche.

Structure syntaxique de la phrase :

[le petit chien]_{GN} [a mangé] [un gros os]_{GN}]_{GV}.

Les frontières des groupes prosodiques et syntaxiques sont corrélées même s'il n'y a pas identité entre les deux types de structure. De nombreuses frontières syntaxiques ne sont pas marquées du tout par la prosodie. Par exemple, pour la phrase « [Il [mange vite]_{AP}]_{VP}]_{IP} » la frontière syntaxique la plus importante est entre *il* et *mange* et elle n'est pas marquée par une frontière prosodique. Cependant lorsqu'il existe une

2. Soulignons que la prosodie est particulièrement marquée dans la parole dirigée aux enfants

frontière prosodique alors il existe également une frontière syntaxique. Un enfant en cours d'acquisition lexicale pourrait donc construire un tel squelette syntaxique sans pour autant connaître le sens des mots. La phrase devient donc :

[Le V W]_{GN} [a X-é]_{GV} [un Y Z]_{GN}

L'enfant sera ainsi amené à chercher un objet pour V W et Y Z, alors qu'il cherchera une action à associer au mot X.

La figure 1.5 page suivante schématise le traitement de la structure partielle proposé par (Christophe et al., 2008) qui pourrait être effectué par les enfants en cours d'acquisition lexicale. Dans une première étape, les enfants réaliseraient l'analyse du signal acoustique en se basant sur les indices prosodiques et phonétiques pour segmenter le signal en unité de sens. Les enfants pourraient ensuite extraire les morphèmes grammaticaux et les mots de contenu déjà présent dans leur lexique (Bergelson and Swingley, 2012) ; les mots grammaticaux aidant à la détection des mots de contenu (par exemple (Shi and Lepage, 2008)). La combinaison de ces différents indices conduirait à la formation d'une représentation syntaxique de la phrase qui peut être partielle si certains mots de contenu ne sont pas encore encodé dans le lexique des enfants.

1.3 Traitement des mots de grammaticaux dans l'acquisition du langage

Plusieurs expériences montrent que les enfants entre deux et trois ans sont capables d'utiliser la structure argumentale des verbes et la morphologie inflectionnelle pour inférer le sens de nouveaux mots (voir (Fisher, 1996; Fisher et al., 2010; Naigles and Kako, 1993; Naigles, 1996; Yuan and Fisher, 2009) en anglais, et (Göksun et al., 2008) en ture).

Les mots grammaticaux sont utilisés pour segmenter les mots de contenu adjacents (Shi and Lepage, 2008) et pour calculer la structure syntaxique en temps réel dès deux ans (Bernal et al., 2010). De plus, plusieurs expériences montrent que les jeunes enfants sont capables d'exploiter ces petits mots pour contraindre le sens des mots aux alentours

1.3. Traitement des mots de grammaticaux dans l'acquisition du langage²¹

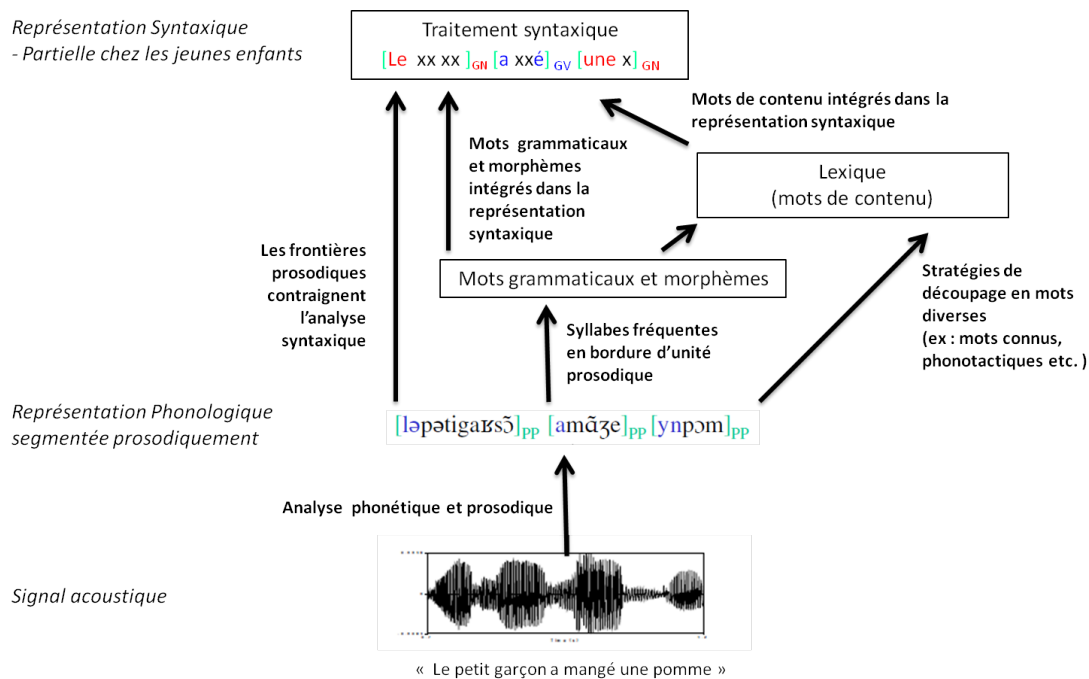


FIGURE 1.5 – Un modèle des premières étapes de traitement de la parole. Figure adaptée de (Christophe et al., 2008)

de deux ans (Fisher et al., 2006; Hall et al., 2003; Waxman and Booth, 2001). Trois expériences récentes montrent que les jeunes enfants associent un nouveau mot présenté dans un contexte de verbe à une action plutôt qu'à un objet ((Bernal et al., 2007) en français à 23 mois présentée plus haut, (Oshima-Takane et al., 2010) en japonais à 20 mois, (Waxman et al., 2009) en anglais à 24 mois). Un même mot inconnu placé dans des contextes syntaxiques différents reçoit un sens différent. De plus, même dans des contextes d'homophonie, Veneziano et Parisse (Veneziano and Parisse, 2010) montrent que des enfants francophones de deux ans distinguent des noms et des verbes homophones connus du type *une ferme/je ferme* selon le contexte syntaxique dans lequel ces mots apparaissent.

Des travaux avec des enfants plus jeunes montrent que les 14 et 16 mois s'attendent à trouver des noms après les déterminants mais qu'ils ne sont pas encore capables d'utiliser les pronoms pour prédire l'apparition des verbes ((Höhle et al., 2004) en allemand, (Shi and Melancon, 2010) en français). L'expérience en français utilisait un paradigme

de fixation visuelle. Les chercheuses ont familiarisé des enfants de 14 mois avec deux pseudo-mots *mige* et *cræle*. La moitié des enfants ont entendu les deux pseudo-mots dans des contextes de noms, c'est-à-dire précédé d'un déterminant (*ton* ou *des*) tandis que l'autre moitié les a entendus dans un contexte de verbe, précédé d'un pronom (*je* ou *il*). Tous les enfants ont été testés avec les mêmes mots précédés de mots de fonction non-utilisés dans l'habituation (*le* ou *tu*). Ainsi deux nouveaux contextes syntaxiques corrects et incorrects ont été créés. Les enfants familiarisés avec le contexte de nom ont écouté plus longtemps les énoncés incorrects précédés d'un pronom ; les enfants ont été surpris lorsque le mot habituellement précédé d'un déterminant apparaissait à la suite d'un pronom. En revanche, les enfants habitués au contexte de verbe ne montrent pas de préférence d'écoute. Ils n'ont donc pas réussi à déduire le contexte d'apparition du nouveau mot. Dans la même optique, deux équipes ont montré que les enfants anglophones de dix-huit mois utilisent les déterminants de leur langue en temps réel pour reconnaître les noms (Kedar et al., 2006; Zangl and Fernald, 2007)³.

Pour résumer, les déterminants commencent à être classés ensemble et associés avec les noms dès l'âge de quatorze mois. D'un autre côté, le lien entre les pronoms et les verbes n'a pas été montré chez les enfants avant 20 mois (Bernal et al., 2007; Oshima-Takane et al., 2010; Waxman et al., 2009).

Plusieurs raisons peuvent rendre compte du fait que les verbes soient plus compliqués à apprendre comparés aux noms. D'une part, ils sont conceptuellement plus complexes ; d'autre part, ils apparaissent dans des contextes plus variés (voir la discussion de (Höhle et al., 2004) et (Shi and Melancon, 2010)). Néanmoins, même les jeunes enfants apprennent quelques verbes (Fenson et al., 1994) et il est possible qu'ils aient déjà assimilé des contextes d'apparition de verbes.

3. Avec un paradigme de regard préférentiel, les enfants regardent plus longtemps vers l'image du nom cible lorsque celui-ci est précédé d'un déterminant par rapport à un mot de fonction incorrect comme *and/et* ou un mot inventé comme *el* (Kedar et al., 2006). Ils regardent aussi plus longtemps à une image congruente lorsque le nom cible est précédé d'un déterminant en comparaison d'un article inventé (Zangl and Fernald, 2007).

1.4 Expérience 1 : exploitation en temps réel des mots grammaticaux pour contraindre l'accès au lexique chez des enfants de 18 mois

Cette étude porte sur la capacité des enfants de dix-huit mois à exploiter le contexte syntaxique pour reconnaître les noms et les verbes. En utilisant la technique de conditionnement de l'orientation de la tête, nous avons testé si les jeunes enfants s'attendaient à ce que les déterminants précèdent les noms et que les pronoms précèdent verbes. La moitié des enfants a été entraînée à tourner la tête pour le nom *balle* et l'autre moitié pour le verbe *mange*. Ils ont été testés en réponse à trois types de phrases différents :

- Contexte correct où le mot cible est précédé du mot de fonction juste :
J'adore les balles en mousse ou *Pierre, il mange du chocolat* ;
- Contexte incorrect où le mot cible est précédé d'un mot de fonction incongruent :
**Pierre, il balle du chocolat* ou **J'adore les manges en mousse* ;
- Phrase distractrice qui ne contient pas le mot cible :
J'adore les fraises au sucre ou *Demain tu chantes chez Paul*.

Étant donné les résultats de la littérature, nous nous attendions à ce que les enfants du groupe nom tournent la tête plus souvent pour les contextes corrects par rapport aux contextes incorrects et donc qu'ils exploitent les déterminants pour reconnaître le nom cible. Avec cette nouvelle technique, les résultats répliqueraient alors ceux de (Shi and Melancon, 2010; Höhle et al., 2004; Kedar et al., 2006; Zangl and Fernald, 2007), nous permettant de vérifier l'efficacité du paradigme. En effet, comme cette méthode utilise une détection de mot purement auditive, les enfants pourraient interpréter la tâche comme une détection de forme phonologique de la cible plutôt qu'une détection de mot à proprement parler. Dans ce cas, aucune différence entre les contextes corrects et incorrects ne sera observée puisque la forme phonologique est présente dans les deux cas. Au contraire, si les enfants sont capables d'exploiter le contexte dans lequel apparaît le mot cible pour le reconnaître plus rapidement, alors les enfants reconnaîtront

spécifiquement le mot *balle* et pas seulement sa forme phonologique.

De plus, si le groupe nom montre un avantage significatif pour le contexte correct par rapport à l'incorrect, le groupe verbe nous permettra de tester si les enfants de dix-huit mois utilisent les pronoms pour faciliter la reconnaissance des verbes ce qui n'a pas encore été montré à cet âge. Si le verbe cible est reconnu aussi souvent lorsqu'il est précédé d'un mot grammatical correct qu'incorrect, alors nous pourrions conclure que les pronoms ne sont pas encore utilisés pour la reconnaissance des verbes. En revanche, si les enfants reconnaissent plus souvent le verbe cible lorsqu'il est précédé d'un pronom par rapport à un déterminant, alors nous pourrions conclure que les enfants exploitent les pronoms pour la reconnaissance des verbes. De plus, la comparaison directe des performances des enfants du groupe nom et du groupe verbe permettra de tester si les noms bénéficient plus du contexte qui les précède que les verbes (comme le suggère la littérature).

Matériel et méthodes

Participants Trente-six enfants de 18 mois de langue maternelle française ont participé à cette étude (âgés de 17 mois 14 jours à 18 mois 25 jours, en moyenne 18 mois 2 jours). L'expérience consistait en 2 sessions utilisant la technique de conditionnement de l'orientation du regard (Conditioned-Head-Turning). La moitié des enfants (18) ont été entraînés avec une cible nom et l'autre moitié (18) ont été entraînés avec une cible verbe. Dix-huit autres enfants ont été testés mais exclus de l'analyse parce qu'ils n'avaient pas compris la tâche de détection (9 dans le groupe Nom et 9 dans le groupe Verbe) : ils n'ont pas tourné la tête plus souvent pour les phrases correctes (phrases dans lesquelles la cible apparaissait dans une position normale) que pour les phrases distractrices (dans lesquelles la cible était absente). Cent-huit enfants supplémentaires ont été testés mais n'ont pas terminé la première session : 60 n'ont pas réussi à passer au delà de l'entraînement sur les groupes de mots, et 48 sur les courtes phrases. Plusieurs raisons expliquent ces échecs : le manque d'intérêt, la peur, le manque de coopération (maladie ou fatigue)

ou parce que les enfants n'ont pas réussi à atteindre le critère de réussite au bout de 40 essais (voir la procédure). Parmi les enfants qui ont réussi à finir la première session, 2 ne sont pas revenus pour la deuxième session, 3 ont été exclus pour des raisons techniques et 22 autres n'ont pas fini la seconde session parce qu'ils n'étaient plus motivés par le renforcement ou parce qu'ils étaient trop agités pour terminer la session. Plusieurs facteurs expliquent ce taux d'exclusion important : l'expérience avait lieu sur deux sessions séparées, chacune de ces sessions était très longue (entre 10 et 20 minutes pour la première session et environ 20 minutes pour la deuxième) ; il est difficile pour des enfants de 18 mois de rester assis calmement sur les genoux pendant toute l'expérience. De plus, les enfants ne pouvaient pas jouer avec les jouets eux-mêmes car il fallait qu'ils soient capables de diviser leur attention entre ce qu'ils voyaient et ce qu'ils entendaient. En jouant eux-mêmes, l'intérêt pour le jeu étant accru, l'attention portée aux stimuli audio aurait été réduite de beaucoup. Enfin, la tâche en elle-même était difficile, les enfants devaient partager leur attention entre ce que l'expérimentateur faisait en face d'eux, les stimuli sonores et le renforcement sur leur gauche. En comparant les taux d'exclusion dans des expériences utilisant un paradigme similaire (détection de mot en utilisant la technique de conditionnement de l'orientation du regard en 2 sessions) (Gout et al., 2004; Millotte et al., 2010), on s'aperçoit que plus l'enfant grandit, plus le taux d'exclusion croît. Cette observation s'explique par le fait que plus l'enfant est âgé, plus il a envie d'interagir et moins il aime rester sur les genoux de ses parents à regarder quelqu'un jouer en face de lui, sans même pouvoir toucher les jouets.

Stimuli Les enfants ont entendu 3 types de phrases différentes (exemple dans la table 1.1 page suivante) :

- douze phrases contenaient le mot cible dans une position correcte c'est-à-dire toujours précédé du mot grammatical correspondant à sa catégorie syntaxique (contexte correct) ;
- douze phrases contenaient le mot cible dans une position incorrecte c'est-à-dire qu'il était précédé d'un mot de fonction impossible avec sa catégorie syntaxique

(contexte incorrect) ;

- douze phrases ne contenaient pas le mot cible du tout (phrases distractrices).

	Groupe Nom Cible : Balle	Groupe Verbe Cible : Mange
Contexte Correct	J'adore les balles en mousse.	Demain, tu manges chez moi.
Contexte Incorrect	Demain, tu balles chez moi	J'adore les manges en mousse
Phrases distractrices	J'adore les fraises au sucre.	Demain, tu chantes chez Paul

TABLE 1.1 – Exemple de stimuli pour les 3 contextes (correct, incorrect et distracteur) pour les 2 groupes d'enfants selon qu'ils avaient été entraînés à reconnaître une cible nom ou verbe.

Les phrases incorrectes ont été construites en échangeant les cibles noms et verbes dans les phrases correctes de sorte que les phrases correctes du groupe nom soient identiques aux phrases incorrectes du groupe verbe à l'exception du mot cible. De la même façon, les phrases correctes du groupe verbe servaient de phrases incorrectes pour le groupe nom. Tous les stimuli ont été enregistrés en utilisant de la parole dirigée aux enfants par un locuteur de langue maternelle française. Toutes les phrases ont été enregistrées par paires (par exemple une phrase correcte du groupe nom avec la phrase incorrecte correspondante du groupe verbe) en utilisant la même intonation et le même rythme. Comme tous les stimuli ont été produits naturellement plutôt que par cross-splicing, nous avons vérifié que les différents contextes ne différaient pas par leur durée, ni par leur fréquence fondamentale ou encore par leur intensité et ce pour le mot cible, le mot de fonction le précédant et le mot qui suivait. Aucune différence n'a été détectée entre les contextes correct et incorrect. La table 1.2 page suivante montre les valeurs moyennes et les erreurs standards pour chaque variable acoustique.

La liste de tous les stimuli utilisés se trouve en annexe A page 221.

Création des groupes et contrebalancement Les enfants ont été assignés soit au groupe Nom, soit au groupe Verbe. Chaque groupe a été divisé en deux moitiés dont la différence portait sur les mots de fonction utilisés dans la session d'entraînement. La moitié des enfants du groupe Nom a été entraînée avec « la balle » et « des balles » et l'autre moitié avec « une balle » et « les balles ». De la même façon, la moitié des enfants

1.4. Expérience 1 : exploitation en temps réel des mots grammaticaux pour contraindre l'accès au lexique chez des enfants de 18 mois 27

		Correct		Incorrect		Différence
		Moyenne	(Err. Std)	Moyenne	(Err. Std)	Correct-Incorr.
Durée (ms)	Mot de fonction	203	(12.1)	213	(12.2)	-10
	Mot Cible	436	(14.9)	425	(23.3)	11
	Mot Suivant	256	(51.6)	282	(57.6)	-26
$F0_{max}$ (Hz)	Mot de fonction	276	(9.3)	274	(7.8)	2
	Mot Cible	346	(9.5)	341	(8.3)	6
	Mot Suivant	300	(11.9)	296	(11.0)	4
RMS (Pa)	Mot de fonction	0.091	(0.008)	0.089	(0.008)	0.002
	Mot Cible	0.100	(0.008)	0.093	(0.004)	0.007
	Mot Suivant	0.075	(0.006)	0.066	(0.006)	0.009

TABLE 1.2 – Analyse acoustique de la durée, de la fréquence fondamentale maximale (F0), de l'intensité sonore (Root mean square, RMS ou racine carré des moyennes des carrés de l'intensité) pour le mot de fonction précédant la cible, la cible et le mot la suivant, en comparant les contextes correct et incorrect. Le valeur du test statistique de student (t-value) pour les différences entre les contextes est toujours inférieure à 1.

du groupe Verbe a entendu pendant l'entraînement « je mange » et « il mange », tandis que l'autre moitié a entendu « tu manges » et « on mange ». Tous les enfants d'un même groupe ont entendu les mêmes phrases de test qui contenaient les 4 déterminants et les 4 pronoms. La moitié des phrases de test contenait donc des mots de fonction entendus pendant la session d'entraînement (familiers), tandis que l'autre moitié des phrases contenait des mots de fonction qui n'avait pas été utilisés dans l'entraînement (non utilisés ou nouveaux). Cette différence de familiarité des mots de fonction nous permettra de tester si les enfants repèrent seulement les bigrammes *mot de fonction* + *mot cible* par exemple « la balle » ou s'ils répondent spécifiquement pour les mots cibles et qu'ils sont donc capables de généraliser leur réponse à d'autres mots de fonction appropriés (contextes nouveaux).

Procédure Le paradigme expérimental contient une variante de la technique du conditionnement de l'orientation du regard à deux sessions comme dans (Gout et al., 2004; Millotte et al., 2010). Chaque session était divisée en deux parties. Durant la première session (entraînement), les enfants étaient entraînés sur des mots isolés puis sur de courtes phrases. Si l'enfant réussissait à atteindre le critère de réussite prédéfini, alors il revenait pour une deuxième session (8 jours d'intervalle entre les deux sessions en

moyenne). Cette seconde session débutait avec une phase de revue pour rappeler à l'enfant la tâche qu'il avait apprise, puis la phase de test proprement dit commençait ensuite (schéma 1.6).

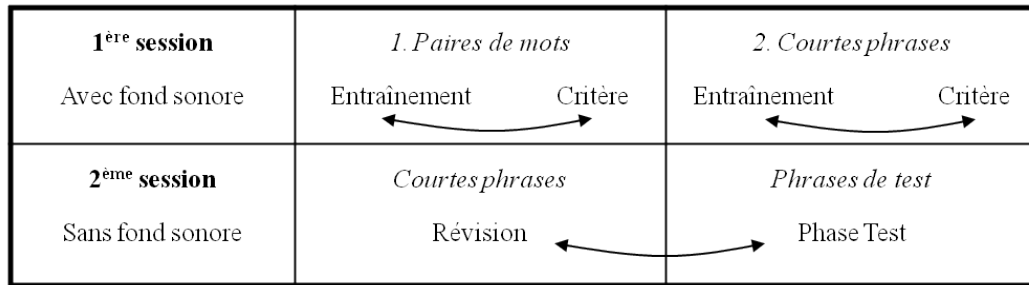


FIGURE 1.6 – Schéma de l'organisation de l'expérience.

Pendant toute l'expérience, l'enfant était assis sur les genoux de ces parents en face d'une petite table. Une assistante était assise en face de l'enfant et était chargée de maintenir l'attention de l'enfant en manipulant silencieusement un assortiment de jouets sur la table. Les haut-parleurs et les nounours animés mécaniquement qui servaient de renforcement étaient situés à 90° à gauche de l'enfant à environ 1.5m de distance. Une caméra vidéo était posée au dessus des haut-parleurs. Dans la salle de contrôle, l'expérimentateur observait l'enfant sur un téléviseur et déterminait si l'enfant regardait en direction de la caméra (et donc des nounours). Pendant toute l'expérience, le parent et l'assistante portaient des écouteurs dans lesquels ils entendaient un masquage (un brouhaha de personnes parlant en même temps dans différentes langues) par un casque antibruit. L'expérimentatrice pouvait à n'importe quel instant demander à l'assistante d'adapter son comportement à celui de l'enfant, grâce à un microphone connecté au casque de l'assistante. Lorsque l'enfant était focalisé sur ce qui se passait sur la table en face de lui, l'expérimentatrice lançait l'essai. Si l'enfant tournait la tête vers les haut-parleurs, l'expérimentatrice appuyait sur un bouton pour signaler l'orientation du regard. Le renforcement était activé par l'ordinateur seulement si l'enfant tournait la tête de façon appropriée c'est-à-dire lorsqu'il avait entendu le mot cible.

Première session : Entraînement

1.4. Expérience 1 : exploitation en temps réel des mots grammaticaux pour contraindre l'accès au lexique chez des enfants de 18 mois **29**

Pendant la première phrase d'entraînement, les enfants entendaient en boucle une série de paire de mots comprenant un mot de fonction et un mot de contenu, par exemple « une balle ... les balles » ou « je mange ... il mange ». Les paires de mots étaient séparées par 1000ms et présentées à un volume sonore confortable (68dB SPL). Pour le groupe Nom, des groupes verbaux étaient présentés comme fond sonore et remplacés par des groupes nominaux pour les essais cibles. A l'inverse pour le groupe Verbe, les groupes nominaux servaient de fond sonore présentés en boucle. Lorsque l'attention de l'enfant était portée sur les jouets en face de lui, l'expérimentatrice lançait un essai cible. Un triplet de groupes cibles (par exemple « je mange ... je mange ... il mange » pour le groupe Verbe) était présenté 12dB plus fort. Lorsque l'intensité sonore augmentait pour les cibles, l'enfant tournait le plus souvent spontanément la tête vers la source sonore, les haut-parleurs. L'expérimentateur appuyait sur la bouton droit de la souris pour signaler que l'enfant se tournait. Si l'enfant tournait la tête pendant qu'il entendait le triplet de cibles (en moyenne 6,13s), alors un nounours apparaissait pendant 3500ms, fenêtre de renforcement, et le volume sonore des cibles diminuait de 3dB. Si l'enfant échouait trois fois de suite, le volume sonore remontait d'un cran de 3dB. Lorsque le même volume sonore était atteint pour les cibles et le fond sonore et que l'enfant réussissait deux fois de suite à répondre pour les cibles, alors la phase de critère débutait. Cette étape permettait de s'assurer que l'enfant répondait effectivement aux mots cible et non à la différence de volume sonore. Dans cette phase, lorsque l'expérimentatrice lançait un essai, l'ordinateur jouait aléatoirement des cibles ou du fond sonore. L'expérimentatrice était donc aveugle au type de stimuli qu'entendait l'enfant et rapportait juste si l'enfant s'était tourné vers les nounours ou non. La réponse de l'enfant était codée comme correcte s'il se tournait pour les cibles et s'il ne se tournait pas pour le fond sonore ; et incorrecte s'il ne se tournait pas pour les cibles ou s'il répondait au fond sonore. Si l'enfant échouait successivement trois fois, il revenait en phase d'entraînement. Si l'enfant répondait correctement sept fois sur huit essais consécutifs, ou pour 80% de 12 essais consécutifs, l'enfant poursuivait dans la seconde partie de l'entraînement : reconnaître le

mot dans de courtes phrases. Si l'enfant ne parvenait pas à atteindre l'un de ces critères sur 40 essais, l'expérience était terminée pour lui. La procédure de la seconde partie de la session d'entraînement était semblable à la première sauf que les stimuli (fond sonore et cibles) étaient remplacés par de courtes phrases (par exemple : « la balle est jolie » ou « je mange une pomme »). Si l'enfant réussissait à atteindre le critère de cette seconde phase à savoir répondre correctement à 7 de 8 essais consécutifs ou à 80 des essais, alors on demandait aux parents de revenir pour une seconde session quelques jours plus tard. Les enfants qui n'avaient pas atteint un de ces critères en 40 essais étaient exclus de l'étude.

Seconde session : Test

Cette session débutait par une phase de rappel utilisant les mêmes courtes phrases que celles de la deuxième partie de la première session, cette fois-ci sans fond sonore. L'expérimentatrice lançait un essai, une phrase était choisie aléatoirement (50% contenant la cible et 50% sans) par l'ordinateur et était présentée à l'enfant. L'expérimentatrice avait connaissance du type de phrase jouée. La première phrase était présentée 12dB plus fort que les phrases de test. Si le comportement de l'enfant était correct, c'est-à-dire s'il tournait la tête si la phrase contenait la cible (et qu'il était récompensé par un nounours qui apparaissait durant 2500 ms) ou s'il l'ignorait quand elle n'était pas présente, la phrase suivante était jouée 4 dB moins fort. Si l'enfant faisait 2 erreurs consécutives, le volume sonore augmentait à nouveau de 4dB. Lorsque l'enfant réussissait 2 fois de suite avec une phrase présentée au même niveau sonore que les phrases de test, alors le programme passait automatiquement en phase Test. L'intervalle entre deux phrases était au minimum de 1000 ms. Dans la phase de test, l'expérimentatrice lançait un essai et l'ordinateur choisissait aléatoirement une phrase parmi les trois types possibles (contexte correct, incorrect ou phrase distractive). L'expérimentatrice était à ce moment aveugle aux conditions présentées à l'enfant. La récompense n'était donnée à l'enfant que s'il tournait la tête pour une phrase contenant le mot cible. L'enfant avait 2500 ms pour tourner la tête à partir du début du mot cible. La récompense apparaissait alors pen-

dant 2500 ms. Si l'expérimentatrice considérait que l'enfant était moins attentif, elle pouvait choisir de repasser en phase de rappel pour récupérer son attention. Pour finir la session, l'enfant devait avoir entendu chacune des trente-six phrases. Le schéma [1.6 page 28](#) résume l'organisation de l'expérience.

Analyse En plus du codage en temps réel, les résultats des enfants qui avaient terminé la deuxième session ont été recodés. Le recodage a été effectué pour éviter les cas où l'enfant s'était tourné juste avant le début du mot cible. Concrètement, les essais pour lesquels le temps de réponse enregistré à partir du début du mot cible était inférieur à 250 ms (ou même négatif), ont été considérés comme des essais pour lesquels l'enfant était déjà tourné. Ce temps a été ajouté pour tenir compte du fait que si l'enfant était tourné à ce moment là, la décision de tourner la tête (temps de saccade) avait été prise avant l'écoute du mot cible. De plus, ce recodage s'assurait que toutes les réponses enregistrées en temps réel correspondaient effectivement à des moments où l'enfant s'était tourné vers les haut-parleurs et non à des moments où il ne se serait pas réellement tourné (fausse alerte, l'enfant semble démarrer une réponse mais ne la termine pas et revient très vite vers le centre) . Environ 15% des phrases ont changé de statut après le recodage manuel.

Résultats

La figure [1.7 page suivante](#) montre le pourcentage moyen d'orientation de la tête en fonction du type de phrases entendues et de la catégorie de la cible. Deux analyses de variance ont été réalisées sur les pourcentages moyens d'orientation de la tête vers les nounours, une par sujets et une par items. L'anova par sujets comportait deux facteurs inter-sujets : Catégorie de la cible (enfants entraînés à détecter une cible Verbe ou une cible Nom), et le facteur de contrebalancement (groupe 1 ou 2 entraîné avec des mots de fonction différents) et deux facteurs intrasujets : Condition (contexte Correct, Incorrect, ou phrase Distractrice), et Familiarité (selon que le mot de fonction a été utilisé pendant l'entraînement, Familier, ou non, Non-Familier). L'anova par items comportait trois facteurs intra-items : Catégorie de la cible, Condition et Familiarité.

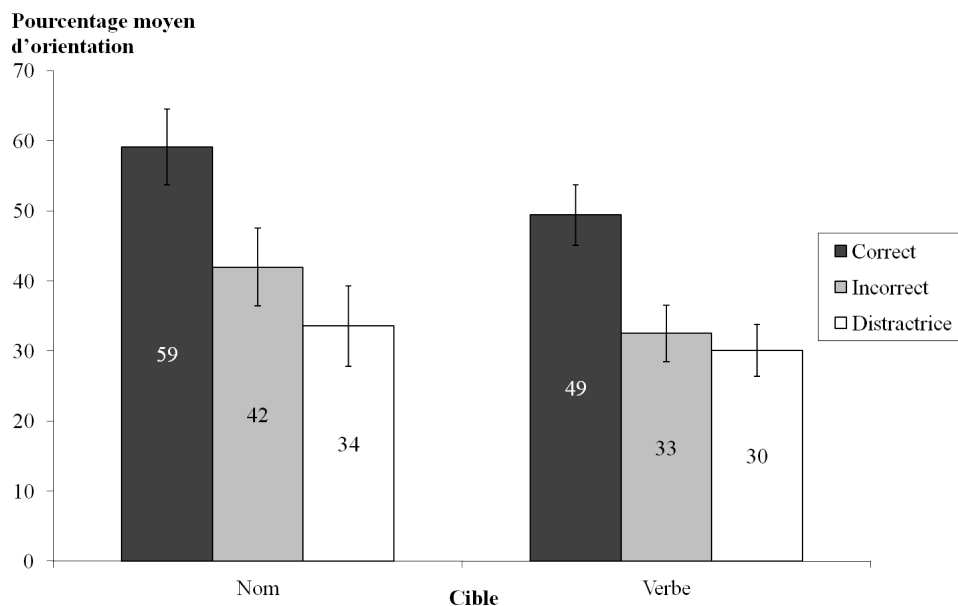


FIGURE 1.7 – Pourcentage du nombre de fois où les enfants s’orientent vers les nounours en fonction de la Catégorie de la Cible (Nom ou Verbe) et de la Condition (Contexte Correct, Incorrect ou Phrase Distractrice).

Les analyses de variance ont révélé un effet principal de Condition ($F_1(2, 64) = 33, p < 10^{-9}$, $F_2(2, 22) = 15, p < 10^{-4}$) avec plus de réponses aux phrases correctes (54%) que pour les phrases incorrectes (37%) et distractrices (32%). La Catégorie de la Cible présente un effet significatif seulement dans l’analyse par item ($F_1(1, 32) = 2.3$, $F_2(1, 11) = 8, p < 0.01$) ce qui reflète le fait que les enfants du groupe Nom ont tendance à répondre en moyenne plus que ceux du groupe Verbe (l’effet du groupe de sujet est particulièrement probable dans une analyse par item). De façon cruciale, le facteur Catégorie de la Cible n’interagit pas avec le facteur Condition ($F_1(2, 64) = 1.1$, $F_2(2, 22) < 1$), ni avec la Familiarité ($F_1(1, 32) = 3, p < 0.09$, $F_2(1, 11) < 1$) ce qui montre que le groupe d’enfant habitué avec des noms et celui avec des verbes se comportent de la même façon. L’effet de Familiarité est seulement marginal dans l’analyse par sujet ($F_1(1, 32) = 3.14, p = 0,09$, $F_2(1, 11) = 1.1$), ce qui reflète une petite tendance à répondre plus pour les mots de fonction utilisés pendant l’entraînement. Aucune interaction entre les facteurs Condition et Familiarité n’a été mise en évidence

($F_1(2, 64) < 1, F_2(2, 22) < 1$). L'effet de Condition est semblable pour les deux types de mots de fonction (voir la figure 1.8 qui dissocie les réponses en fonction de la familiarité des mots de fonction). Bien que l'interaction entre les facteurs Condition et Familiarité ne soit pas significative, nous avons voulu restreindre l'analyse aux mots de fonction qui n'avaient pas été employés dans la phase d'entraînement afin d'étudier uniquement la généralisation. Les mêmes anovas, restreintes, ont révélé que le facteur Condition était toujours significatif ($F_1(2, 64) = 13.9, p < 10^{-5}, F_2(2, 22) = 5.7, p < 0.01$). Donc, les enfants ne reconnaissent pas seulement les bigrammes *mot de fonction/mot cible* sur lesquels ils ont été entraînés mais sont bien capables d'identifier un mot cible lorsque celui-ci est précédé d'un mot de fonction avec lequel ils n'ont pas été entraînés.

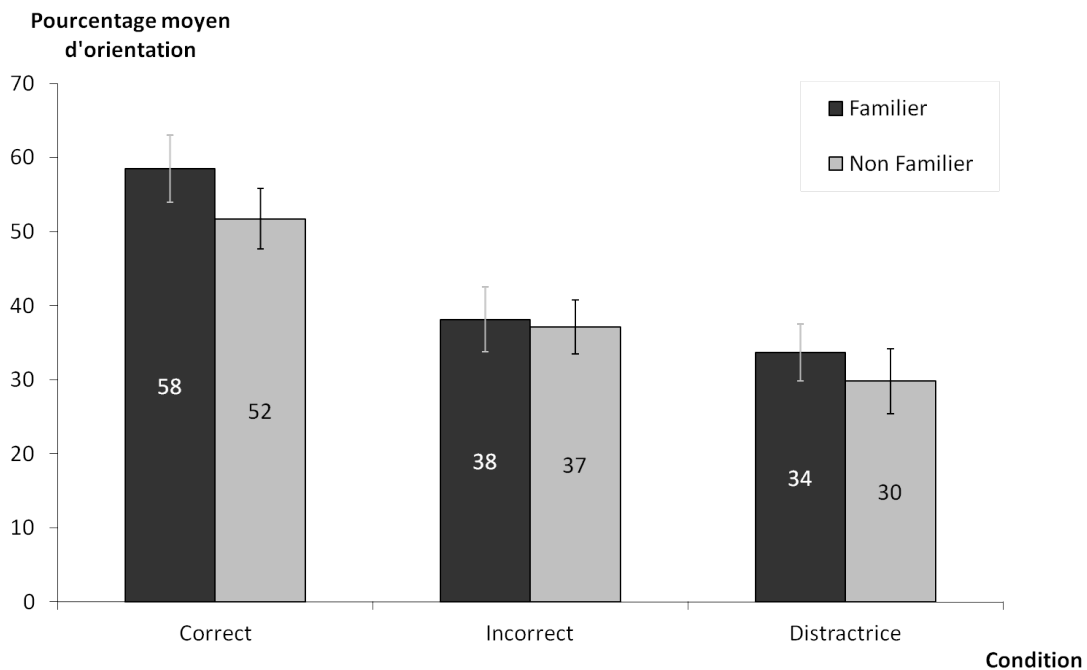


FIGURE 1.8 – Pourcentage d'orientation des enfants en fonction de la Condition (Contexte Correct, Incorrect, Phrase Distractrice) et des mots de fonctions utilisés pendant l'entraînement (Familiier ou Non Familier).

Puisque les deux groupes d'enfant se comportent de la même façon, nous nous sommes ensuite intéressés aux différences de comportement entre les différentes conditions. Nous avons réalisé trois comparaisons post-hoc, avec une correction de Bonferroni :

la différence entre les contextes Corrects et les phrases distractrices est particulièrement significative (taille d'effet 22.5%, $F_1(1, 32) = 59$, $p < 10^{-7}$, $F_2(1, 11) = 27$, $p < 0.001$) tout comme la différence de comportement entre les contextes Corrects et Incorrect (taille d'effet de 17,0%, $F_1(1, 32) = 40$, $p < 10^{-5}$, $F_2(1, 11) = 15$, $p < 0.008$). Le dernier résultat est particulièrement frappant puisque les deux types de phrases, bien qu'elles contiennent la même forme phonologique du mot cible provoquent des comportements très différents. Le contexte syntaxique dans lequel le mot cible apparaissait a influencé fortement le comportement des enfants. La dernière comparaison entre les contextes Incorrects et les phrases Distractrices n'était pas significative (taille d'effet 5.4%, $F_1(1, 32) = 3.4$, $p = 0.21$, $F_2(1, 11) = 1.3$, $p > 0.8$) : les enfants n'ont pas tourné leur tête significativement plus pour les phrases incorrectes qui contenait pourtant la cible que pour les phrases distractrices qui ne contenaient pas la cible.

Comme les enfants disposaient d'une fenêtre de 2.5 s pour donner leur réponse, il est possible qu'ils aient exploité le contexte suivant le mot cible pour répondre. Si c'était le cas, les enfants pourraient ne pas reconnaître le mot cible parce que celui-ci ne fait pas sens avec les mots qui le suivent. Par exemple, dans la phrase « Je balle une petite pomme », les enfants pourraient avoir du mal à accéder au nom « balle » non seulement parce qu'il est précédé d'un pronom mais aussi parce que le sens du nom ne va pas avec le reste de la phrase. Dans le but d'évaluer les réponses des enfants avant qu'ils n'accèdent aux mots suivant la cible, nous avons réalisé une analyse post-hoc supplémentaire. Nous avons restreint l'analyse aux réponses données avant la fin du mot cible plus un temps de décision (300 ms), et un temps pour la préparation motrice (300 ms)⁴. Ces réponses rapides représentaient 47% des données. La figure 1.9 page suivante montre les résultats par condition pour toutes les réponses (avec les mêmes données que dans les deux figures précédentes) et pour les réponses données en temps réel. Comme dans l'analyse principale, nous avons effectué une analyse de variance par sujet avec le facteur intra-sujet Condition, le facteur de Contrebalancement inter-sujet, et la

4. La fenêtre a été calculée séparément pour chaque item car la durée des mots cibles varie. La durée moyenne des mots était de 430.5 ms, la fenêtre des réponses rapides durait en moyenne 1030.5 ms.

Catégorie de la Cible, facteur inter-sujet (groupe 1 ou 2) également. Un effet principal de Condition $F_1(2, 64) = 6.6, p < 0.002$ montre que les enfants répondent significativement plus dans les contextes Corrects par rapport aux contextes Incorrects et aux phrases distractrices. Aucun autre facteur ni interaction ne se sont révélés significatifs. Des t-test post-hoc avec correction de Bonferroni comparant les contextes Corrects et Incorrects et aussi Corrects et phrases Distractrices montrent des différences significatives $t(35) = -2.6, p < 0.01$ et $t(35) = 2.96, p < 0.05$ respectivement. La différence entre les contextes Incorrects et les phrases Distractrices n'est pas statistiquement significative $t(35) < 1$.

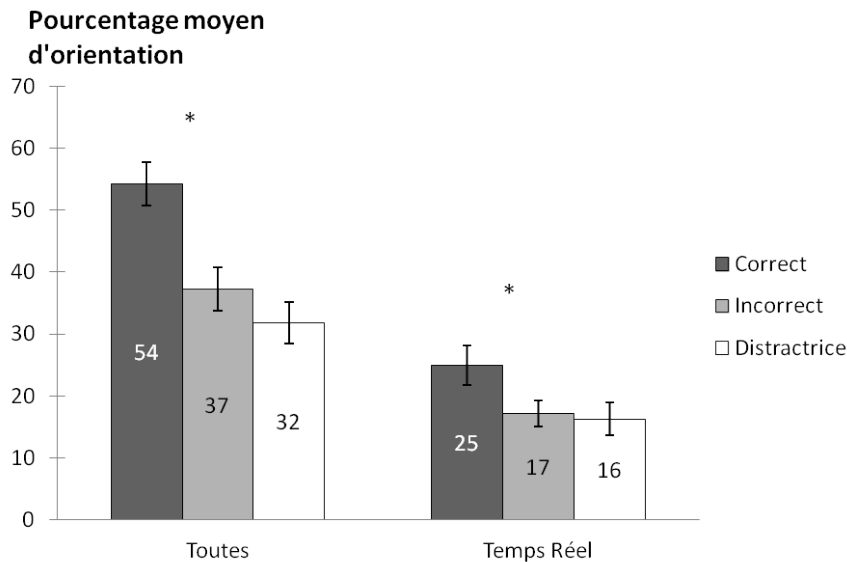


FIGURE 1.9 – Pourcentage moyen de réponses données pendant toute la fenêtre de réponse (Toutes) ou seulement les réponses données en temps réel.

Discussion

Nous avons utilisé un paradigme de conditionnement de l'orientation du regard pour déterminer si les enfants de 18 mois étaient capables d'utiliser les mots de fonction précédant un mot cible pour construire des attentes quant à la catégorie syntaxique d'un mot connu (Nom ou Verbe) et ainsi d'améliorer sa détection. Plus précisément, nous avons testé si les déterminants facilitaient la détection d'une cible Nom et si les

pronoms personnels facilitaient celle d'une cible Verbe.

La cible Nom a été identifiée plus souvent lorsqu'elle était utilisée dans un contexte syntaxique correct en comparaison d'un contexte syntaxiquement erroné dans lequel la forme phonologique de la cible apparaissait après un mot de fonction inapproprié. Donc les cibles Nom étaient mieux reconnues lorsqu'elles étaient précédées d'un déterminant par rapport à un pronom personnel, tandis que les cibles Verbe à l'inverse étaient reconnues plus souvent lorsqu'elles étaient précédées d'un pronom personnel par rapport à un déterminant. Nos résultats montrent que dans ce paradigme les enfants n'ont pas effectué une tâche de reconnaissance purement phonologique. Si c'était le cas, ils auraient du reconnaître la cible indépendamment du contexte dans lequel elle était présentée. Nous testons donc la détection d'un mot en particulier et non la reconnaissance de sa forme phonologique. Nos résultats sont en accord avec ceux obtenus par des paradigmes d'association d'image et de mot pour les noms ((Kedar et al., 2006; Zangl and Fernald, 2007)) qui montrent que des enfants anglais de 18 mois s'attendent à trouver un nom après avoir entendu un déterminant. L'avantage de notre paradigme est qu'il nous permet de tester de façon symétrique des noms et des verbes. Pour les verbes, comme pour les noms, nous observons que les enfants répondent significativement plus souvent pour la cible lorsqu'elle est placée dans une position syntaxiquement correcte par rapport à une position incorrecte. Les enfants de 18 mois de langue maternelle française s'attendent à trouver un verbe après avoir entendu un pronom personnel mais pas après un déterminant.

Ce résultat est nouveau étant donné que, de toutes les études qui ont essayé de montrer que les enfants étaient sensibles au contexte syntaxique dans lequel apparaissaient les verbes, seules celles réalisées avec des enfants d'au moins 20 mois ont réussi (Bernal et al., 2007) en français, (Oshima-Takane et al., 2010) en japonais et (Waxman et al., 2009) en anglais. Deux études portant sur des enfants plus jeunes ont échoué : (Shi and Melancon, 2010) montrent que des enfants de langue maternelle française de 14 mois ne reconnaissent que la classe de déterminant mais pas celle des pronoms. De façon simi-

laire, Höhle et ses collègues (Höhle et al., 2004) montrent que des enfants allemands de 16 mois qui entendent un nouveau mot précédé d'un déterminant le reconnaissent lorsqu'il apparaît dans une position de nom dans une phrase, mais ils ne sont pas capables de reconnaître un verbe (initialement précédé d'un pronom) lorsque celui-ci occupe une position de verbe dans une phrase. Au contraire, nos résultats ne montrent pas d'asymétrie entre les noms et les verbes chez des enfants français de 18 mois. Donc, dès 18 mois, les enfants en train d'apprendre leur langue maternelle sont capables d'exploiter les pronoms pour améliorer l'identification d'une cible Verbe.

Les mots de fonction utilisés pendant l'entraînement (Familiers) n'ont pas suscité plus de réponses que ceux qui n'avaient pas été présentés pendant l'entraînement (Non Familiers). Donc la tâche a été effectivement comprise comme une tâche de détection d'item lexical « balle » ou « mange » et non comme la détection des bigrammes *mot de fonction + mot cible* entendus pendant la session d'entraînement. Les analyses de variance restreintes aux mots de fonction Non-Familiers montrent le même effet du facteur Condition que les anovas globales : la cible est plus souvent reconnue lorsqu'elle est précédée du mot de fonction correct plutôt que d'un mot de fonction incorrect, même si ces mots de fonction n'ont pas été présentés pendant l'entraînement.

De plus, une analyse restreinte aux réponses en temps réel, réponses initiées avant la fin du mot cible, montre le même effet que les résultats globaux (figure 1.9 page 35). Même lorsque la décision est prise sur la base seule du mot cible et du contexte le précédant, les jeunes enfants traitent différemment les phrases correctes et incorrectes. Les enfants exploitent donc le contexte qui précède immédiatement un mot pour le reconnaître en temps réel.

Nous avons vérifié que la différence de reconnaissance entre les contextes corrects et incorrects ne pouvait pas s'expliquer par un effet d'apprentissage durant l'expérience. En effet, les enfants pourraient avoir appris à ne pas tourner la tête pour les contextes incorrects parce que ces phrases n'étaient pas renforcées pendant la phase de test. Si c'était le cas, les enfants tourneraient initialement la tête en réponse aux contextes

incorrects et remarqueraient en cours de session que ce type de phrase n'était pas renforcé et qu'il fallait donc arrêter de se tourner en les entendant. En d'autres mots, le taux de réponse devrait être plus important au début qu'à la fin pour les phrases incorrectes. Une analyse post-hoc montre que ce n'est pas le cas. Les enfants répondent de façon similaire pendant le premier et le dernier tiers de l'expérience. La différence de réponse entre les contextes corrects et incorrects ne peut pas être due à un effet d'apprentissage.

Une explication alternative à nos résultats pourrait être que les enfants reconnaîtraient des groupes de mots fréquents stockés en mémoire. Ils réagiraient donc plus au mot cible lorsque celui-ci appartient à une séquence de mot familière plutôt qu'à une suite non familière de mots comme dans les contextes incorrects dans lesquels le mot cible et le mot de fonction le précédant n'apparaissent jamais ensemble. Si cette hypothèse est juste, les enfants devraient mieux découvrir un mot familier lorsqu'il est inséré dans une suite fréquente de mots. Ceci implique de trouver une corrélation positive entre le pourcentage de l'orientation de la tête et la fréquence de la suite de mot « mot de fonction + mot cible » dans ce qu'entendent les enfants. Nous avons exploré cette hypothèse en calculant les fréquences d'occurrence de suites de mots à partir du corpus de type CHILDES en français (Demuth and Tremblay, 2008). Aucune corrélation n'a été trouvée entre la fréquence d'occurrence des suites de mots et le taux de bonne réponse (table 1.3 page suivante). Cette analyse renforce notre interprétation en termes de connaissance de catégories syntaxiques et affaiblit l'hypothèse d'un effet lié uniquement à des suites de mots stockés en mémoire.

Finalement, nos analyses tendent à montrer que les enfants de dix-huit mois reconnaîtraient les mots en temps réel en s'appuyant sur les mots qui les précèdent. De plus, une comparaison directe entre les contextes incorrects qui contiennent la forme sonore de la cible et les phrases distractrices qui en sont dépourvues révèle une différence non significative. Cette absence de différence est particulièrement frappante puisque la forme phonologique de la cible est présente dans les contextes incorrects et pas dans les phrases distractrices ; on aurait pu s'attendre à ce que les enfants soient tentés de répondre à la

Item	Fréquence	Pourcentage d'orientation global	Pourcentage d'orientation restreint aux mots grammaticaux
La balle	40	66	52
Des balles	1	46	40
Une balle	6	45	52
Les balles	0	71	73
Je mange	4	44	48
Tu manges	17	40	30
Il mange	14	46	52
On mange	17	42	40
R^2		0.04	0.04

TABLE 1.3 – Fréquence des suites de mots entendues par les enfants (corpus CHILDES (Demuth and Tremblay, 2008)) et leur corrélation avec le pourcentage d'orientation de la tête vers la cible. La corrélation a été calculée pour tous les essais mais aussi seulement pour les essais utilisant des mots de fonction non-utilisés dans la phase d'entraînement.

forme phonologique. Au contraire, ils se comportent comme si la cible dans un contexte incorrect (« je balle ») était différente de la cible réelle qu'ils recherchaient. Ce résultat suggère qu'ils ne reconnaissent pas la cible si elle apparaît dans un contexte incorrect ou que sa reconnaissance est particulièrement ralentie.

La reconnaissance et l'utilisation des mots de fonction pour construire des attentes quant à la catégorie syntaxique du mot qui suit pourraient contraindre l'accès au lexique en temps réel. Nous interprétons nos résultats comme soutenant un modèle d'accès au lexique contraint par la syntaxe. La figure 1.10 page suivante présente ce modèle en l'adaptant à ce cas de non concordance entre ce qui est prédit et ce qui est effectivement entendu. Ce modèle peut être appliqué à notre expérience chez les enfants : alors qu'ils entendent un contexte de nom, les enfants chercheraient dans leur lexique mental uniquement un nom commençant par les premiers phonèmes entendus, tandis que s'ils entendent un contexte de verbe, ils rechercheraient un verbe commençant par les mêmes phonèmes. Selon cette interprétation, en entendant des phrases du type « ... je balle ... » les enfants devraient commencer à traiter « je ba », s'attendre à trouver un verbe commençant par la syllabe /ba/ et essayer de trouver dans leur lexique le verbe *balle*

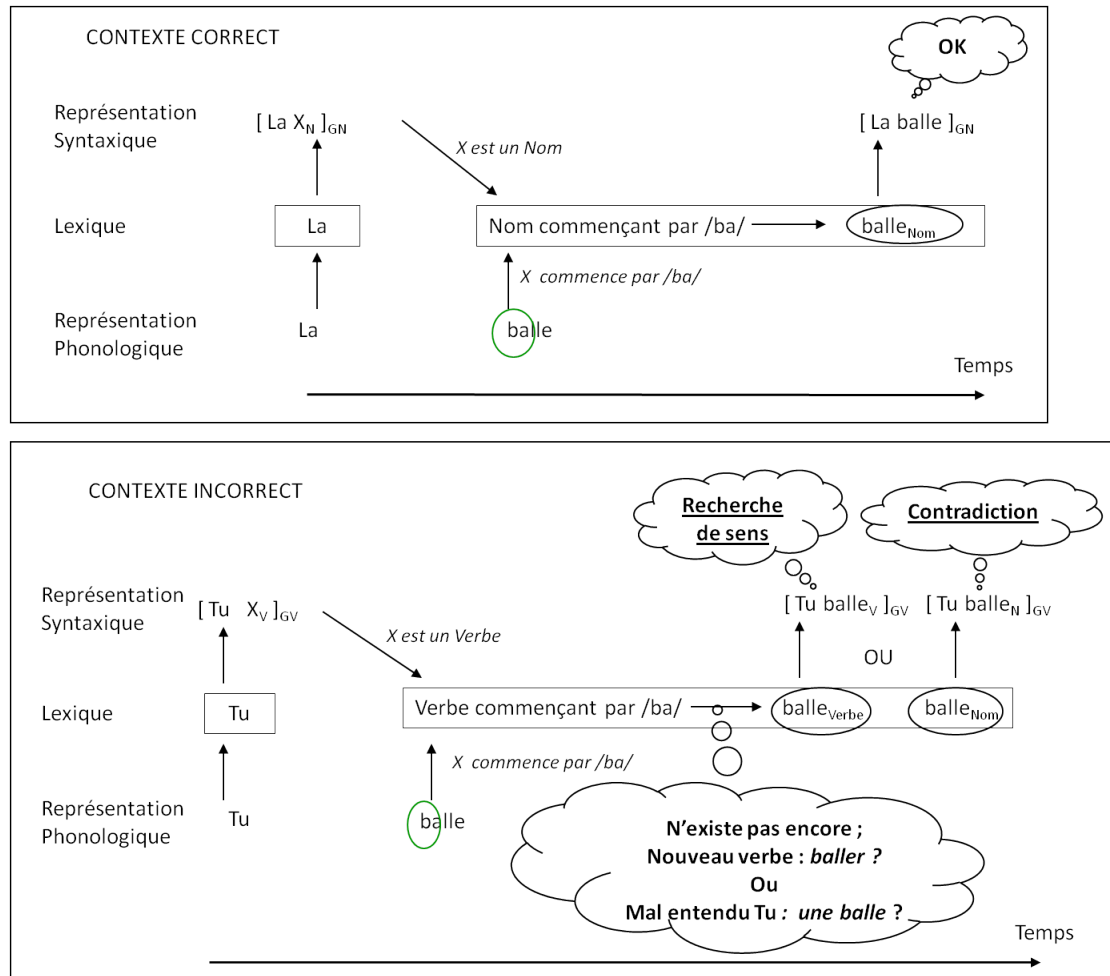


FIGURE 1.10 – Modèle où les prédictions syntaxiques contraignent l'accès lexical. En haut, le contexte syntaxique d'apparition du mot balle est correct. En bas, le contexte est incorrect.

parmi tous les verbes commençant par /ba/. Cette recherche devrait alors échouer. Les enfants pourraient alors effectuer 2 choses :

- soit rechercher un nouveau sens pour ce mot : « un nouveau verbe que je ne connais pas encore : baller ? Quel est son sens ? » et créer alors une nouvelle entrée dans leur lexique ;
- soit échouer à trouver un verbe cohérent et finalement accéder au nom *balle* tardivement (après la fenêtre de 2.5sec), et alors essayer de résoudre cette contradiction entre la catégorie syntaxique attendue et celle du mot récupéré dans leur lexique. Il est cependant possible de proposer un modèle dans lequel l'analyse syntaxique

a lieu après l'accès au lexique. La figure 1.11 page suivante présente un modèle alternatif d'interaction en temps réel entre les informations syntaxiques et lexicales. Dans ce modèle, il y a toujours une première tentative d'identification lexicale sur la base d'informations phonologiques. Les contraintes syntaxiques interviennent seulement plus tard dans le traitement, lorsque les mots identifiés sont mis ensemble pour former une structure syntaxique cohérente. Ce modèle permet d'émettre des prédictions pour notre expérience. Si les enfants entendent le mot cible *balle*, quel que soit le contexte, ils devraient essayer d'accéder à l'item lexical et donc on s'attendrait à ce qu'ils répondent au moins quelquefois à la cible dans des contextes incorrects (par rapport au taux basal de réponse mesuré pour les phrases distractrices). Dans une seconde étape, le mot reconnu serait inséré avec les autres mots précédemment identifiés, dans notre cas le mot de fonction précédant la cible. Si ce mot de fonction est congruent alors l'accès lexical se poursuit correctement avec le reste de la phrase. Lorsque ce mot de fonction est incongruent, alors une erreur se produit et le système doit revenir en arrière pour résoudre la contradiction. Ce processus ralentirait l'accès lexical du mot cible et de celui qui le suit. Ce modèle nous permet de prédire au moins quelques réponses pour la cible dans un contexte incorrect puisque la forme phonologique cible est présente, même si leur nombre pourrait être plus réduit par rapport aux contextes corrects puisque le processus d'intégration n'est pas interrompu dans ce cas. Ici, nous observons un taux de réponse très réduit pour les contextes incorrects par rapport aux contextes corrects, à tel point que les réponses pour les contextes incorrects ne sont pas significativement différentes des réponses données pour les phrases distractrices qui ne contiennent pas le mot cible.

Nos résultats sont totalement cohérents avec le premier modèle présenté dans lequel les informations syntaxiques contraignent l'accès lexical en temps réel. Ils sont en désaccord avec le deuxième.

Notre étude montre que les enfants francophones de 18 mois sont capables d'exploiter les pronoms pour améliorer le traitement des verbes. Elle montre des résultats similaires

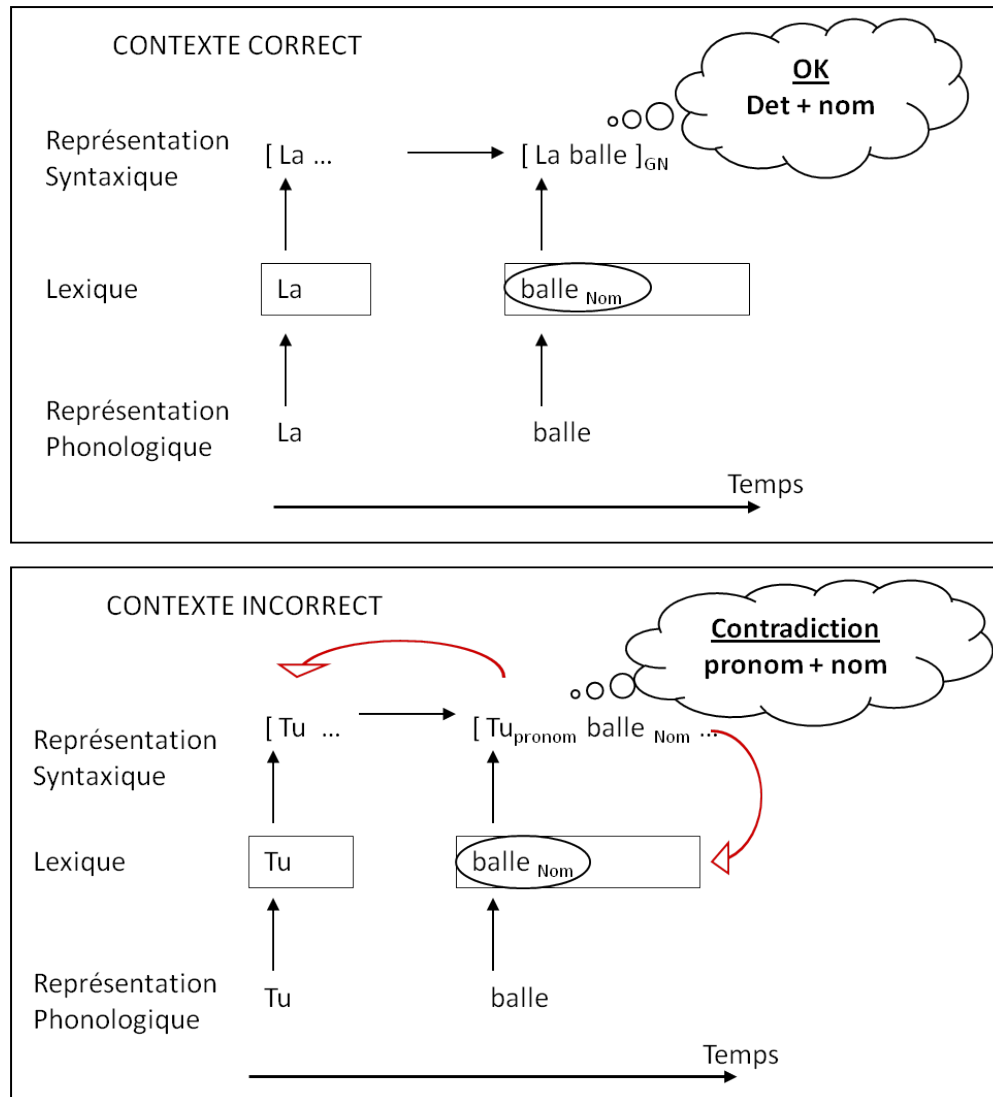


FIGURE 1.11 – Modèle de traitement de la parole en temps réel : L'accès au lexique est effectué à partir des informations phonologiques seulement. La syntaxe jouerait un rôle second. En haut, la prédiction de ce modèle dans le cas où le mot balle apparaît dans un contexte syntaxique correct ; en bas, dans un contexte incorrect.

à d'autres études dans lesquelles les enfants de 18 mois sont déjà capables d'exploiter des informations morphologiques spécifiques des verbes. Par exemple, dans l'étude de (Golinkoff et al., 2001), des enfants de 18 et 21 mois ont vu simultanément deux actions pendant qu'ils entendaient soit la forme grammaticale d'un verbe (par exemple dancing), soit une forme agrammaticale (dancely) ou encore une forme qui n'existe pas (dancelu). Ils regardaient significativement plus longtemps l'image représentant l'action pendant

qu'ils entendaient la forme grammaticale du verbe en comparaison des deux autres formes ; ce qui suggère que la morphologie verbale est utile pour repérer un verbe connu. De plus, les enfants de 18 mois sont sensibles à des dépendances à longue distance comme celle existant en anglais entre l'auxiliaire et la terminaison verbale -ing (Santelmann and Jusczyk, 1998). Ces auteurs ont utilisé un paradigme de préférence de l'orientation du regard avec des enfants de 15 et 18 mois. Ils montrent que seuls les enfants de 18 mois regardent plus longtemps pour les passages contenant la construction « is verbe-ing » en comparaison de passages contenant « can verbe-ing » (voir aussi (Höhle et al., 2006) pour des résultats similaires en allemand). De la même façon, (Van Heugten and Shi, 2010) montrent que des enfants francophones de 17 mois écoutent plus longtemps des phrases contenant des dépendances grammaticales comme « la coupile va conduire » que des phrases contenant des dépendances agrammaticales comme « les coupiles va conduire ». Comme les enfants sont sensibles à des dépendances discontinues de ce type, il est possible qu'ils les exploitent pour catégoriser des mots inconnus (voir (Mintz, 2003) pour l'hypothèse des cadres fréquents - frequent frames en anglais et (Chemla et al., 2009) pour le français). En prenant en compte ces résultats montrant une sensibilité à la morphologie verbale et notre démonstration que les pronoms sont exploités en temps réel pour le traitement des verbes, nous suggérons que les enfants de 18 mois sont capables d'utiliser le contexte morphosyntaxique dans lequel apparaissent les verbes pour améliorer leur traitement.

En résumé, nos résultats et ceux de la littérature dans différentes langues, suggèrent qu'à l'âge de 18 mois, les enfants ont déjà une connaissance précise des contextes morphosyntaxiques dans lesquels les noms et les verbes apparaissent. Il est donc possible qu'ils utilisent cette connaissance dans le but de catégoriser des mots inconnus et ensuite d'inférer si le sens est plus probablement une action si c'est un verbe ou un objet si c'est un nom.

Pour conclure, nous montrons que les enfants de 18 mois reconnaissent plus souvent un mot cible lorsque celui-ci est précédé d'un mot de fonction correct. Les enfants

tournent la tête vers les nounours plus souvent lorsque la cible verbe est précédée d'un pronom personnel que lorsque le verbe est précédé d'un déterminant ; et de façon similaire pour une cible nom précédée d'un déterminant par rapport à un pronom. Les enfants, alors qu'ils sont en train d'acquérir leur lexique, exploitent en temps réel le contexte syntaxique dans lequel un mot apparaît pour le reconnaître. Ces résultats sont compatibles avec un modèle de traitement du langage dans lequel le contexte syntaxique contraint l'accès au lexique en temps réel.

1.5 Accès au lexique

Les mots grammaticaux sont donc exploités en temps réel pour calculer la catégorie syntaxique de mots connus par les enfants. Par principe de continuité, ce mécanisme pourrait aussi être utilisés par les adultes pour contraindre l'accès au lexique. Cette étape de la reconnaissance d'un mot est la mise en correspondance entre la forme phonologique et la représentation de sa forme lexicale. Il s'agit d'associer un sens à la forme sonore entendue en recherchant dans son lexique interne. Dès les premiers phonèmes d'un mot, s'effectue une recherche de l'item le plus probable pour le prédire. Une cohorte de mots commençant par les mêmes phonèmes est pré-activée et à mesure que la somme d'informations augmente, le ou les items les plus probables sont sélectionnés, c'est le modèle de la cohorte (Marslen-Wilson, 1975). Les alternatives les plus activées sont celles qui ressemblent le plus à la forme sonore. Sur quels indices repose cette prédiction lexicale ? Qu'est ce qui permet de sélectionner les mots les plus probables dans le lexique ?

Il est évident que le contexte pragmatique dans lequel le mot apparaît va jouer un rôle dans l'accès lexical et l'isolation de l'item correct. Cependant, la syntaxe et particulièrement le calcul de la catégorie grammaticale et le genre des mots (masculin, féminin) pourraient fournir des indices pertinents quant à la sélection lexicale du sens juste. Plusieurs études montrent que les adultes (Dahan et al., 2000; Lew-Williams and Fernald, 2010) et les enfants (Van Heugten and Shi, 2009) utilisent en effet le genre des mots pour

contraindre leur accès lexical. Ainsi Dahan et collègues (Dahan et al., 2000) ont réalisé une expérience analysant la position du regard (eye tracking). Les sujets voyaient quatre objets sur un même écran comme sur la figure 1.12, puis une instruction leur demandait de cliquer avec la souris sur la cible. La direction du regard a été analysée durant la recherche de l'objet. En l'absence de marquage du genre, réalisé en utilisant le pluriel,

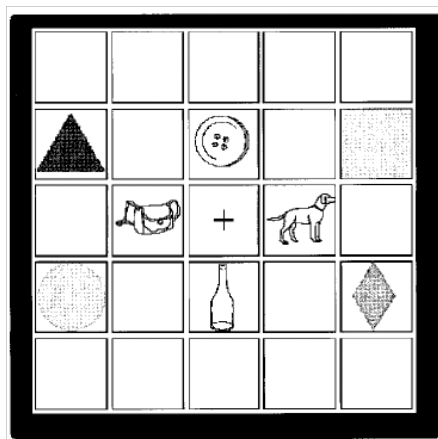


FIGURE 1.12 – Exemple d'un essai. Les couleurs ont été omises dans ce schéma. Figure issue de (Dahan et al., 2000). Parmi les quatre objets présentés, on trouve la cible, par exemple *bouton*, un compétiteur qui commence par la même syllabe *bouteille* et deux distracteurs.

les sujets ont regardé plus la cible (*les boutons*) et le compétiteur, (commençant par les mêmes phonèmes que la cible par exemple *les bouteilles*) par rapport aux distracteurs non reliés phonologiquement (*les chiens et les sacs*), et ce, à mesure que les premiers phonèmes du mot étaient prononcés. Dans une deuxième condition dans laquelle les mots étaient spécifiés pour leur genre, alors seule la cible (par exemple *le bouton*) était fixée plus souvent par rapport au compétiteur partageant les premiers phonèmes avec la cible mais qui était d'un genre différent (*la bouteille*) et aux distracteurs sans lien phonologique avec la cible. Le déterminant spécifiant le genre d'un mot contraint donc l'accès lexical puisque les noms de genre différent ne sont pas sélectionnés même s'ils commencent par les mêmes phonèmes.

La catégorie syntaxique des mots réduirait de façon importante la taille de l'ensemble des mots du lexique pouvant apparaître dans cette position. Le modèle de traitement

du langage dans lequel la syntaxe contraint l'accès au lexique postule donc qu'il serait possible dès les premiers phonèmes d'un mot de chercher ce mot uniquement parmi une partie du lexique spécifiée par sa catégorie; par exemple en entendant *la ba*, on chercherait un nom féminin commençant par la syllabe /ba/ au lieu de chercher parmi tous les mots commençant par cette même syllabe.

Un tel mécanisme est rappelé dans la figure 1.13. Dans ce modèle, dès que le mot de fonction est analysé et les premiers phonèmes du mot qui le suit sont entendus, des attentes syntaxiques sont créées et sont alors utilisés pour contraindre la recherche de l'item dans le lexique. Ce mécanisme pourrait jouer un rôle crucial dans des cas

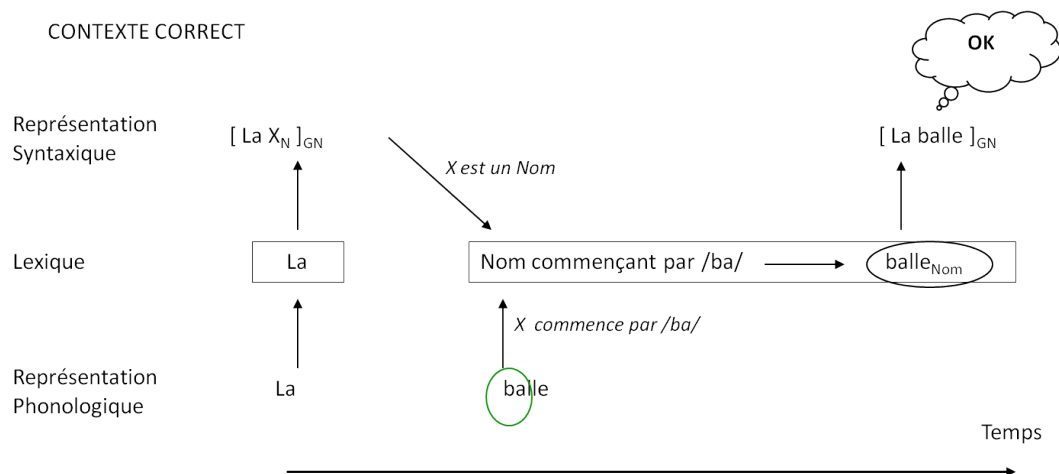


FIGURE 1.13 – Modèle représentant l'intégration des informations syntaxiques dans l'accès lexical.

d'ambiguïté lié à l'homophonie. Par exemple, le mot *porte* peut être un nom *la porte* mais aussi un verbe *je porte*. L'enchaînement des phonèmes ne suffit pas à isoler un sens pour le mot.

1.6 Expérience 2 : exploitation des indices syntaxiques pour contraindre l'accès au lexique en temps réel chez l'adulte

La prochaine expérience comportementale teste donc, si le contexte syntaxique précédant un mot et qui spécifie sa catégorie grammaticale, nom ou verbe, exerce également une contrainte sur l'accès lexical chez les adultes. Dans cette tâche de détection de mot cible dans une phrase auditive, les sujets sont poussés à répondre rapidement. Dans les phrases, se trouve soit la cible par exemple *chat*, soit un compétiteur de la même catégorie grammaticale *château*, soit un compétiteur de catégorie grammaticale différente *chatouille*. Si la catégorie grammaticale des mots, calculée en temps réel, est effectivement utilisée pour prédire le mot qui suit et restreindre le nombre de mots du lexique pré-activés, alors les sujets devraient effectuer plus de fausses alarmes sur les mots de la même catégorie syntaxique que la cible. Selon cette hypothèse, un mot d'une catégorie syntaxique différente, même s'il partage les mêmes premiers phonèmes, ne devrait pas pouvoir être activé.

Matériel et méthodes

Matériel Quinze noms et treize verbes monosyllabiques fréquents en français ont été sélectionnés pour servir de mot-cible. Une contrainte supplémentaire a été appliquée : il devait exister au moins un nom et un verbe fréquent commençant par la même syllabe que notre mot cible. Quatre phrases ont été construites pour chaque item :

- deux phrases contenant le mot-cible :
 - « Le **chat** de mes voisins est très câlin. » ;
 - « Je regarde le **chat** qui saute sur le muret d'en face. » ;
- une phrase contenant un piège de la même catégorie grammaticale que la cible :
 - « Le **chapeau** de paille de ma soeur la protège bien des insolations. » ;
- une phrase contenant un piège de la catégorie grammaticale opposée (verbe si la

cible était un nom et inversement nom si la cible était un verbe) :

« Il **chatouille** sa nièce dès qu'il la voit. »

Le sens des phrases ne contraint pas l'apparition du mot cible mais ne l'exclue pas non plus. La position de la cible piège variait selon les phrases entre début, milieu et fin de la phrase. Les différents types de phrases pour un même item ont été répartis dans quatre blocs différents. Les blocs étaient contrebalancés pour le type de phrases contenu. Pour chaque sujet, les phrases d'un bloc étaient jouées dans un ordre semi-aléatoire afin d'éviter que deux mêmes cibles se suivent.

Procédure Le programme régissant le déroulement de l'expérience a été écrit en python. Le passage de l'expérience a eu lieu dans des cabines insonorisées. Le test durait vingt-cinq minutes, entraînement compris. Les participants ont passé l'expérience individuellement. Ils voyaient d'abord un mot-cible qu'ils devaient ensuite détecter le plus rapidement possible dans une phrase présentée de façon auditive. Ils étaient prévenus que certaines phrases contenaient des pièges, à savoir des mots commençant comme la cible. Ils devaient presser la barre d'espace le plus rapidement possible uniquement lorsqu'ils détectaient le mot-cible. Durant l'entraînement contenant dix phrases, le programme corrigeait les réponses des sujets et leur fournissait leur temps de réaction. S'ils répondaient plus d'une seconde après le début du mot cible dans la phrase, il leur était demandé de répondre plus rapidement. S'ils avaient répondu de façon incorrecte, la phrase était rejouée pour qu'ils comprennent leur erreur. A la fin de l'entraînement, l'expérimentateur s'assurait de la bonne compréhension de la tâche puis le test commençait. Pendant le test, les sujets ne recevaient plus de correction. Le mot-cible apparaissait à l'écran pendant 1.5 secondes et était suivi d'un blanc de 0.5 seconde à la suite duquel la phrase était jouée. L'essai suivant commençait 3 secondes après la fin du mot critique de l'essai précédent. Une courte pause était proposée par le programme au milieu de l'expérience : les sujets appuyaient sur la barre-espace pour recommencer lorsqu'ils étaient prêts.

Participants Vingt-et-une personnes, douze femmes et neuf hommes, d'âge moyen 23.9 ans (de 20 à 36 ans) et de langue maternelle française ont participé à cette étude.

Résultats

Seules les réponses fausses alarmes ont été conservées pour l'analyse, c'est-à-dire lorsque les sujets ont répondu pour les compétiteurs. La figure 1.14 montre le pourcentage moyen de fausse alarme dans les deux conditions de compétiteurs pour les deux catégories de la cible : nom (en gris foncé) ou verbe (gris clair). En moyenne les su-

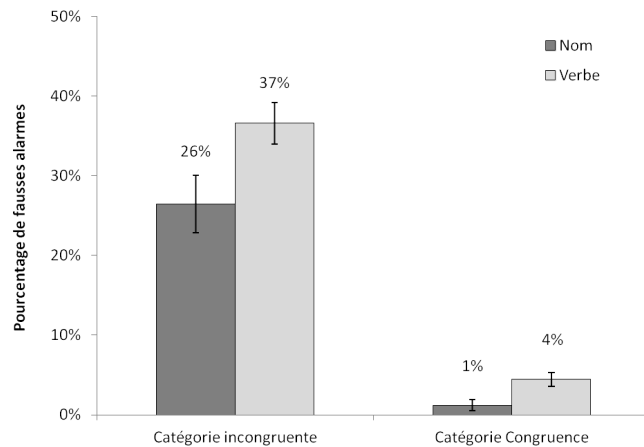


FIGURE 1.14 – Pourcentage moyen de fausse alarme pour les différentes catégories de compétiteur. Lorsque la cible à détecter était un nom, le graphe est représenté en gris foncé et lorsque c'était un verbe, en gris clair. La catégorie grammaticale du compétiteur est congruente si le compétiteur et la cible à détecter partagent la même catégorie, elle est incongruente si le compétiteur et la cible sont de catégories syntaxiques différentes.

jets ont produit 32% de fausses alarmes lorsque la cible et le compétiteur partageait la même catégorie syntaxique tandis qu'ils en produisent environ 3% lorsque le compétiteur provient de la catégorie grammaticale opposée (à savoir nom si la cible était verbe et inversement pour les cibles noms). Une analyse de variance par sujet a été réalisée sur le nombre de fausses alarmes. Le facteur principal Condition à 2 niveaux (catégorie du compétiteur congruente ou incongruente) est très significatif $F(1, 20) = 158, p < 10^{-10}$ et appuie le fait que les sujets se trompent beaucoup plus lorsque la cible et le compétiteur possèdent la même catégorie syntaxique. Le facteur principal Catégorie à 2 niveaux

(cible nom ou verbe) est aussi significatif $F(1, 20) = 11, p < 0.004$, montrant que les sujets répondent globalement plus aux cibles verbes. L'interaction des deux facteurs n'est pas significative $F(1, 20) = 2.3$ ce qui montre que le comportement de fausse alarme pour un compétiteur de la même catégorie grammaticale que la cible n'est pas différent entre les cibles noms et les cibles verbes.

Discussion

L'analyse des fausses alarmes des sujets montre que les sujets répondent au mot-piège uniquement lorsque celui-ci partage la catégorie grammaticale de la cible, de la même façon pour les noms et les verbes. Ainsi les sujets qui répondent assez vite pensent détecter *chat* alors que *chapeau* était présenté, par contre ils ne se laissent pas piéger par le verbe *chatouiller* qui commence pourtant par les mêmes phonèmes.

Les premiers phonèmes d'un mot ne pré-activent pas tous les mots qui débutent par ces phonèmes. En effet, dès qu'il existe des indices syntaxiques pour spécifier la catégorie grammaticale du mot, seul un sous ensemble de mots de la même catégorie commençant les mêmes phonèmes sont activés dans le lexique. La différence observée dans les taux de fausse alarme spécifique à la catégorie grammaticale montre que cet indice est utilisé en temps réel pour reconnaître un mot. Dans le cadre de l'accès lexical, cette caractéristique syntaxique constitue un indice important puisqu'il permet de réduire le nombre d'items pré-activés lors de la prédiction. Les données sont donc en faveur du modèle présenté figure 1.13 page 46 selon lequel la syntaxe contraint l'accès au lexique. En plus des déterminants ((Gregory et al., 2012) pour une expérience de facilitation de production pour nommer un objet) et du genre dont on sait qu'il contraint l'accès lexical chez les adultes (Dahan et al., 2000; Lew-Williams and Fernald, 2010) et les enfants (Van Heugten and Shi, 2009), la catégorie syntaxique joue donc un rôle similaire. (Magnuson et al., 2008) a utilisé un paradigme basé sur un lexique artificiel pour arriver à une conclusion similaire. La catégorie grammaticale d'un mot et son genre permettent d'optimiser et d'accélérer l'accès lexical et pourraient être cruciaux dans des cas d'homophonie comme souligné

dans le début de cette section.

Dans le cadre de l'acquisition lexicale, les enfants ne connaissent pas forcément tous les mots auxquels ils sont confrontés. Ce mécanisme de calcul de la catégorie syntaxique pourrait être pertinent même en absence de lexique pour contraindre les sens possibles d'un mot, à savoir objet pour un nom ou action pour un verbe. Si les adultes sont capables de réaliser une analyse syntaxique en temps réel en absence d'information lexicale, alors il est possible que les enfants le fassent aussi. On peut en effet supposer jusqu'à preuve du contraire, que les enfants et les adultes partagent une même architecture cognitive.

1.7 Expérience 3 : Traitement syntaxique sans lexique chez l'adulte : utilisation des mots de fonction en Jabberwocky

La première étape consiste donc à vérifier qu'une catégorisation sans lexique est réalisable par des adultes. Millotte (Millotte, 2005) a montré que les adultes étaient capables de prédire la catégorie syntaxique d'un mot même sans lexique dans des structures grammaticales assez simples. Après un pronom, les adultes analysent le mot qui suit comme un verbe, tandis qu'après un article ils l'analysent comme un nom. Pour démontrer cette analyse des catégories syntaxiques en l'absence de lexique, elle a utilisé une langue inventée qui ressemble au français. Le Jabberwocky (d'après *Alice au pays de merveilles* de Lewis Carroll) est une variante de n'importe quelle langue puisqu'elle conserve la phonotactique de sa langue originelle⁵, ses mots grammaticaux, ses indices morphosyntaxiques et sa prosodie. Par contre, les mots de contenu sont tous remplacés par des pseudo-mots. De cette façon, les adultes qui l'écoutent se trouvent dans une situation proche de celle d'enfants de 18 mois qui n'ont pas accès au sens des mots mais dont on sait qu'ils repèrent les mots grammaticaux et la prosodie. Un exemple de phrases entendues :

5. l'organisation des sons pour former des mots suivant les contraintes propre à chaque langue

Phrase avec une cible verbe : [Tu **bamoules**][saman ti] [à mon ada]

Phrase avec une cible nom : [Une **bamoule**] [dri se froliter] [dagou]

Les sujets de cette expérience ont vu la cible *bamoule* spécifiée avec sa catégorie grammaticale (nom ou verbe) : soit *la bamoule*, soit *bamouler*. Ensuite, ils entendaient une phrase contenant la cible soit en position de nom c'est-à-dire précédée par un déterminant, soit en position de verbe précédée par un pronom. Les sujets devaient appuyer sur un bouton lorsqu'ils reconnaissaient exactement le mot cible. Par exemple, ils devaient répondre si *une bamoule* avait été présentée visuellement et que *bamoule* apparaissait en tant que nom dans la phrase auditive ; de même pour les cibles-verbe. Grâce aux indices morphosyntaxiques et à la prosodie, les participants ont été capables de catégoriser des mots dont ils ne comprenaient pas le sens, et ce de façon efficace puisque 90% des réponses étaient correctes en moyenne. Après un article, ils s'attendaient à trouver un nom, tandis qu'après un pronom, ils pensaient trouver un verbe. L'hypothèse de calcul des catégories syntaxiques des mots inconnus à partir du squelette syntaxique par les enfants est donc plausible.

Les adultes réalisent une catégorisation syntaxique en temps réel très efficace pour des structures grammaticales simples. Cependant un certain nombre de phrases pourraient être équivoques : par exemple, en français les mots de fonction *le*, *la* *les* sont ambigus puisqu'ils peuvent à la fois être des articles comme dans *le chien* et des clitiques objet comme dans *je le mange*. Dans de tels cas, le mot de fonction seul n'est pas informatif, il faut que l'enfant puisse utiliser le contexte syntaxique au delà du seul mot précédant le mot inconnu.

Qu'est-ce qui nous permet d'être sûr que *le/la/les* de telle phrase est un clitique accusatif ou un article ? Existe-t-il une interprétation par défaut qui pourrait être modifiée par la fin de la phrase ? Comme *le/la/les* apparaissent les trois quart du temps en tant qu'article, dans un corpus de parole dirigé aux enfants : CHILDES (Demuth and Tremblay, 2008), cette fonction pourrait être donnée par défaut à ce mot. Cette valeur par défaut pourrait être révisée si un indice venait à l'encontre de cette analyse.

On peut aussi envisager que l'on calcule la fonction de ce mot grammatical en temps réel grâce aux indices précédant le mot, qu'ils soient prosodiques ou syntaxiques. Si ces indices ne suffisent pas, les deux fonctions pourraient être concurrentes jusqu'à ce qu'une information apparaisse départageant ainsi les deux possibilités.

Notre expérience propose de placer des adultes dans une situation comparable à celle qu'a utilisée Millotte. Nous avons donc construit des phrases en Jabberwocky contenant *le/la/les* utilisés soit en tant qu'article soit en tant que clitique accusatif. Dans cette expérience, nous testons si les adultes sont capables de catégoriser des mots absents de leur lexique, même dans des contextes grammaticaux potentiellement complexes, et ce en temps réel. Si les adultes y parviennent, alors il est possible que les bébés effectuent un traitement syntaxique semblable.

Nous avons testé deux structures syntaxiques différentes contenant les homophones *le/la/les*. Elles testent particulièrement l'influence des mots grammaticaux dans la construction de la structure syntaxique puisque les phrases d'une même paire ne diffèrent que par le mot grammatical qui est placé devant *le/la/les*, tandis que la structure prosodique est identique.

Matériel et méthodes

Matériel Seize pseudo-mots cibles ont été inventés. La moitié d'entre eux était monosyllabique, l'autre bisyllabique. Pour créer les phrases en Jabberwocky, des phrases en français contenant les contraintes structurales et prosodiques propres à chaque condition ont d'abord été construites. Pour chacun des seize items cibles, trois paires de phrases ont été créées, une pour chacune des deux structures grammaticales ainsi qu'une dernière testant l'influence de la structure prosodique, expérience qui sera détaillée séparément à la fin de ce chapitre. Dans la première phrase de chaque paire, appelée par la suite « phrase nom », la cible apparaît en tant que nom ; dans la seconde, la cible apparaît en tant que verbe, ce type de phrase sera nommé « phrase verbe » par la suite. Quarante-deux phrases ont ainsi été composées avec leur équivalent en Jabberwocky. Dans

ces dernières, tous les mots de la classe ouverte ont été remplacés par des non-mots ayant le même nombre de syllabes, de sorte que le locuteur puisse lire ces phrases avec une prosodie naturelle pour les enregistrer.

Condition 1 : Adverbe/le-la-les VS. Pronom/le-la-les

Phrase nom

[Comme la douceur] [est annoncée], [les feuilles] [vont repousser] [sur les arbres]

[Comme la **crépouille**] [est dacrinée], [les meuches] [gont doriper] [sur les atrules]

Phrase verbe

[Tu la détends] [de son stress] [en la massant]

[Tu la **crépouilles**] [de son prasse] [en la mèrant]

Chaque paire de phrase contient un mot grammatical monosyllabique suivi de *le/la/les* qui peuvent être soit article, soit clitique objet. Le mot grammatical ambigu est lui-même suivi de la cible à détecter. Les fins de phrase d'une même paire sont différentes. La seule différence jusqu'au mot cible entre les phrases où la cible apparaît en tant que nom et celles où elle est utilisée en tant que verbe est donc la nature du premier mot grammatical monosyllabique.

Condition 2 : que/le-la-les VS. qui/le-la-les

Phrase nom

[Tu ne prendras pas] [le pantalon] [que la vendeuse] [te conseille]

[Tu ne douras pas] [le nabalon] [que la **crépouille**] [te maleille]

Phrase verbe

[Sens-tu] [cette chaleur] [qui la réchauffe] [tout doucement] ?

[Doules-tu] [cette bouneur] [qui la **crépouille**] [tiau doulement] ?

Les phrases de la condition 2 contiennent toutes des propositions relatives. Dans les phrases Nom, la cible est précédée de *que + le/la/les*, le mot grammatical ambigu est alors un déterminant, tandis que dans les phrases verbes elle est précédée de *qui + le/la/les* où *le/la/les* sont alors des clitiques objet.

Toutes les phrases ont été enregistrées par une locutrice française dans une cabine

insonorisée à une fréquence d'échantillonnage de 44100Hz puis converties à 22050Hz, éditées et découpées grâce aux logiciels Adobe Audition et Praat (Boersma and Weenink, 2010). Le début et la fin des mots cibles ont été marqués sur le logiciel praat. Afin d'avoir une prosodie correcte et naturelle, les phrases ont été enregistrées deux par deux. La phrase en français était d'abord prononcée et servait de modèle pour la prosodie de la phrase en Jabberwocky. La prosodie de toutes les phrases est donc grammaticale.

Les sujets ont tous entendu toutes les phrases une seule fois. La cible précédant la phrase entendue était contrebalancée à travers les groupes auquel ils appartenaient. Dans chacune des structures testées, pour une phrase-test donnée, la moitié des sujets a vu un mot-cible de catégorie syntaxique identique à celui entendu dans la phrase test, tandis que l'autre moitié a vu un mot-cible de catégorie syntaxique différente de celui entendu dans la phrase-test. Les items ont été séparés en deux groupes appariés pour la structure syllabique et le nombre de syllabes. Pour un item donné, les deux phrases test (nom et verbe) apparaissent en version coïncidente pour la moitié des sujets, et non coïncidentes pour l'autre moitié des sujets. Deux listes expérimentales ont ainsi été créées. A l'intérieur d'une liste expérimentale, les deux phrases test d'un même item (phrase nom et phrase verbe) étaient placées l'une dans la première moitié de l'expérience, l'autre dans la seconde moitié. Chaque moitié de liste contenait le même nombre de phrases nom et verbe ainsi que le même nombre d'essais congruents et incongruents. Les sujets du groupe 1 entendaient d'abord la première moitié de la liste puis la seconde, ceux du groupe 2 d'abord la seconde puis la première. De même les groupes 3 et 4 entendront les deux moitiés de la liste 2.

De plus, douze phrases d'entraînement ont été créées pour permettre aux sujets d'assimiler la consigne et de répondre lorsqu'ils repéraient le mot de la bonne catégorie grammaticale et pas seulement lorsqu'ils entendaient la forme phonologique de la cible.

La liste de tous les stimuli utilisés se trouve en annexe [A page 221](#).

Procédure Le programme régissant le déroulement de l'expérience a été écrit en python. Le passage de l'expérience a eu lieu dans des cabines insonorisées. Le test durait

une vingtaine de minutes, entraînement compris. Les participants ont passé l'expérience individuellement. Une explication orale portant sur le déroulement de l'expérience complétée par des consignes écrites débutait l'expérience. Les sujets réalisaient tout d'abord un entraînement pour se familiariser avec le Jabberwocky et la tâche. L'expérimentateur restait dans la pièce durant l'entraînement pour répondre aux éventuelles questions des sujets et vérifier que les consignes avaient bien été comprises. Les essais se déroulaient de la manière suivante : tout d'abord un non-mot-cible inventé, en Jabberwocky, s'affichait à l'écran, puis les sujets entendaient au travers d'un casque une phrase en Jabberwocky. Leur tâche était d'appuyer sur la barre espace le plus rapidement possible s'ils reconnaissaient le mot-cible dans la phrase. Durant l'entraînement, les sujets comprenaient qu'ils devaient répondre uniquement lorsque le mot exact était entendu et non un mot ressemblant mais d'une catégorie grammaticale différente. Par exemple, s'ils devaient répondre à *gouner*, ils ne devaient pas appuyer sur le bouton s'ils entendaient : « Le goune palombe jigoulamment tous les joibes ». Durant l'entraînement l'ordinateur corrigeait les réponses des sujets et leur fournissait leur temps de réaction. S'ils répondaient plus d'une seconde après le début du mot cible dans la phrase, il leur était demandé de répondre plus rapidement. S'ils avaient répondu de façon incorrecte, la phrase était rejouée pour qu'ils comprennent leur erreur. A la fin de l'entraînement, le test commençait. Dans cette partie d'une quinzaine de minutes, les participants se retrouvaient seuls et n'étaient plus corrigés. Au milieu de l'expérience, le programme se mettait en pause et le sujet devait appuyer sur une touche pour reprendre le test. A la fin du test, l'expérimentateur relevait les impressions des sujets face à la tâche, la difficulté apparente ainsi que d'éventuelles structures typiques de phrases reconnues dans le test.

Participants Vingt-quatre personnes, sept hommes et dix-sept femmes d'âge moyen 22.6 ans (de 19 à 34 ans), ont participé à cette expérience. Aucun n'avait entendu d'autre langue que le français durant ses trois premières années de vie. Aucun n'a rapporté de problème auditif ou de langage. Les sujets étaient répartis de façon égale entre les quatre groupes.

Résultats

La figure 1.15 montre le pourcentage de réponse Nom et Verbe selon le type de phrase en fonction des deux types de construction syntaxique testées. Les sujets ont détecté correctement la cible Nom dans des phrase Nom dans 91% des cas, ils ont cru détecté cette même cible dans les phrases verbe dans 15% des cas, ce sont les fausses alarmes.

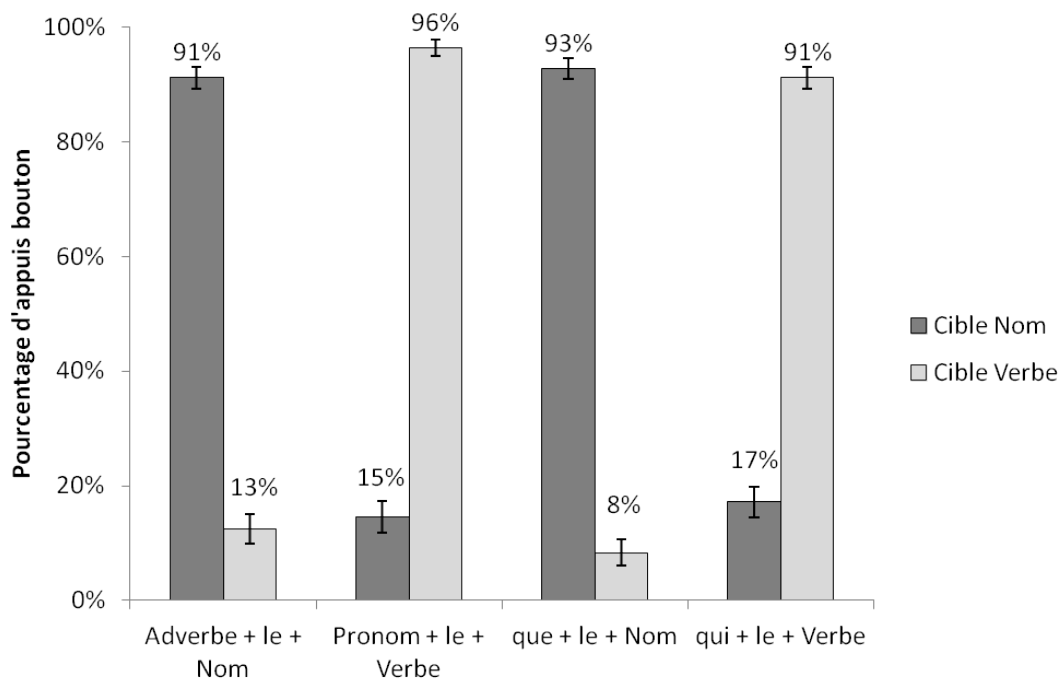


FIGURE 1.15 – Pourcentage moyen de réponse de type nom ou verbe donnée en fonction du type de cible entendue. Les sujets ont répondu dans 91% des cas lorsque la cible à détecter était un nom et que la phrase contenait effectivement cette cible dans une position de nom, c'est le taux de réussite. Cette même cible Nom a été détectée dans 13% des cas lorsqu'elle était présentée dans un contexte de verbe (du type *je le + cible*), c'est le taux de fausse alarme pour la cible Nom. De la même façon pour la cible verbe, les sujets l'ont détecté dans 96% des cas lorsqu'ils entendaient effectivement cette cible dans un contexte de verbe et ils l'ont repéré dans 15% des cas lorsqu'elle était présentée dans un contexte de Nom du type *Adverbe le + cible*. Les sujets repèrent bien mieux la cible dans les phrases où elle est présentée dans le contexte syntaxique correct.

Deux analyses de variances ont été réalisées sur les pourcentages de bonnes réponses : une par sujets et une par items. L'anova par sujets (F_1) comportait quatre facteurs inter-

sujet : la condition de test (deux types de structures grammaticales), la catégorie de la phrase présentée (nom ou verbe), la coïncidence entre la cible à trouver et le mot présent dans la phrase auditive (coïncidence ou non), le nombre de syllabe du mot cible (une ou deux). L'anova par item (F_2) comptait les mêmes facteurs intra-item : condition, catégorie de la phrase, coïncidence de la cible visuelle et audio mais pas le nombre de syllabe. Un effet principal de coïncidence très fort ressort est confirmé par l'analyse $F_1(1, 23) = 760, p < 10^{-16}$ et $F_2(1, 15) = 626, p < 10^{-12}$ reflétant le fait que les sujets ont plus reconnu la cible dans les phrases où elle apparaissait dans le contexte attendu. Les sujets reconnaissent la cible nom lorsqu'elle est placée soit après *un adverbe + le/la/les*, soit après *que + le/la/les*, de même les sujets appuient plus souvent lorsque la cible verbe est précédée de *pronom + le/la/les* ou de *qui + le/la/les*. Un effet de catégorie de la phrase est mis en évidence $F_1(1, 23) = 14, p < 0.002$ et $F_2(1, 15) = 8.3, p < 0.02$, les sujets appuient un peu plus souvent pour les phrases verbes. Cependant, il est difficile d'attribuer un sens à cet effet puisque sont mélangés les détections correctes et incorrectes.

Les sujets ont donc majoritairement répondu lorsqu'ils entendaient la cible dans un contexte congruent avec la catégorie syntaxique spécifiée, tandis qu'ils n'appuyaient pas sur le bouton dans le cas du même mot employé dans une position syntaxique différente. Les sujets ont pu utiliser toute la phrase et pas seulement ce qui précède le mot cible pour calculer la catégorie syntaxique du mot. Nous avons donc restreint l'analyse aux réponses données pendant l'écoute du mot cible et jusqu'à 150ms après la fin de ce mot pour tenir compte du temps de réaction des sujets afin de nous assurer de l'intégration des seuls mots présents avant la cible.

La figure 1.16 page ci-contre présente le pourcentage de réponses données en temps réel pour les cibles noms et verbes en fonction du type de phrases entendues.

Un quart des détections correctes ont eu lieu en temps réel (26%). Une analyse de variance a été effectuée la totalité des réponses en temps réel (détection correcte et fausses alarmes) par sujet en déclarant les facteurs intrasujets condition (2 constructions gram-

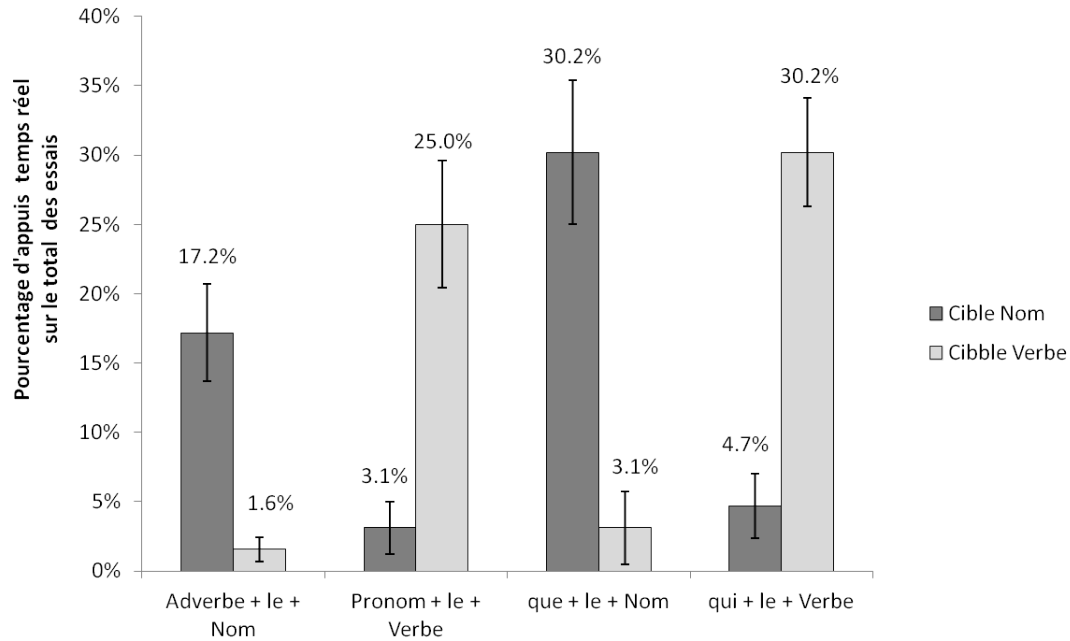


FIGURE 1.16 – Pourcentage de réponses en temps réel en fonction de la condition et de la catégorie de la cible entendue et vue. Les sujets repèrent toujours mieux la cible lorsqu'elle est présentée dans des contextes congruents avec sa catégorie syntaxique.

maticales), la catégorie syntaxique de la cible dans la phrase entendue (nom ou verbe) et la coïncidence (cible à détecter de la même catégorie que l'item présenté visuellement ou non). L'effet principal de la coïncidence, $F(1, 23) = 57, p < 1.10^{-7}$, indique que les réponses sont très majoritairement données lorsque la cible à détecter et celle présente dans la phrase audio partagent la même catégorie syntaxique. L'analyse révèle que le facteur condition est significatif $F(1, 23) = 10, p < 0.005$; la construction grammaticale *que le / qui le* a été mieux traitée en temps réel par rapport à la structure *quand le / je le*. Le facteur principal catégorie est aussi significatif $F(1, 23) = 6.2, p < 0.02$ qui reflète le fait, comme dans l'analyse globale, que les phrases verbes ont donné lieu à plus d'appuis bouton par rapport aux phrases noms. L'interaction significative entre les facteurs condition et coïncidence, $F(1, 23) = 6.3, p < 0.02$, renvoie au fait que la différence entre les réponses aux phrases coïncidentes et non coïncidentes est plus importante pour la condition 2 contenant les propositions relatives par rapport à la condition 1 contenant

l'adverbe ou le pronom.

Discussion

Le but de cette expérience était de savoir si les adultes étaient capables en absence de lexique d'utiliser les indices prosodiques et syntaxiques à leur disposition pour repérer spécifiquement une cible d'une catégorie grammaticale spécifiée. Les contextes utilisés ici sont potentiellement ambigus puisque la cible est précédée du mot de fonction *le/la/les* qui peut être soit déterminant, soit clitique objet. Ce mot grammatical n'était donc pas informatif à lui seul pour préciser la catégorie du mot qui le suivait. De plus la forme phonologique de la cible apparaissait systématiquement dans les phrases soit positionnée dans un contexte de nom soit dans un contexte de verbe. Les sujets ne détectant la cible que dans la moitié des cas, lorsque celle-ci apparaissait dans un contexte cohérent avec sa catégorie grammaticale, les sujets n'ont donc pas réalisé une tâche de reconnaissance phonologique.

En outre, même en absence de lexique et suivant un mot de fonction potentiellement ambigu, ils ont été capables de détecter la cible en temps réel spécialement dans les contextes syntaxiques dans lesquels elle pouvait effectivement apparaître. Le même mot-cible, selon qu'il est spécifié en tant que nom ou verbe, sera reconnu dans des types de phrases différentes. Ainsi une cible-nom sera détectée spécifiquement dans les contextes complexes de type *que le* et *adverbe le* et une cible verbe dans le contexte *qui le* et *pronom le*. La reconnaissance d'un mot entraîne donc plus que la seule reconnaissance de la forme phonologique puisque dans tous les types de phrase se trouve effectivement la cible et qu'elle n'est reconnue que dans des contextes syntaxiques particuliers. Les adultes sont capables d'utiliser en temps réel le contexte syntaxique dans lequel apparaît un mot pour le détecter, comme le montre l'analyse des réponses très rapides (avant la fin du mot). Nos résultats vont donc dans le sens qu'une prédiction pourrait avoir lieu quant à la catégorie syntaxique du mot qui suit. Cette prédiction se fonderait sur l'analyse des mots grammaticaux qui précèdent le mot-cible. Les sujets ont forcément analysé

plus qu'un mot grammatical, puisque le mot qui précédait immédiatement la cible était ambigu, et identique pour les phrases nom et verbe. Les sujets sont donc au minimum capables d'analyser des contextes de deux mots précédant la cible.

L'hypothèse du squelette syntaxique suppose en plus que la structure prosodique peut influencer directement l'analyse syntaxique. Afin de tester cette hypothèse en absence de lexique, nous avons réalisé une autre expérience en Jabberwocky.

1.8 Expérience 3 bis : Traitement syntaxique sans lexique : Utilisation des indices prosodiques. La Corrélation entre prosodie et syntaxe

De nombreuses recherches portant sur la prosodie de groupe intonational qui présente des frontières très marquées, montrent qu'elle est utilisée pour résoudre des ambiguïtés syntaxique dans des tâches impliquant un traitement en temps réel (paradigme de dénomination cross-modal) ou différé (jugement de grammaticalité, tâche de compréhension, jugement explicite sur la structure syntaxique) (entre autres (Marcus and Hindle, 1990; Beach, 1991; Blasko and Hall, 1998; Weber et al., 2006; Frazier et al., 2006)). Les frontières prosodiques servent aussi à détecter des pseudo-mots (Millotte et al., 2008). Il semble que la taille absolue des frontières n'importe pas tant que le rapport de taille entre plusieurs types de frontières (Carlson et al., 2001; Clifton et al., 2002; Price et al., 1991) pour l'analyse syntaxique. Les frontières prosodiques seraient donc interprétées hiérarchiquement les unes par rapport aux autres pour guider l'analyse syntaxique. Qu'en est-il en absence de lexique ? Les indices donnés par les frontières de groupes phonologique et intonational sont-ils suffisants pour permettre une prédiction de la catégorie syntaxique des items en temps réel ?

Cette dernière expérience teste spécifiquement l'influence de la taille des frontières prosodiques sur la détection d'un pseudo-mot spécifié pour sa catégorie grammaticale. Les deux membres d'une paire de phrases contiennent les mêmes constructions

de pseudo-mots jusqu'à la cible et ne diffèrent que par leur structure prosodique. Dans la première structure prosodique congruente avec l'apparition d'un nom, une prosodie de liste a été utilisée avec des frontières très marquées de groupe intonational entre les éléments de la liste.

[La fraise], [la framboise] [et les groseilles] [sont des fruits rouges]

[La draïse], [la baruche] [et les nicrons] [sont des chouï drones]

La deuxième structure prosodique dans laquelle le mot cible apparaît en position de verbe, ne contient que des frontières de plus petite taille de groupe phonologique entre le premier groupe nominal et le groupe verbal.

[Sa mère] [la retient] [en l'attrapant] [par la main]

[Sa lure] [la baruche] [en l'apramant] [par la blain]

Dans toutes les phrases le pseudo-mot-cible était toujours précédé du mot grammatical ambigu *le/la/les* qui n'était pas informatif à lui seul pour deviner la catégorie grammaticale de la cible.

Matériel et méthodes

La construction des stimuli a été décrite dans la section 1.7 page 53. Les paires de phrases commencent par un groupe nominal suivi d'un groupe contenant *le/la/les* (soit article, soit clitique objet) et la cible. La prosodie de ces 2 phrases est très différente. Dans les phrases Nom, la prosodie de liste est réalisée par des frontières de taille importante de type groupe intonational entre les éléments de la liste. Dans les phrases Verbe, la frontière prosodique entre les 2 premiers groupes est plus petite, il s'agit d'une frontière de groupe phonologique.

Phrase nom : frontières de groupe intonational

[La petite soeur], [le frère] [et le cousin] [s'amusent ensemble]

[La fichoute leille], [le **lïpe**] [et le foulin] [s'arulent ennoble]

Phrase verbe : frontières de groupe phonologique

[La jolie fille] [le prend] [avec tendresse]

[La fili jouille] [le **lipe**] [avec craillasse]

La procédure et les participants sont décrits dans le paragraphe 1.7 page 53. La liste de tous les stimuli utilisés se trouve en annexe A page 221.

Résultats

La figure 1.17 présente le pourcentage d'appuis sur le bouton en fonction du type de phrase entendu, d'une part pour toutes les réponses (à gauche) et d'autre part, pour les réponses données en temps réel (à droite).

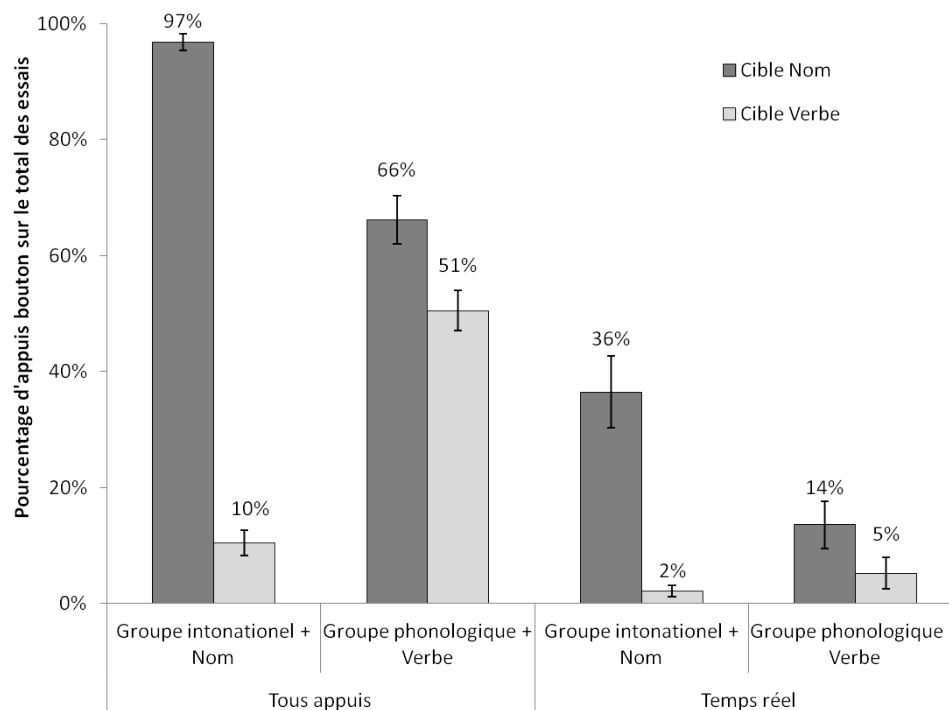


FIGURE 1.17 – Pourcentage d'appuis bouton en fonction de la phrase entendue. Analyse de tous les appuis bouton (à gauche) et des appuis en temps réel (à droite).

L'analyse de variance pour tous les appuis bouton contient 3 facteurs intrasujet (Catégorie de la cible présente effectivement dans la phrase audio, Coïncidence de la cible visuelle et auditive, Nombre de syllabe de la cible). Elle révèle un effet principal de coïncidence $F(1, 23) = 91, p < 10^{-8}$ soulignant le fait que les réponses diffèrent significativement entre les phrases nom et les phrases verbe. Un effet principal de catégorie

$F(1, 23) = 4.3, p < 0.05$ est également mis en évidence. L'interaction entre la catégorie et la coïncidence $F(1, 23) = 270, p < 10^{-13}$ signale le fait que la prosodie de liste utilisée pour les cibles nom permet de reconnaître le nom tandis que la prosodie associée aux cibles verbe contenant une plus petite frontière prosodique ne permet pas l'identification de la cible. (Interaction triple aussi). La même analyse de variance restreinte aux réponses données en temps réel (pendant que le mot était prononcé) ne tient pas compte du nombre de syllabe de la cible. Les facteurs catégorie et coïncidence sont significatifs, respectivement $F(1, 23) = 22, p < 10^{-4}$ et $F(1, 23) = 21, p < 10^{-4}$, ce qui montre d'une part que les phrases nom et verbe donnent des réponses différentes et d'autre part que les sujets réussissent la tâche avec plus de réponses correctes que de fausses alarmes. L'interaction entre les deux facteurs $F(1, 23) = 28, p < 10^{-5}$ reflète le même phénomène que décrit précédemment pour toutes les réponses.

Discussion

Les sujets devaient reconnaître un pseudo-mot cible dont la catégorie syntaxique était spécifiée dans des phrases en Jabberwocky. Les phrases contenant les cibles-nom utilisaient une prosodie de liste, présentant donc de grandes frontières de groupe intonational entre les éléments coordonnés : [La fichoute leille], [le **lipe**] [et le foulin] [s'arulent ennoble]. Les phrases contenant les cibles-verbe présentaient des frontières de groupe phonologique : [La fili jouille] [le **lipe**] [avec craillesse]. Les résultats montrent que les individus sont capables de reconnaître spécifiquement la cible nom dans les phrases contenant les frontières de groupe intonational, tandis qu'ils sont incapables de reconnaître la cible verbe dans les phrases contenant une plus petite frontière prosodique de type groupe phonologique. La forme des réponses est la même que les réponses soient fournies en temps réel ou non, même si le nombre de réponse est moins important en temps réel, surtout pour les phrases verbes où les sujets ont plus de mal à prendre une décision.

Dans certaines circonstances les pauses entre les différents constituants phonolo-

giques d'une phrase sont très saillantes tandis que dans d'autres, elles ne sont marquées que très légèrement. Il est possible que les sujets entendant une petite frontière phonologique l'ait analysée soit comme une frontière de groupe intonational peu marquée par le locuteur soit comme une frontière de groupe phonologique normalement marquée, il était donc difficile de conclure quant à la nature de la frontière. La frontière très marquée dans la prosodie de liste, elle, ne peut être analysée que comme frontière de groupe intonational. La frontière très marquée dans la prosodie de liste, elle, ne peut être analysée que comme frontière de groupe intonational. Elle permet de spécifier tout de suite le type de contenu des groupes (groupe nominal ou verbal par exemple) au contraire de la frontière de groupe phonologique. Ainsi en entendant une grosse frontière les sujets analysent les deux premiers groupes prosodiques comme des groupes nominaux et sont capables de prédire la catégorie syntaxique du mot de contenu qui se trouve dans le deuxième groupe : il s'agit forcément d'un nom et pas d'un verbe. Par contre, en entendant une frontière de groupe phonologique, il est impossible de savoir sans information lexicale si la cible est un nom ou un verbe puisque les sujets répondent 51% du temps lorsque la cible à détecter est un verbe et 66% lorsque la cible à détecter était un nom. Le taux de réponse en temps réel pour la prosodie ne contenant que des frontières de groupe phonologique est très faible, les sujets ont donc tendance à attendre la suite de la phrase pour répondre, ce qui montre que ce type de frontière n'est pas suffisant pour prendre une décision. Les sujets attendent d'accumuler des indices supplémentaires pour juger si la cible entendue était effectivement de la bonne catégorie, dans nos phrases en absence de lexique, les indices sont insuffisants pour effectuer une catégorisation efficace.

1.9 Conclusion

L'ensemble des expériences comportementales présentées dans ce chapitre montre que les mots de fonction et la prosodie sont utilisés en temps réel pour contraindre l'accès lexical chez les enfants et les adultes. Nous proposons un modèle de traitement de la parole intégrant différentes étapes : calcul d'une structure syntaxique partielle,

reconnaissance des premiers phonèmes d'un mot et intégration de ces éléments pour prédire l'item lexical. Cette représentation du traitement de la parole chez l'adulte est adaptée à l'enfant en cours d'acquisition lexicale qui utiliserait alors les mots de fonction et les morphèmes grammaticaux pour restreindre le sens possible des mots inconnus. En effet, à l'appui de cette hypothèse, les adultes sont capables d'effectuer des calculs de la catégorie grammaticale à partir des mots de fonction et de la prosodie en absence de lexique, et ce, dans des contextes potentiellement complexes utilisant l'ambivalence de *le/la/les* soit déterminant, soit clitique objet. Il est donc possible que les enfants se servent du squelette syntaxique, c'est-à-dire des mots grammaticaux et de la prosodie, pour contraindre le sens des mots inconnus, puisque les adultes en sont capables.

Résumé du chapitre

- Nous montrons que dès 18 mois, les enfants utilisent les mots de fonction pour détecter un mot connu. En effet, ils s'attendent à trouver un nom après un déterminant mais pas après un pronom, parallèlement ils cherchent un verbe après un pronom mais pas après un déterminant.
- Une première expérience comportementale avec des adultes révèle qu'ils se servent des déterminants et des pronoms pour contraindre l'accès au lexique. Dans le cadre des modèles de cohorte, nous montrons que les adultes activent spécifiquement une partie de leur lexique dès les premiers phonèmes d'un mot en restreignant leur nombre grâce à la catégorie grammaticale.
- Les adultes utilisent également les mots grammaticaux pour calculer en temps réel la catégorie syntaxique d'un pseudo-mot en absence de lexique, dans des contextes potentiellement complexes. Ils repèrent qu'un pseudo-mot est un nom s'il suit des contextes du type *adverbe + le/la/les* ou *que + le/la/les* et que c'est un verbe s'il suit des contextes du type *pronom + le/la/les* ou *qui + le/la/les* est un verbe.
- Une dernière expérience comportementale suggère que la taille des frontières prosodiques est utilisée pour calculer la catégorie syntaxique d'un pseudo-mot en absence de lexique. Les frontières de groupe intonational dans des prosodies de liste sont informatives pour prédire un nom. Par contre, l'interprétation des frontières de groupe phonologique est ambiguës et ne permettent pas de découvrir la catégorie grammaticale d'un pseudo-mot en absence de lexique.
- Nous présentons un modèle d'accès au lexique basé sur les informations acoustiques (premiers phonèmes et prosodie) et syntaxiques (catégorie syntaxique et genre) pour contraindre l'accès au lexique.

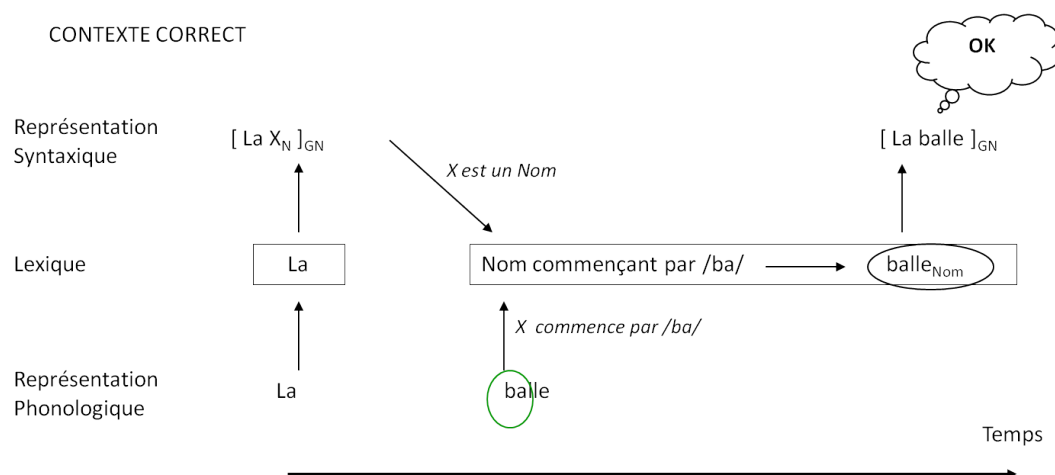


FIGURE 1.18 – Modèle représentant l'intégration des informations syntaxiques dans l'accès lexical.

Décoder les activations cérébrales dues aux structures syntaxiques et prosodiques

Sommaire

2.1	La Coordination : un nouvel outil pour étudier la prosodie et la syntaxe	69
2.2	Expérience 4 : Bases neurales du traitement des structures syntaxique et prosodique	72

2.1 La Coordination : un nouvel outil pour étudier la prosodie et la syntaxe

Les indices prosodiques et syntaxiques sont fortement corrélés. Il est donc difficile d'isoler spécifiquement et uniquement les calculs syntaxiques. Pour contourner le problème posé par la prosodie, certaines expériences utilisent l'écrit pour étudier la syntaxe (Noppeney and Price, 2004; Weber and Indefrey, 2009) et particulièrement celle sur laquelle nous nous appuyons dans les expériences suivantes (Pallier et al., 2011). Cependant, même lors d'une lecture silencieuse, il est tout à fait plausible que nous plaquions automatiquement une prosodie. Janet Fodor (Fodor, 1998) propose d'ailleurs une théorie

dans ce sens¹ dont on retrouve des corrélats dans la mesure des mouvements oculaires (Ashby and Clifton, 2005) et au niveau neuronal grâce aux potentiels évoqués à la surface du scalp enregistrés par électro-encéphalogramme (ERP)(Stolterfoht et al., 2007).

A l'oral, il existe des stratégies pour essayer de s'affranchir de l'influence de la prosodie sur la structure syntaxique par exemple en utilisant de la parole synthétisée avec une prosodie plate (exemples à fournir), ou en appliquant une prosodie de liste (Humphries et al., 2005; Humphries et al., 2006). Dans l'expérience ci-après une nouvelle méthode est développée et fait varier de façon orthogonale la prosodie et la syntaxe afin d'en dissocier les effets. Ce paradigme repose sur un fragment de la syntaxe en français : la coordination, c'est-à-dire des structures du type « Baptiste et Catherine et Perrine ». La théorie de Wagner (Wagner, 2010) sur la coordination en anglais - processus semblable à la coordination en français - nous fournit un cadre théorique pour étudier de façon conjointe les structures syntaxiques et prosodiques.

Ainsi, il est possible de découper la proposition précédente de 3 façons différentes selon la place et la taille des indices prosodiques :

- Baptiste | et Catherine | et Perrine
- Baptiste || et Catherine | et Perrine
- Baptiste | et Catherine || et Perrine

Les barres verticales représentent les frontières prosodiques ; leur nombre représente leur taille. Plus le nombre de barres verticales est grand, plus la frontière est marquée. Ces 3 façons de découper la proposition reflètent 3 arbres syntaxiques différents (figure 2.1 page suivante).

Aucune modification à l'écrit ne vient illustrer ces différences prosodiques et syntaxiques. Dans un exemple comme le précédent, rien ne prouve que l'auditeur fasse l'analyse grammaticale correcte des 3 structures, ni même qu'il soit conscient de leurs existences et des différences pragmatiques qu'elles impliquent. En effet, pourquoi un lo-

1. Fodor 1998 : The Implicit Prosody Hypothesis (IPH) : In silent reading, a default prosodic contour is projected onto the stimulus, and it may influence syntactic ambiguity resolution. Other things being equal, the parser favors the syntactic analysis associated with the most natural (default) prosodic contour for the construction.

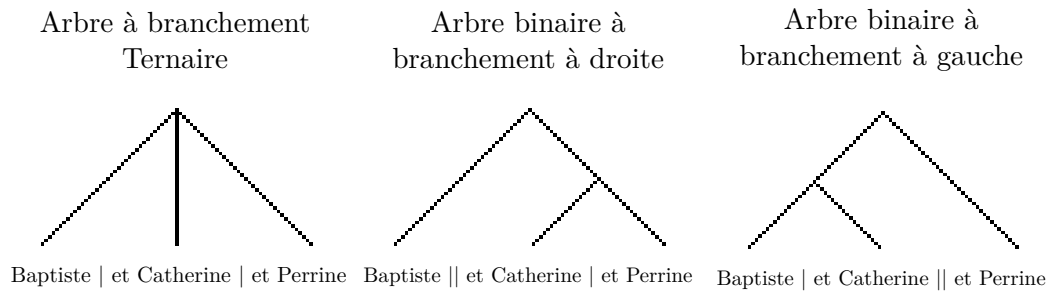
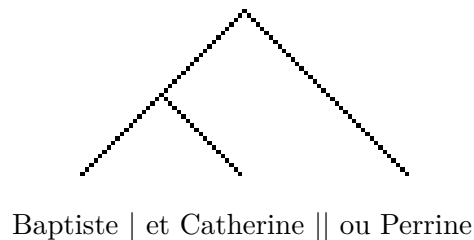


FIGURE 2.1 – Différents arbres possibles selon l’emplacement et la taille des frontières prosodiques.

cuteur introduirait-il un niveau d’analyse syntaxique qui n’apporte aucune information supplémentaire au message transmis ; cela viole le principe d’économie de Grice (Grice, 1975). Nous avons donc utilisé une alternance de connecteurs logiques « et » et « ou » qui force les groupes d’éléments et l’enchâssement² de ces groupes de façon à produire des différences entre les conditions de vérité de ces structures. Par exemple :



L’utilisation de connecteurs logiques différents rend alors obligatoire le traitement de la structure comme un arbre binaire à branchement à gauche. Les 2 premiers éléments sont groupés ensemble et comparés au troisième élément. Grâce à la coordination, la structure syntaxique varie sans différences lexicales ou sémantiques majeures.

Les indices prosodiques, dans notre cas la taille des frontières, contraignent, ici, la structure syntaxique. La théorie de Wagner permet de prédire le nombre de structures syntaxiques possibles étant donné le nombre d’éléments coordonnés. On peut ainsi coordonner 4 éléments selon 11 structures différentes étant donné le type de frontières

2. l’emboîtement

prosodiques utilisé et leur emplacement dans la proposition. Ces 11 structures peuvent être représentées par 5 arbres syntaxiques différents. La structure la plus simple comporte 3 frontières de même type. On peut complexifier cette structure en introduisant différents niveaux d'enchâssement des éléments, c'est-à-dire qu'on peut créer des sous-ensembles d'éléments et les emboîter les uns dans les autres selon différentes structures hiérarchiques. Ces différents niveaux d'enchâssement sont représentés sur la figure 2.2 page ci-contre par les 2 types de parenthèses en gras ou classique. L'enchâssement est une dimension particulière de la structure syntaxique qui a été étudiée tant du point de vue de la linguistique théorique et computationnelle ((Kayne, 1994), (Abney and Johnson, 1991)) que de la psycholinguistique (Bader and Meng, 1999) et sur laquelle nous allons nous appuyer dans cette étude.

Ainsi, la coordination fait correspondre un même arbre syntaxique à plusieurs niveaux d'enchâssements prosodiques.

2.2 Expérience 4 : Bases neurales du traitement des structures syntaxique et prosodique

Cette propriété nous a permis de faire varier de façon orthogonale la structure syntaxique (3 arbres différents) et la complexité prosodique (2 niveaux d'enchâssements ou nombre de frontières prosodiques différentes) en jouant uniquement sur les indices prosodiques, à savoir l'emplacement des frontières pour la syntaxe et leurs tailles pour la complexité prosodique. De façon cruciale, Wagner a montré que les différentes structures prosodiques, y compris celles contenant les enchâssements les plus complexes, comme par exemple « rouge ou (vert et (jaune ou bleu)) », sont réellement produites par des sujets³.

Pour cette expérience, trois types d'arbres syntaxiques différents ont été choisis parmi

3. Une expérience de production a été réalisée durant sa thèse (Wagner, 2005) sur 4 sujets pour lesquels il met en évidence des différences acoustiques reflétant les différentes structures. Il montre en outre que les frontières prosodiques ne sont pas produites d'une façon catégorielle, c'est-à-dire qu'elles n'ont pas de durée précise pour chaque type de frontière. La taille des frontières est modulée en fonction du nombre de frontières différentes à produire.

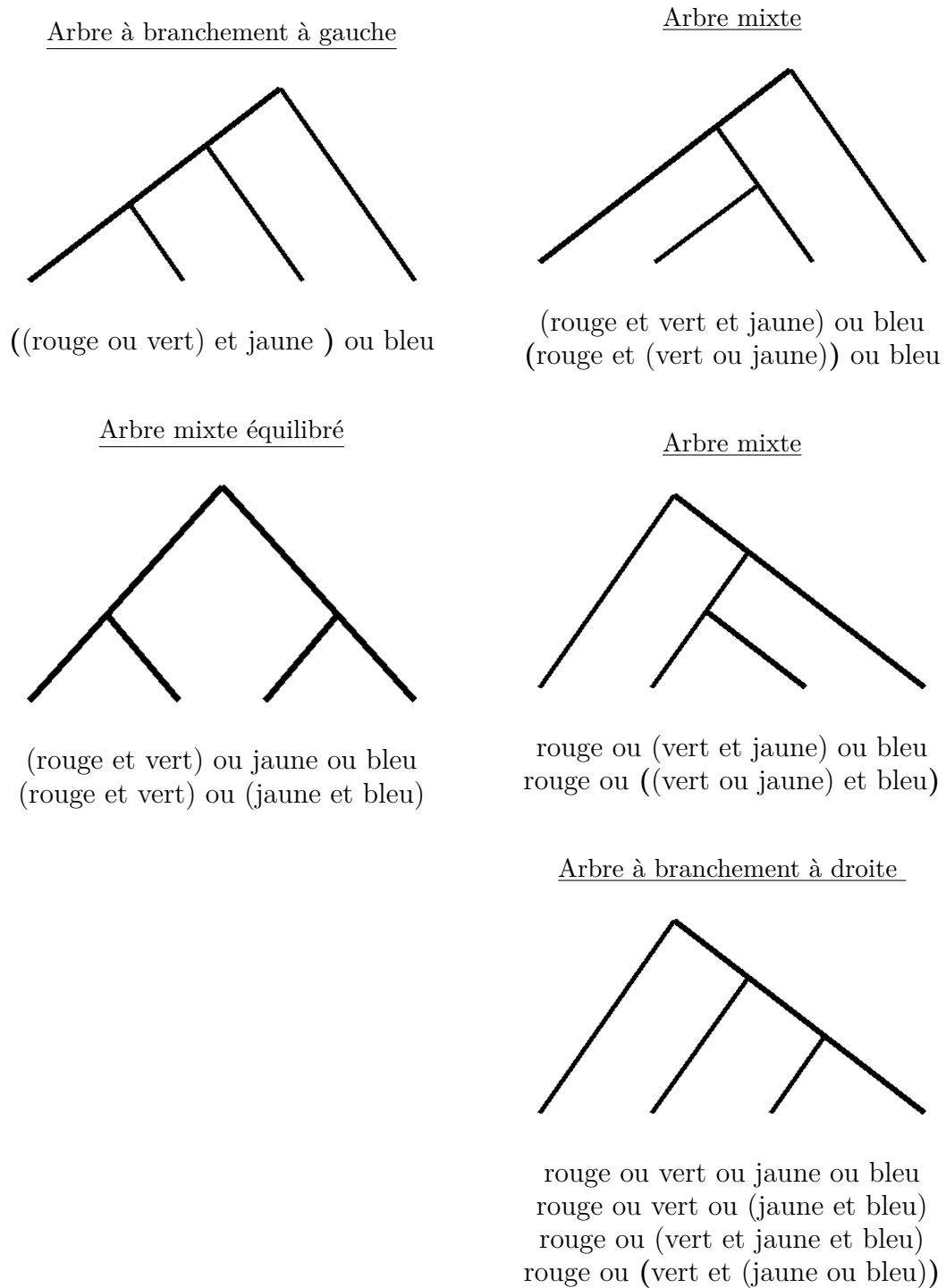


FIGURE 2.2 – Différentes structures possibles pour une coordination de 4 éléments avec leurs arbres syntaxiques associées. Les parenthèse en gras indique un niveau d'enchâssement supplémentaire par rapport aux parenthèses simples.

les 5 disponibles. Ils contiennent chacun 2 niveaux d'enchâssement possibles pour l'orthogonalité du dessin expérimental. Ont ainsi été éliminés les arbres n'ayant qu'un seul niveau d'enchâssement prosodique : l'arbre à branchement à gauche puisqu'il n'existe qu'une seule structure prosodique associée, et l'arbre mixte équilibré puisqu'il ne contient qu'un seul type d'enchâssement possible, le même dans les 2 structures prosodiques associées.

Il existe de nombreuses études sur les représentations neurales de la prosodie. La plupart d'entre elles s'intéressent aux différences entre prosodie émotionnelle et phrasale (Dirk et al., 2009). Parmi les quelques expériences qui s'intéressent spécifiquement à la prosodie phrasale (ou linguistique) la plupart contraste de la parole normale à de la parole dans laquelle on a supprimé les informations prosodiques (Humphries et al., 2005). Le problème inhérent à ce type de dessins expérimentaux est qu'il étudie le traitement neural de stimuli particulièrement peu écologiques. Il n'existe que 2 expériences IRMf publiées qui manipulent la prosodie dans un contexte naturel (Doherty et al., 2004; Ischebeck et al., 2008). Le premier (Doherty et al., 2004) s'intéresse à des aspects d'information portés par la structure (question par rapport à affirmation), domaine différent de celui que nous avons manipulé. Ischebeck et ses collègues (Ischebeck et al., 2008) ont manipulé le nombre de frontières de groupe intonational (1 ou 2) dans des phrases naturelles ou dans des phrases fredonnées (prononcée avec seulement des 'm'). Cette expérience est celle qui se rapproche le plus de notre manipulation de la structure prosodique. Seules les régions auditives sont plus activées par un plus grand nombre de frontières prosodiques de type intonational, de façon bilatérale.

Quant à la structure syntaxique, elle est habituellement étudiée en comparant des structures plus ou moins complexes, par exemple des structures à branchement à droite avec des structures à branchements à gauches plus complexes, ou encore des structures enchâssées. Un grand nombre d'expériences en neurosciences cognitives du langage se basent sur l'hypothèse qu'une région cérébrale fait partie du réseau de la syntaxe si elle est activée différemment pour différentes structures syntaxiques (par exemple (Santi and

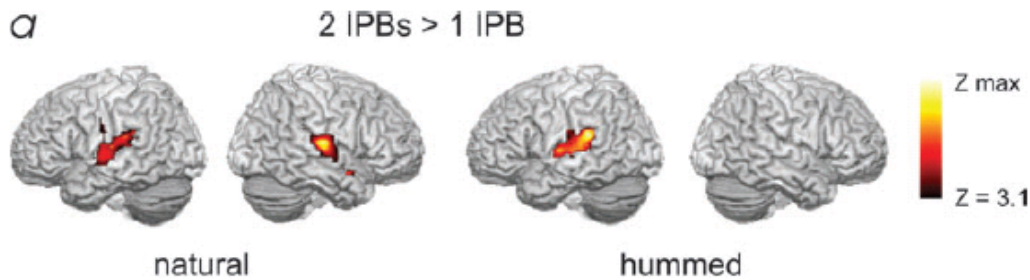


FIGURE 2.3 – Carte des régions cérébrales plus actives pour les phrases contenant 2 frontières prosodiques de type intonationnelles par rapport à celles n'en contenant qu'une. A gauche dans la parole naturelle, à droite dans les phrases fredonnées. (Ischebeck et al., 2008)

Grodzinsky, 2010)). Cette hypothèse apparemment simple se heurte à 2 points cruciaux. Premièrement, la structure syntaxique covarie avec un certain nombre de propriétés linguistiques comme la sémantique, la mémoire de travail, la prosodie, la fréquence des mots, etc. Deuxièmement, la plupart des dessins expérimentaux en IRM fonctionnelle utilisent la soustraction univariée logique, c'est-à-dire qu'on soustrait les activités cérébrales évoquées par 2 conditions différentes. Or pour que la logique soit valide jusqu'au bout, il faut que les 2 structures comparées soient différentes ou qualitativement - la condition A doit contenir un élément X qui est absent de la condition B - ou quantitativement - la condition A doit nécessiter plus de traitement que la condition B (Caplan, 2009). Dans le cas de structures syntaxiques différentes, la distinction qualitative n'est pas applicable et il n'est pas évident, à priori, de déterminer quelle structure est plus complexe qu'une autre. Même si les structures avec un enchâssement central (center embedding) entraînent un coup computationnel et neuronal plus important que les structures à branchement à droite (Stromswold et al., 1996), il est possible que la difficulté soit due à la longue distance entre le verbe et son argument⁴ plutôt qu'à la structure hiérarchique en tant que telle. Les structures à branchement à droite et à gauche sont également différenciées du point de vue computationnel dans la littérature, mais la difficulté relative de ces 2 structures dépend des hypothèses admises sur la nature

4. notre dessin expérimental nous permet de nous affranchir de cette relation de distance

du mécanisme de parser humain (Abney and Johnson, 1991), nature qui est toujours en débat actuellement.

Pour éviter ce type d'hypothèse, il existe des stratégies différentes au sein de la littérature en IRMf, comme par exemple des paradigmes d'habituation (pour une revue sur ce type de paradigme en stimulation visuelle voir (Krekelberg et al., 2006)). Ce type de paradigme repose sur l'observation que la répétition d'un stimulus induit une diminution d'activité cérébrale (Repetition suppression)⁵. Ce type de paradigme a été appliqué à des questions linguistiques. Par exemple le priming sémantique provoque une habituation en ERP (Lau et al., 2008) et en IRMf (Devauchelle et al., 2009). Quant à l'habituation syntaxique, elle a été montrée très clairement de façon comportementale en production, par exemple (Smith and Wheeldon, 2001) et en compréhension, par exemple (Potter and Lombardi, 1998). Par contre en IRMf lorsqu'on contrôle la syntaxe pour tous les autres facteurs covariants habituellement, sa pure répétition ne produit pas d'habituation (Devauchelle et al., 2009). Cependant, en séparant plusieurs aspects de la syntaxe comme le mouvement et l'enchâssement (Santi and Grodzinsky, 2010), il est possible de trouver des zones dont l'activité est réduite par la répétition du mouvement comme l'aire de brodmann 45 à gauche (gyrus frontal inférieur antérieur)⁶ (voir la figure 2.4 page suivante pour les stimuli et 2.5 page ci-contre pour les résultats).

Dans cette expérience, nous avons utilisé une stratégie différente en utilisant une procédure de classification multivariée (Cox and Savoy, 2003; Mitchell et al., 2004; Haynes and Rees, 2005; Kamitani and Tong, 2005; Norman et al., 2006). Cette approche utilise un algorithme de classification⁷ (par exemple une machine à support vectoriel, une régression logistique ou une analyse de discriminant linéaire) pour apprendre à partir

5. Cette diminution d'activité a été observée avec différentes techniques : enregistrement électrophysiologique cellulaire (Carandini et al., 1998) pour la sensibilité à l'orientation d'un objet, enregistrement des potentiels évoqués (Dehaene-Lambertz and Dehaene, 1994) en langage pour la répétition de syllabes, en IRMf (Dehaene-Lambertz et al., 2006) pour la répétition de phrase. L'idée sous jacente est qu'une même population neuronale activée plusieurs fois de suite verra son activité réduite pour des stimulations successives. Cependant les bases neurales de ces différentes manifestations de l'habituation sont toujours à l'étude (Grill-Spector et al., 2006)

6. Le gyrus frontal inférieur postérieur (aire de brodmann 44), le sillon précentral inférieur droit et le gyrus temporal supérieur gauche sont sensibles à la répétition de l'exact même arbre donc au mouvement et à l'enchâssement

7. qu'on appellera classifieur par la suite

	"Canonical" word-order (Subject Movement)	"Non-Canonical" word-order (Object Movement)
CE	I. <CE, -MOV> The boy [who <u> </u> is chasing the tall girl] is Derek	II. <CE, +MOV> The boy [who the tall girl is chasing <u> </u>] is Derek
RB	III. <RB, -MOV> Derek is the boy [who <u> </u> is chasing the tall girl]	IV. <RB, +MOV> Derek is the boy [who the tall girl is chasing <u> </u>]

FIGURE 2.4 – Stimuli présentés dans l’expérience de (Santi and Grodzinsky, 2010). Des-
sin expérimental croisant le mouvement et l’enchâssement central. CE : central embed-
ding/enchâssement central. RB : right-branching/branchement à droite. MOV : move-
ment/mouvement.

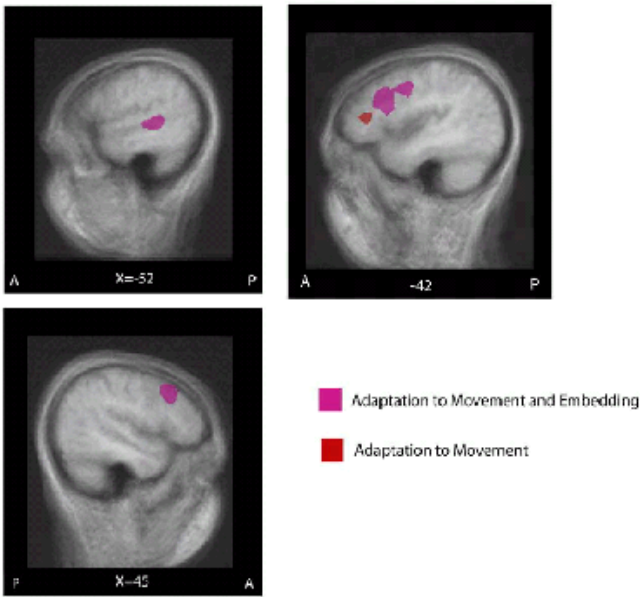


FIGURE 2.5 – Résultats de l’expérience de (Santi and Grodzinsky, 2010), adapté. Ré-
gions dont l’activité est réduite par la répétition d’aspects particuliers de la syntaxe :
le mouvement et l’enchâssement. En rouge, les régions dont l’activité est réduite par la
répétition du mouvement. En rose, les régions dont l’activité est réduite par la répétition
de mouvement et d’enchâssement.

d'un ensemble de données (étape d'entraînement) comment différencier les volumes fonctionnels provenant de 2 conditions expérimentales différentes. Dans l'étape de test, le classifieur attribue à de nouvelles images (inutilisées pendant la phase d'entraînement) la condition expérimentale dont elles sont issues. Si le classifieur est plus efficace que le hasard, alors il est capable de discriminer les conditions à partir des données. Cette technique nous fournit l'avantage d'utiliser les pattern multivariés des voxels et sera donc plus puissante que l'approche classique par soustraction qui a également été réalisée par soucis de contrôle de cette nouvelle méthode d'analyse des données. Cette classification est ici utilisée pour la première fois dans le domaine de la syntaxe et plus globalement du traitement du langage en général. Il est important de souligner que cette classification a été réalisée au niveau du groupe c'est-à-dire à travers les sujets, alors qu'habituellement, ce type d'analyse est effectué au niveau individuel⁸. L'entraînement du classifieur a lieu non plus au niveau de l'individu mais prend en compte les données d'une partie des sujets et teste l'efficacité de la classification sur l'autre partie des sujets.

L'utilisation de la coordination, fragment réduit de la grammaire française (qui inclue une seule règle de récursion) peut être vue comme un lien entre les travaux sur les grammaires artificielles (Bahlmann et al., 2008) et ceux utilisant l'intégralité de la grammaire naturelle. Les grammaires artificielles sont intéressantes par leur simplicité et le fait qu'elles permettent à l'expérimentateur de contrôler maximale-ment différentes dimensions à travers les stimuli. Cependant, comme leur nom l'indique, elles sont artificielles et il n'est pas certain qu'elles utilisent les mêmes ressources neurales qui permettent la parallélisation du calcul du langage naturel. La coordination en tant que fragment de la syntaxe naturelle partage la simplicité des grammaires naturelles.

Elle nous a permis de faire varier la structure syntaxique avec des conséquences minimales sur les autres facteurs covariant traditionnellement avec elle. En faisant varier de façon orthogonale la structure syntaxique et prosodique, nous avons été en mesure

8. On entraîne le classifieur sur une partie des données d'un sujet et on essaie de prédire les conditions dont sont issues les données sur l'autre partie des données de ce même sujet.

de séparer l'influence de chacun de ces facteurs. Enfin l'utilisation d'une analyse non conventionnelle de classification nous permet d'être plus sensible que les analyses traditionnelles puisque cette méthode nous permet de classer à partir des pattern d'activation multivariés les images cérébrales provenant des différentes conditions.

Matériel et méthodes

Participants

Vingt-huit sujets sains, droitiers, de langue maternelle française (âgés de 18 à 36 ans, moyenne, 10 femmes et 18 hommes) ont participé à cette expérience. Aucun des participants n'a rapporté de déficit auditif ni de trouble neurologique. Ils ont tous signé un consentement éclairé et reçu 80 euros en échange de leur participation au protocole. L'expérience a été approuvée par le comité éthique du Commissariat à l'Energie Atomique.

Stimuli

Chaque essai est constitué de 2 stimuli : un auditif et un visuel présentés de façon simultanée. Les stimuli auditifs sont composés de 4 groupes de mots coordonnés suivant la forme « A et/ou B et/ou C et/ou D ». Les groupes de mots utilisés (A, B, C, D dans l'exemple) sont ou des noms ou des groupes nominaux contenant un article et un nom. Les stimuli visuels sont des images contenant chacune les 4 éléments de la proposition auditive. Un sous-ensemble des 4 objets apparaît à un endroit donné de l'image (exemple figure 2.6 page suivante, le cercle bleu correspond au sous-ensemble). Dans la moitié des essais, la phrase entendue correspond à l'image vue (image correcte) et dans l'autre moitié, la phrase entendue ne reflète pas l'image (image incorrecte).

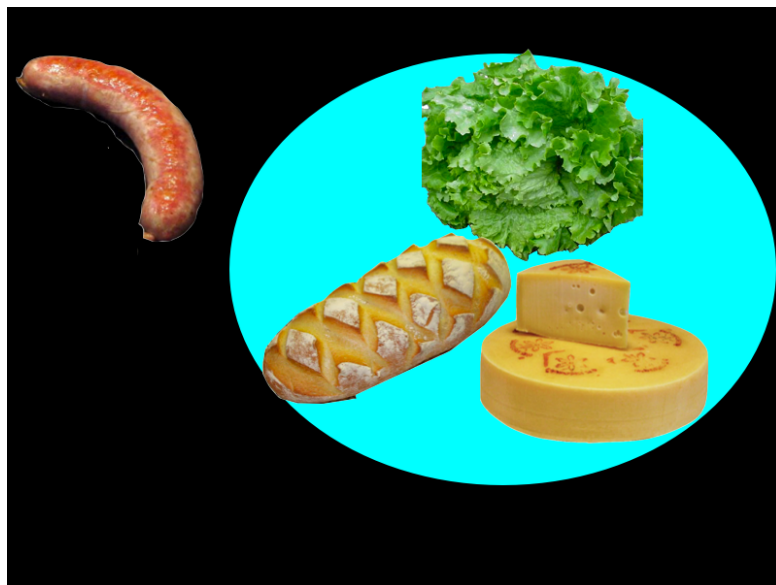


FIGURE 2.6 – Exemple d'image vue

Quatre ensembles d'objets ont été créés pour éviter la monotonie (des animaux, des objets/meubles de la maison, des aliments et des couleurs). Deux facteurs expérimentaux, la structure prosodique et la structure syntaxique, ont été croisés dans ce dessin expérimental de la façon suivante. La profondeur de la structure prosodique a été modifiée en introduisant ou un seul type de frontière prosodique dans la moitié des essais, ou deux types de frontières différant par leur taille (une grosse frontière et une plus petite) pour l'autre moitié des essais. La structure syntaxique a elle-aussi varié, non seulement grâce aux conjonctions de coordination utilisées mais aussi grâce à l'endroit où les frontières prosodiques ont été insérées. Trois types d'arbres syntaxiques ont été sélectionnés (figure 2.2 page ci-contre) : branchement à droite ($A(B(C D))$) (syntaxe 1), branchement complexe à gauche ($((A(BC))D)$) (syntaxe 2), branchement complexe à droite ($A((BC)D)$) (syntaxe 3) .

Six conditions différentes ont ainsi été créées en combinant les 2 facteurs prosodie et syntaxe (figure 2.2 page suivante).

Au total, 48 propositions ont été construites. Chaque proposition a été utilisée 2 fois (une fois avec une image correcte et une fois avec une image incorrecte) soit un total de 96

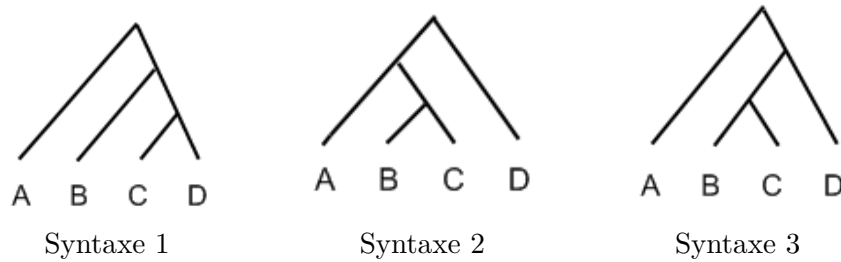


FIGURE 2.7 – Les 3 arbres syntaxiques utilisés

Condition	Syntaxe	Prosodie
Syntaxe 1 - Prosodie 1	(A (B (C D)))	A # B C D
Syntaxe 1 - Prosodie 2	(A (B (C D)))	A ## B # C D
Syntaxe 2 - Prosodie 1	((A (B C)) D)	A B C # D
Syntaxe 2 - Prosodie 2	((A (B C)) D)	A # B C ## D
Syntaxe 3 - Prosodie 1	(A ((B C) D))	A # B C # D
Syntaxe 3 - Prosodie 2	(A ((B C) D))	A ## B C # D

FIGURE 2.8 – Les 6 conditions créées. Les parenthèses () représentent les branches de l'arbre. Les # représentent les frontières prosodiques, ## est une frontière plus grande que #.

paires de stimuli audio-visuels. Les stimuli auditifs ont été enregistrés par une locutrice française entraînée. Afin que les stimuli soient le plus naturels possibles, la décision a été prise de ne pas normaliser les fichiers sonores obtenus. Ainsi la durée des stimuli (moyenne 3.4 sec, écart-type = 0.52sec) mais aussi d'autres variables acoustiques varient afin que les différentes frontières prosodiques et leur type soient saillants. Un deuxième locuteur natif a vérifié en suivant les informations prosodiques que le groupement des mots était clair et correspondait aux conditions souhaitées.

Procédure

Un essai débute par la proposition auditive et l'apparition simultanée de l'image à l'écran. L'image reste à l'écran pendant 6 secondes. Les sujets doivent vérifier si ce qu'ils entendent correspond avec ce qu'ils voient. Avant d'entrer dans le scanner, l'expérimentateur explique aux sujets qu'un magicien va essayer de deviner ce qu'ils voient. Ils doivent juger si la proposition du magicien (la proposition auditive) correspond à

l'image présentée. Si la proposition du magicien correspond à l'image, alors ils doivent appuyer sur le bouton droit (correct) sinon sur le bouton gauche (incorrect). Les paires de stimuli ont été assemblées de façon à retarder le plus possible le moment où le sujet pouvait déterminer la réponse correcte. Il est important de noter que pour réussir cette tâche, les sujets devaient forcément calculer la structure syntaxique et prosodique des phrases. Ainsi pour les essais où les sujets répondent correctement, il est certain qu'ils ont effectivement utilisé les informations prosodiques pour construire la structure voulue. Les participants ont entendu 2 blocs de 48 essais. Avant d'entrer dans le scanner les sujets se sont vu expliquer la tâche et les différentes situations visuelles qu'ils allaient devoir juger (figure 2.2).

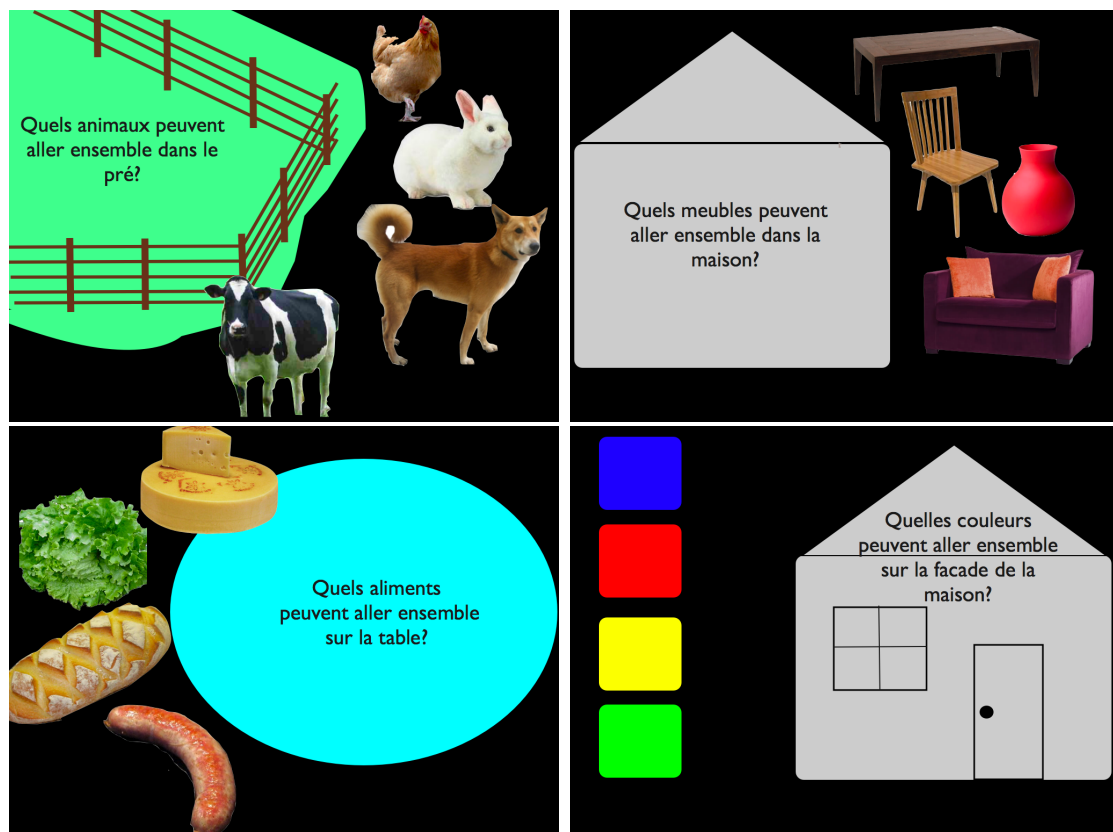


FIGURE 2.9 – Différentes situations visuelles rencontrées dans l'expérience

L'entraînement, hors du scanner, permettait aux sujets : d'abord de gérer le rythme avec lequel les stimuli apparaissaient en pressant un bouton pour passer à l'essai suivant.

Les réponses données étaient corrigées par le programme et l'expérimentateur répondait aux questions si nécessaire. Dans une deuxième partie de l'entraînement, les essais passaient au même rythme que pendant le test, les sujets apprenaient à répondre vite, en moins de 6 secondes (avant l'essai suivant) et sans être corrigé.

Acquisition des données

Les acquisitions ont été réalisées sur une IRM 3 Tesla Siemens Trio. Pour chaque participant, une image anatomique en T1 a d'abord été acquise en utilisant une séquence écho gradient 3D (TI = 900 msec, TE = 2.98 msec, TR = 2.3 sec; Taille de Voxel = 1x1x1.1 mm, champ de vue = 256, Matrice = 256x256, angle de rotation=9°, 160 coupes). Ensuite, un total de 380 scans fonctionnels ont été acquis en 2 sessions de 190 scans chacune, en utilisant une séquence EPI 2D bold (TR = 2 sec, TE = 30 msec, angle de rotation = 76°, Matrice = 64x64; Champs de vue = 192; Taille de Voxel = 3x3x3 mm; 38 coupes axiales ont été acquises selon une séquence ascendante pour couvrir le cerveau). Chaque session a duré environ 6 minutes pour 48 essais avec un SOA moyen de 7.5 sec. L'ordre dans lequel les essais ont été présentés ainsi que les variations de SOA ont été optimisés en utilisant Optseq (<http://surfer.nmr.mgh.harvard.edu/optseq/>).

Analyses des données

Le logiciel SPM8 (Wellcome Department of Cognitive Neurology, www.fil.ion.ucl.ac.uk/spm, (Friston et al., 2007)) a été utilisé pour toutes les étapes de preprocessing des données ainsi que pour l'analyse des données au niveau individuel (de premier niveau) et pour l'analyse de variance classique de groupe qui nous permet de comparer les résultats obtenus avec la méthode de classification à celle utilisée de façon standard. Les volumes fonctionnels ont été corrigés pour les mouvements puis réalignés avec l'image anatomique pour chaque sujet. Les images résultantes ont été ensuite normalisées spatialement (en utilisant des déformations non linéaires) sur le template standard défini par le standard neurologique de Montréal réalisé en moyennant 152 cerveaux. Les images ont ensuite été

lissées en utilisant un kernel gaussien isotropique de 4 mm.

Les données individuelles ont été analysées en utilisant un modèle général linéaire en déclarant 7 régresseurs : un par condition expérimentale (6) plus un régresseur regroupant tous les essais incorrects. Chaque évènement a été déclaré en tant que mini-bloc de 2 secondes dont le début correspondait à 2 secondes avant la réponse. Nous avons choisi le temps de réaction comme marqueur de nos conditions et non le début de l'essai, d'une part, parce que nous prenons ainsi en compte la variance de ce facteur, d'autre part parce que l'intérêt de cette expérience est le traitement de la structure syntaxique et non pas les processus acoustiques de bas niveau (Caplan et al., 2002). Le temps de réaction semble plus lié aux processus de calculs syntaxiques nécessaires à la prise de décision pour notre tâche qu'au moment où l'information acoustique est traitée par les aires auditives primaires. Les évènements ont en plus été modélisés en utilisant des modulateurs paramétriques pour tenir compte de la réponse motrice et de la durée du fichier sonore. Pour chaque sujet nous avons ainsi obtenu 6 images de contraste représentant nos 6 conditions expérimentales. Ces images ont ensuite servi de base aux analyses de groupe : premièrement à l'analyse de variance traditionnelle et deuxièmement à l'analyse de classification multivariée. Afin d'obtenir le même nombre de voxels par sujets, les images de contraste ont été masquées par l'intersection des masques individuels obtenus grâce à l'analyse du premier niveau SPM. De plus, l'amygdale, les régions olfactives, le cuneus et la région occipitale (toutes définies par l'atlas AAL) ont été exclus de l'analyse, nous laissant ainsi avec un masque composé de 54 808 voxels.

L'analyse de groupe utilisant la classification multivariée a été réalisée par une autre équipe de neurospin (INRIA Parietal, Univ Paris-Sud, 91191 Gif/Yvette, France) dans le but de tester l'application d'outils développés dans notre centre de recherche. Grâce aux méthodes multivariées de décodage de données d'IRMf (Mitchell et al., 2004), on s'affranchit du problème des comparaisons multiples tout en tenant compte de la structure multivariée des données d'IRMf. En effet dans ce type de données, il existe des corrélations naturelles entre les voxels voisins mais aussi entre des régions fonctionnellement

connectées mais plus éloignées spatialement. Ces observations expliquent pourquoi ce type de méthode d'analyse de données est statistiquement plus sensible que les analyses statistiques standards de parametric mapping (Kamitani and Tong, 2005; Haynes and Rees, 2006).

Cependant le décodage d'images d'IRMf repose sur une fonction de prédiction qui souffre de la dimension des données ; le nombre d'éléments (ici 54 808 voxels) est beaucoup plus grand que l'échantillon d'images (ici 132). En conséquence, les méthodes d'inférence sont sujettes au surapprentissage⁹ (overfit) ce qui conduit à de mauvaises performances lorsqu'on essaie de prédire la classification de nouvelles données.

Pour résoudre ce problème, une solution standard consiste à réduire le nombre d'éléments (ou voxels) avant la classification (Haynes and Rees, 2006). On peut par exemple utiliser une partie des données comme ensemble d'entraînement du modèle et choisir les voxels les plus informatifs de ce sous-ensemble de données pour le modèle de test qui cherchera alors parmi ce petit nombre de voxels, lesquels sont sensibles aux conditions dans l'ensemble des données. A ce jour, la méthode la plus courante pour sélectionner les voxels pour réduire leur nombre est basée sur une analyse de variance (anova) qui évalue chaque voxel du cerveau indépendamment. Les éléments ainsi sélectionnés peuvent être redondants et ne sont pas contraints par des informations spatiales. Les voxels sélectionnés peuvent provenir de régions distantes et sans relation anatomique ou fonctionnelle sans qu'on puisse voir émerger de structures cohérentes. Il est important de comprendre qu'une telle réduction du nombre de voxels part du principe que l'information contenue dans un seul voxel est pertinente pour effectuer la prédiction à un niveau inter-individuel. Or malheureusement, il n'existe pas de correspondance parfaite au niveau du voxel entre 2 individus puisque les variations anatomiques sont grandes et les procédures de normalisation imparfaites.

9. Le surapprentissage consiste dans le fait d'apprendre sur des données avec beaucoup d'informations et donc beaucoup de bruit. Le modèle va apprendre à classer les images à partir de beaucoup de données dont la plupart sont du bruit. A partir de cet apprentissage, le modèle va essayer de généraliser c'est-à-dire de prédire la provenance de nouvelles données. Comme le modèle a appris beaucoup de choses inutiles, la prédiction va contenir beaucoup d'erreurs. On a alors une accumulation d'erreurs due au bruit dans les données.

Dans un contexte de décodage inter-sujet, il existe un meilleur moyen de réduire le nombre de voxels : on va agglomérer les voxels ensemble plutôt que d'en sélectionner certains. On va donc remplacer le signal du voxel par des moyennes locales (ou parcelles) (Flandin et al., 2002; Thirion et al., 2006; Michel et al., 2010). Le modèle prédictif utilisera alors ces parcelles basées sur la moyenne du signal IRMf réduisant ainsi le nombre d'éléments pris en entrée. En partant d'environ 50 000 voxels, le nombre de parcelles ne sera plus alors que de quelques centaines voire quelques milliers.

L'agglomération d'éléments est réalisée en utilisant l'algorithme de Ward (Ward, 1963), qui est une méthode de clustering agglomérative hiérarchique sans modèle. Cet algorithme est basé sur un critère de minimisation de la variance afin de s'assurer que les parcelles moyennées sont effectivement une représentation précise du signal. Des contraintes de connectivité fonctionnelle ont été ajoutées à cet algorithme pour prendre en compte les informations spatiales et fonctionnelles. Les parcelles ainsi estimées forment des ensembles de voxels logiques et les moyennes obtenues donnent une bonne représentation des données.

Le signal moyen de chaque parcelle est alors entré dans la classification de régression logistique parcimonieuse (sparse logistic regression) (Hastie et al., 2003; Ryali et al., 2010) implémentée comme décrit dans (Fan et al., 2008). L'algorithme de classification est d'abord entraîné sur tous les sujets sauf 1 (ici 21 sujets pour l'entraînement). Ensuite ce classifieur est utilisé pour classer, selon les conditions, les images du dernier sujet non utilisé pendant l'entraînement. Cette procédure de cross-validation en laissant un sujet à part pour l'entraînement nous permet de mesurer l'efficacité du classifieur.

Les paramètres du classifieur correspondent d'une part au nombre de parcelles pour l'étape d'agglomération des voxels, d'autre part au paramètre de régularisation pour la régression logistique. Ces 2 paramètres sont calculés sur l'ensemble de données d'entraînement (21 sujets) en utilisant une procédure de cross-validation en 3 pans (7 sujets par pan). Le nombre de parcelles varie entre 400, 800, 1200 et 1600 et le paramètre de régularisation de la régression logistique est optimisé sur une grille logarithmique de

10 valeurs entre 1 et 10^7 . Cette méthode de triple cross-validation est appelée cross-validation imbriquée.

L'efficacité du classifieur est déterminée en pourcentage de classification correcte sur l'étape de test (sur le dernier sujet). Si le classifieur répond au hasard, alors le pourcentage de classification correct sera de 50% car les classes sont contrebalancées dans cette étude.

Résultats

Résultats comportementaux

Tout d'abord, nous avons calculé le pourcentage de réponses correctes par participant. Les sujets ayant obtenu moins de 60% de réponses correctes (sur 96 essais) ont été exclus des analyses (6 sujets sur 28). Les participants restants ont obtenu des scores entre 65% et 90% de réponses correctes (moyenne = 77%, Ecart-type = 7%). Les résultats : scores et temps de réaction sont représentés sur la figure [2.10 page suivante](#).

Les taux de réponses ont été analysés en utilisant un modèle linéaire à effet mixte avec une fonction de lien binomiale et les temps de réaction l'ont été grâce à un modèle mixte linéaire (Gelman and Hill, 2007; Baayen, 2008). Dans les 2 analyses, 4 facteurs fixes ont été modélisés : syntaxe (3 niveaux), prosodie (2 niveaux), contexte (nombre de *et* et de *ou*), durée ainsi que l'interaction entre la syntaxe et la prosodie ; un facteur aléatoire a été déclaré : les participants. Le résultat de ces modèles est présenté sur la table [2.1 page suivante](#).

La prosodie 2 (avec 2 types de frontières différentes) donne des temps de réaction significativement plus longs et des scores plus bas. La condition syntaxe 2 (arbre à branchement complexe à gauche) donne des temps de réaction plus courts par rapport à la condition syntaxe 1 (branchement à droite). Les effets du contexte (sur le temps de

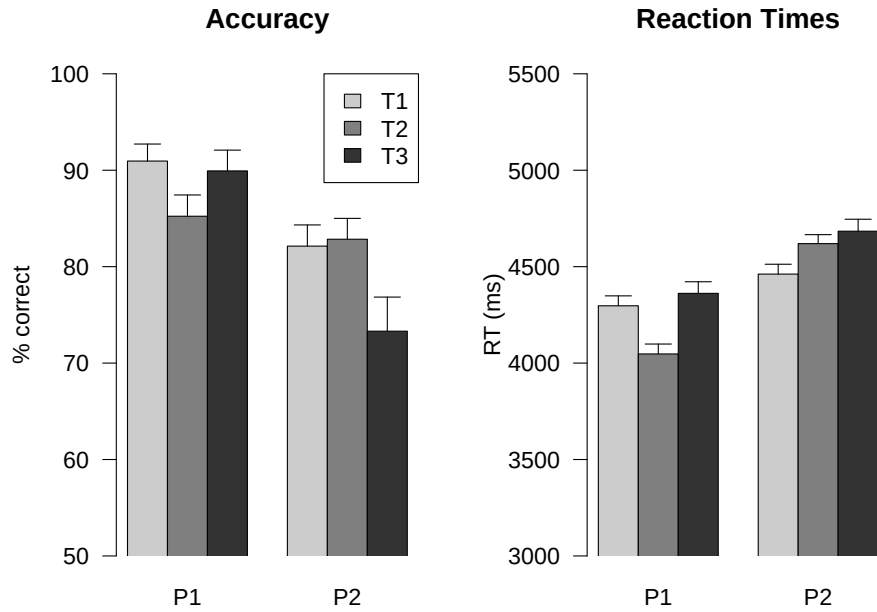


FIGURE 2.10 – Résultat de la tâche comportementale. À gauche les scores, taux de bonnes réponses en fonction des différentes structures prosodiques et syntaxiques et à droite, les temps de réactions en fonction des mêmes structures. P1 et P2 sont les 2 conditions prosodiques, P1 ne contient qu'un niveau de frontière prosodique, P2 contient 2 niveaux de structure enchâssées. T1, T2, T3 sont les types de structure syntaxique testés, T1 étant l'arbre à branchement à droite, T2 l'arbre à branchement complexe à gauche et T3 l'arbre à branchement complexe à droite.

Facteur	Score			Temps de réponse		
	Estimation (%)	Erreur Std.	p-value	Estimation (ms)	Erreur Std.	p-value
(Interception- ordonnée à l'origine)	0.76	1.20	0.52	2496	277	0.00
Syntaxe 2 vs 1	-0.40	0.26	0.11	-133	50	0.01
Syntaxe 3 vs 1	-0.11	0.27	0.67	58	47	0.21
Prosody 2 vs 1	-0.74	0.24	0.00	216	48	0.00
Contexte 2 vs 1	-0.42	0.18	0.02	-116	43	0.01
Contexte 3 vs 1	0.20	0.20	0.32	-112	40	0.01
Contexte 4 vs 1	0.29	0.30	0.34	71	66	0.29
Durée	0.48	0.34	0.16	560	79	0.00
Syntaxe 2 : Prosodie 2	0.24	0.39	0.53	105	82	0.22
Syntaxe 3 : Prosodie 2	-0.45	0.36	0.21	-95	77	0.22

TABLE 2.1 – Effets fixes dans un modèle d'analyse à effet mixte des scores et des temps de réponse. Les p-value ont été calculées en utilisant des simulations Monte carlo avec chaînes de Markov avec une taille d'échantillon de 10 000. (Gelman and Hill, 2007)

décision et sur les scores) et de durée des stimuli (sur les temps de décision) expliquent également de façon significative les données.

Résultats IRMf : Anova Classique

L'analyse de variance (anova within subject) réalisée au niveau du groupe à partir des images individuelles des 6 conditions a permis de mettre en évidence des effets dus à la structure syntaxique (carte de contraste représentant une statistique F à gauche de la figure 2.11) et des effets dus à la structure prosodique (carte de contraste représentant un test statistique t à droite de la figure 2.11) à un seuil non corrigé $p < 0.001$.

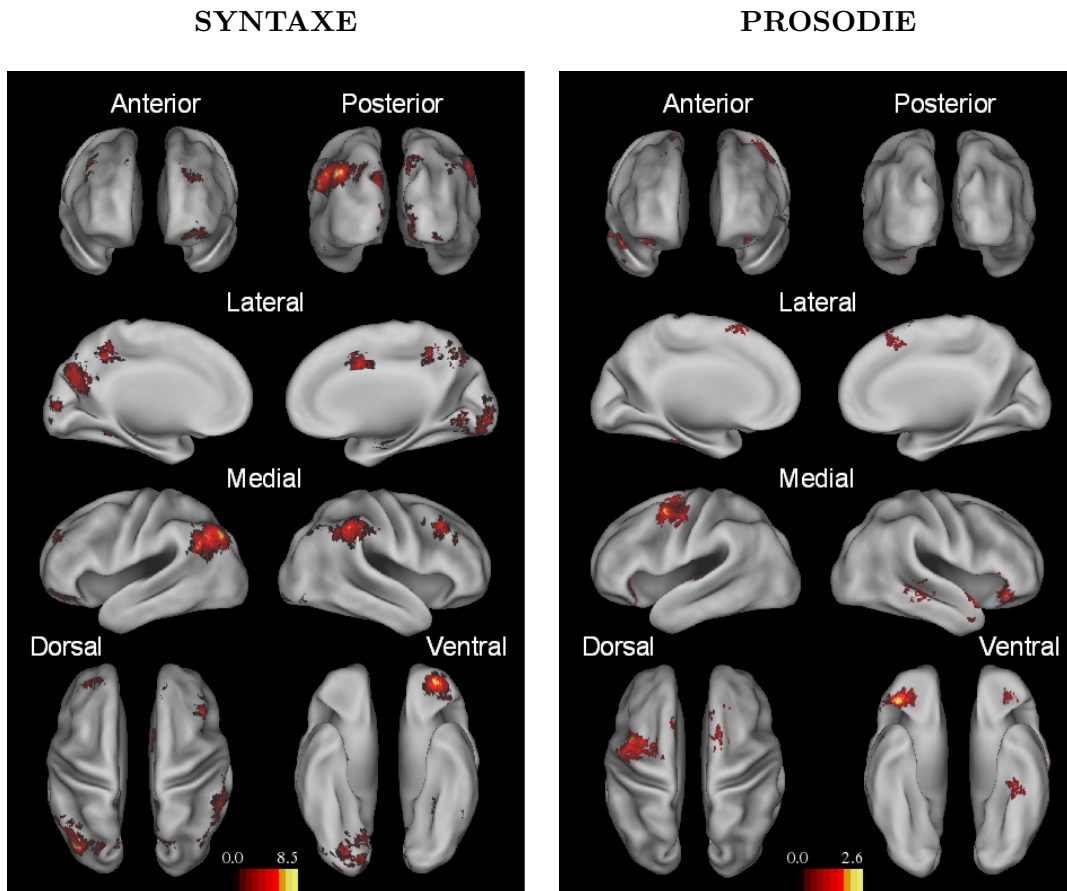


FIGURE 2.11 – Cartes statistiques de sensibilité aux structures prosodiques et syntaxiques. Seuil $p < 0.001$ non corrigé.

Le réseau mis en évidence qui réagit à la structure syntaxique, sans se préoccuper d'ordonner les 3 types de structure, contient bilatéralement le gyrus angulaire, le gyrus inférieur frontal gauche dans sa partie orbitalis, sur la vue médiale le cunéus gauche et le précunéus bilatéralement et enfin le cervelet de façon bilatérale. Le réseau révélé

par le contraste entre les 2 niveaux de hiérarchie prosodique comporte le gyrus frontal inférieur droit dans sa partie orbitalis, le gyrus précentral à gauche (BA6), une partie du cervelet gauche, le sillon temporal supérieur droit dans sa partie médiane ainsi que le pôle temporal associé, l'aire supplémentaire motrice bilatéralement et une région à cheval entre l'insula gauche et le gyrus frontal inférieur.

Les contrastes nous révèlent un réseau différent pour les 2 types de structure. Cependant la sensibilité de ce type d'analyse n'est pas très bonne ; les résultats ne sont pas suffisamment robustes pour résister à des corrections statistiques¹⁰ de type family wise error (FWE) ou false discovery rate (FDR). Il est donc pertinent de tenter d'obtenir une carte d'activation plus robuste grâce à notre nouvelle méthode de classification.

Résultats IRMf : Analyse de Classification

Plusieurs tests de classification ont été réalisés et sont illustrés par les matrices de confusion sur la figure 2.12 page suivante : prosodie 1 vs prosodie 2 (sur $6 \times 22 = 132$ images), syntaxe 1 vs 2 vs 3, syntaxe 1 vs syntaxe 2 (sur $4 \times 22 = 88$ images), syntaxe 1 vs syntaxe 3 (sur 88 images), syntaxe 2 vs syntaxe 3 (sur 88 images). L'algorithme de classification a classé correctement les images comme provenant des 3 arbres syntaxiques différents. Il y a une chance sur 3 de labelliser correctement les images, le niveau du hasard sera donc à 33%. La première matrice de confusion montre que lorsque l'image provenait de la condition arbre à droite, elle a été classée dans 50% des cas dans la bonne catégorie. La classification est donc meilleure que le hasard. De la même façon lorsque l'image était issue de la condition branchement complexe à gauche, le label correct a été appliqué dans 55% des cas et dans la condition branchement complexe à droite dans 52% des cas. Le classifieur a aussi été capable de faire la différence entre les images provenant des différentes conditions prosodiques : prosodie 1 vs prosodie 2 dans 68% des cas ($p=0.00002$ selon un test exact binomial). Les tests des types de structure deux par deux sont également significativement meilleur que le hasard : condition syntaxe

10. Il ne reste plus d'activation lorsque des corrections de type FWE ou FDR sont appliquées à ces données.

2.2. Expérience 4 : Bases neurales du traitement des structures syntaxique et prosodique 91

1 par rapport à la syntaxe 2 (69% ; $p\text{-value}=0.0002$), syntaxe 1 et syntaxe 3 (64% ; $p\text{-value}=0.007$) et entre la syntaxe 2 et 3 (66% ; $p\text{-value}=0.002$).

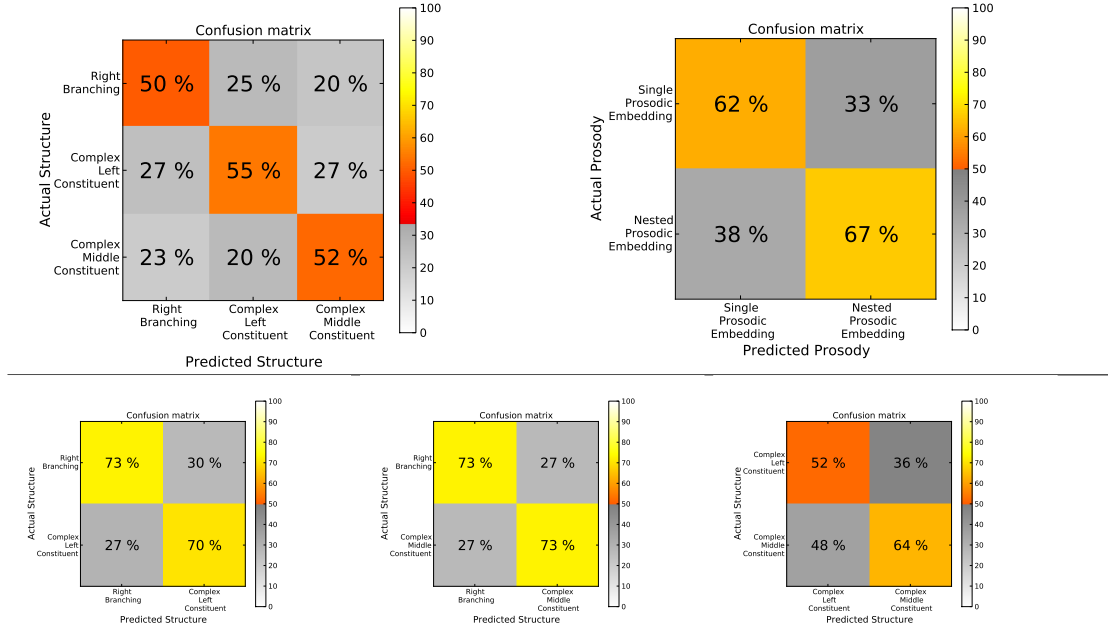


FIGURE 2.12 – Les matrices de confusion pour la classification selon les 3 niveaux de structures syntaxiques (en haut à gauche), selon les deux niveaux de prosodie (en haut à droite), et les classifications des arbres syntaxiques deux à deux (en bas)

Les classifications montrent des taux de classification supérieurs au hasard, c'est-à-dire que les images provenant des différentes conditions prosodiques et syntaxiques contiennent suffisamment d'informations pour pouvoir les classer correctement en fonction des conditions. Cette classification correcte ne nous permet pas en tant que telle de savoir quelles zones cérébrales sont impliquées dans chacune de nos conditions. Cette information de localisation est contenue dans les coefficients de chaque parcelle obtenus par l'analyse de régression logistique. La régression logistique est un classifieur linéaire c'est-à-dire que l'étiquette (la condition) est attribuée à chaque image en se basant sur une combinaison linéaire d'éléments (voxels ou parcelles). La régression logistique parcimonieuse privilégie les vecteurs de poids avec le moins d'éléments nuls possibles c'est-à-dire que seules quelques parcelles (avec des poids non nuls) vont réellement contribuer à la fonction de décision. Ce sont ces quelques parcelles avec des poids non nuls qui vont

être utilisées par l'algorithme pour assigner une étiquette (la condition dans notre cas) à une image.

Ces poids définis au niveau de la parcelle peuvent être projetés sur une image cérébrale en assignant à chaque voxel de l'image cérébrale le poids de la parcelle à laquelle il appartient. Comme tous les éléments passés en entrée de la régression logistique sont standardisés (même variance), la valeur des poids peut être interprétée comme l'importance de la parcelle dans la classification de l'image. De plus, la régression logistique parcimonieuse se concentre particulièrement sur les quelques régions cérébrales les plus prédictives (Yamashita et al., 2008; Carroll et al., 2009).

Pour chaque test de classification, chaque voxel a reçu une valeur représentant le nombre de fois où la valeur de ce voxel n'est pas nulle (de 0 à 22 nombre total de sujets). Pour la classification en fonction de la prosodie, le seuil de l'image a été défini à 0.5 (c'est-à-dire que les voxels sont importants pour la classification dans au moins 50% des sujets). Pour créer une seule image pour les 3 conditions syntaxiques, nous avons seuillé chacune des 3 images à 0.25 puis seuls les voxels apparaissant dans au moins 2 des 3 images ont été gardés.

La carte cérébrale résultant de l'activation des 3 catégories syntaxiques est présentée dans l'image A. 2.13 page ci-contre, et la carte résumant les activations relatives aux conditions prosodiques dans l'image B. 2.13 page suivante. La classification syntaxique est principalement due aux régions corticales frontales gauches médiale et inférieure, au lobe temporal supérieur de façon bilatérale, au gyrus droit supra-marginal, au précuneus et de façon bilatéral à l'hippocampe et au para-hippocampe. La classification prosodique, quant à elle, repose principalement sur les régions inférieures du gyrus frontal droit et l'insula antérieure, le lobe temporal antérieur droit, le cervelet droit, le gyrus précentral dorso-latéral gauche, le gyrus frontal inférieur gauche et le gyrus angulaire gauche.

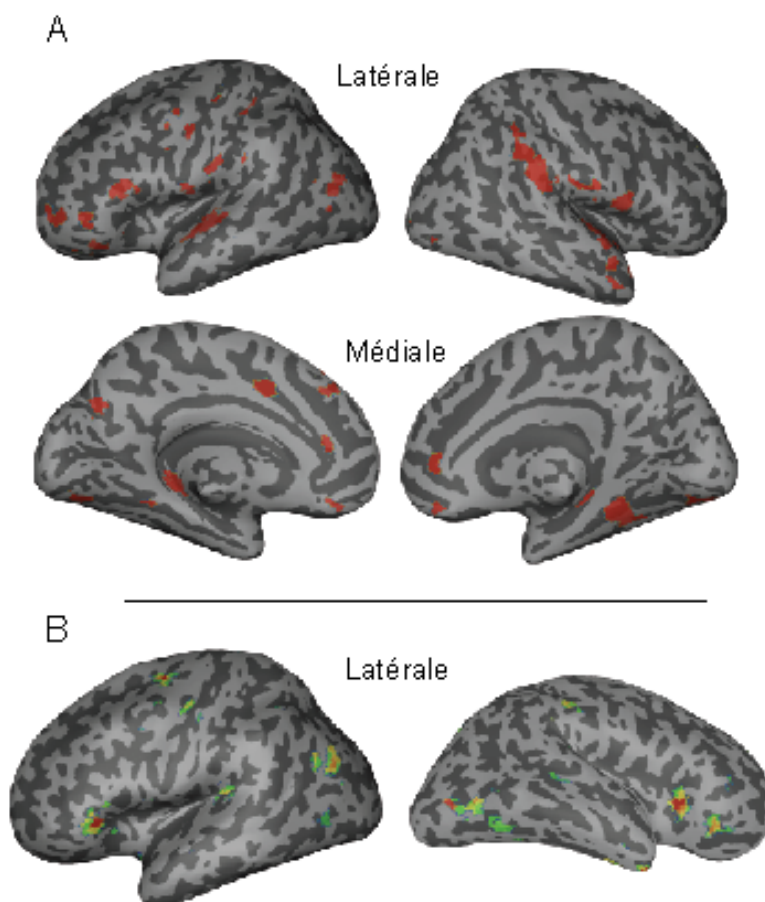


FIGURE 2.13 – Cartes résultant A - des 3 classifications syntaxiques et B - des classifications prosodiques

Discussion

Paradigme et analyse

Cette expérience utilise un nouveau paradigme linguistique basé sur la coordination en français pour tester de façon orthogonale la structure syntaxique et prosodique. Nous avons, ici, réussi à entraîner un classifieur pour labelliser les images cérébrales en fonction des différentes structures testées. La classification d'images cérébrales IRMf a déjà été utilisée pour décoder des propriétés linguistiques (propriétés sémantiques des noms) par (Mitchell et al., 2008). Cependant, notre étude est la première à utiliser cet outil de classification pour étudier des classes d'objets plus abstraites et complexes : la représentation de la structure syntaxique et prosodique de phrases. La nouveauté de

notre procédure de classification est double. Premièrement, c'est la première fois qu'une classification a lieu au niveau du groupe d'individus; habituellement ce type de classification a lieu à un niveau individuel c'est-à-dire qu'on entraîne le classifieur sur une partie des données d'un individu et qu'on teste sa performance de classification sur une autre partie des données de ce même individu. Deuxièmement, la nouveauté provient de la façon de réduire le nombre de voxels. Contrairement à la sélection habituelle des voxels les plus représentatifs qui perd une partie non négligeable de l'information notamment anatomique et fonctionnelle, nous avons créé des parcelles de voxels auxquelles nous avons assigné la moyenne des activations des voxels qui composent chacune. Grâce à cette approche basée sur un couplage entre un fragment de linguistique du français (la coordination) et l'utilisation de discrimination de pattern multivariés, nous avons pu identifier des réseaux cérébraux distincts pour les représentations de la structure syntaxique et de la structure prosodique.

Notre algorithme de classification a classé les images cérébrales selon les différents arbres syntaxiques testés : branchement à droite, branchement complexe à gauche et branchement complexe à droite. Il a également discriminé les images provenant de 2 structures prosodiques distinctes : un type de frontière prosodique vs 2 types de frontières enchâssées. Les cartes cérébrales obtenues grâce aux coefficients du classifieur pour chaque voxel, montre un ensemble de régions distinctes pour les 2 types de structure testée : syntaxique et prosodique. Il est important de souligner que les structures prosodiques mais aussi syntaxiques reposent toutes les 2 sur des indices prosodiques. Il est alors d'autant plus remarquable de mettre en évidence des réseaux qualitativement distincts pour les structures prosodique et syntaxique. Ces réseaux encodent donc des représentations internes abstraites créées par les sujets et non des propriétés inhérentes aux stimuli.

Structure syntaxique

La classification et l'analyse univariée concordent sur l'implication du gyrus angulaire droit, du gyrus frontal inférieur gauche dans sa partie orbitalis (BA47) et du précuneus gauche dans le traitement différentiel des différentes structures syntaxiques. La classification engage également le sillon temporal supérieur gauche et la région BA45 ou gyrus frontal inférieur gauche, zones qui sont assez souvent retrouvés dans la littérature (par exemple (Santi and Grodzinsky, 2010; Pallier et al., 2011)).

Un certain nombre des régions identifiées dans cette expérience comme faisant la distinction entre différents types d'arbres syntaxiques correspondent à des régions révélées par d'autres expériences de neuroimagerie, en particulier la région postérieure dorsale de l'aire de Broca (Santi and Grodzinsky, 2010) et la partie antérieure du lobe temporal supérieur (Pallier et al., 2011). Même si le précuneus n'est pas habituellement considéré comme jouant un rôle dans le traitement des phrases ou de la syntaxe, un certain nombre d'expériences d'imagerie (Caplan et al., 2003; Xu et al., 2005; Shetreet et al., 2009) et une étude lésionnel portant sur un grand nombre de patient (Caplan et al., 2007) ont montré que cette région pourrait être impliquée dans le traitement syntaxique.

Nos résultats sont donc en accord avec la littérature. Cependant, à la différence de celle-ci notre paradigme nous permet de tester spécifiquement et uniquement la structure syntaxique, ce qui pourrait expliquer les différences avec d'autres expériences en imagerie (Santi and Grodzinsky, 2007; Pallier et al., 2011) qui impliquent le pôle temporal gauche et la partie antérieure de l'aire de Broca (BA45 et 47) dans le traitement de la structure syntaxique. Nous émettons ici l'hypothèse que ces 2 régions, que nous ne mettons pas en lumière dans notre expérience, ne sont pas responsables du traitement syntaxique proprement dit, mais que leur activité reflète plutôt d'autres facteurs confondus avec la syntaxe dans la plupart des expériences. Par exemple, (Thompson-Schill et al., 1997) propose que l'aire de Broca serait associée à un processus de sélection d'informations qui serait entre autre utilisé par le langage et qui varie avec les conditions syntaxiques dans les expériences sus-cités. (Hale, 2001) montre qu'un autre facteur varie classiquement

avec la complexité syntaxique : la surprise.

Structure prosodique

Les deux analyses impliquent un réseau de régions de l'hémisphère droit : le gyrus inférieur frontal dans sa partie orbitalis (BA47), le pôle temporal, le sillon temporal supérieur et des régions du cervelet, et de l'hémisphère gauche : le gyrus précentral, l'insula, le gyrus frontal inférieur orbitalis (BA47) dans le traitement de la structure prosodique.

Dans la littérature, on retrouve l'implication de l'hémisphère droit dans le traitement de la structure prosodique (Weintraub et al., 1981). Cependant, on manque de données claires en neuroimagerie c'est-à-dire des données dans lesquelles aucun facteur ne serait confondu avec la structure prosodique. L'implication de l'hémisphère droit de l'aire de Broca (BA47) et du pôle temporal droit suggère une analogie intrigante, même si visiblement simpliste, entre les représentations de la structure abstraite (syntaxe à gauche et prosodie à droite). Le rôle du cervelet dans la production de la prosodie est bien documenté (Ackermann and Hertrich, 2000). En utilisant la TEP (topographie par émission de positron) Strelnikov et ses collègues (Strelnikov et al., 2006) ont montré que le cervelet droit était particulièrement activé pour des phrases segmentées prosodiquement en comparaison de phrases non segmentées, malgré le fait que leur dessin expérimental crée de nombreuses différences à travers les conditions (lexicale, syntaxique, sémantique et tâche logique). L'aire de Brodmann 6 à gauche a été identifiée par (Hesling et al., 2005) en comparant les réponses cérébrales à de la parole utilisant une prosodie expressive par rapport à de la parole avec une prosodie plate.

L'expérience (Ischebeck et al., 2008), présentée en introduction de ce chapitre, manipulait le nombre de frontières de groupe intonational. Cependant, la manipulation de la structure prosodique covariait avec la structure syntaxique contrairement à nos stimuli dans lesquelles les structures prosodique et syntaxique varient de façon orthogonale. Une seconde différence est liée au niveau d'abstraction. Dans l'expérience d'Ischebeck,

il n'existe qu'un seul type de frontière prosodique présent en un ou deux exemplaire mais sa qualité ne varie pas. Le paramètre acoustique relatif à la frontière prosodique est constant, contrairement à notre expérience. Nos stimuli varient selon un niveau de représentation abstraite de la structure prosodique. Les propriétés acoustiques sont différentes pour les 2 types de frontières (une plus grosse que l'autre). Enfin, (Ischebeck et al., 2008) ont utilisé la logique de soustraction univariée standard pour analyser leurs données c'est-à-dire qu'ils ont contrasté les activations de 2 conditions prosodiques. Grâce ce procédé, ils n'ont pu mettre en évidence des effets prosodiques que dans le cortex auditif. Ces activations pourraient être liées à des différences acoustiques de bas niveau entre les différentes conditions. Contrairement aux résultats obtenus dans notre expérience, ils ne trouvent aucun effet prosodique dans d'autres régions cérébrales.

Plusieurs chercheurs suggèrent que le gyrus angulaire (ou région supramarginale) ferait parti du réseau de traitement de la parole (Tyler and Marslen-Wilson, 2008; Lau et al., 2008). Nos résultats suggèrent que cette région ferait correspondre la structure syntaxique aux domaines prosodiques. Caplan et ses collègues (Caplan et al., 2002) ont montré que le gyrus angulaire est sensible à la complexité syntaxique. Cependant comme la manipulation de la structure syntaxique comporte souvent des changements dans la structure prosodique, il est possible de réinterpréter ces résultats (et d'autres similaires dans la littérature) en terme de complexité prosodique. Notre expérience est la première à séparer les deux dimensions tout en gardant des stimuli écologiques et ainsi nous avons pu distinguer des régions purement syntaxiques et des régions qui permettent de faire correspondre les structures syntaxique et prosodique.

Pour conclure, le réseau de traitement de la structure syntaxique que nous observons, chevauche mais n'est pas identique aux activations classiquement observées. Plusieurs raisons peuvent expliquer ce résultat : des différences de paradigmes linguistiques employés, des différences dans l'approche de traitement de données (la classification) ou une combinaison de ces 2 facteurs. Concernant les régions sensibles à la structure prosodique, les données de la littérature sont peu nombreuses et nous mettons donc en

évidence un nouveau réseau qui confirme cependant le tableau général émergeant de la neuropsychologie et des quelques études de neuroimagerie.

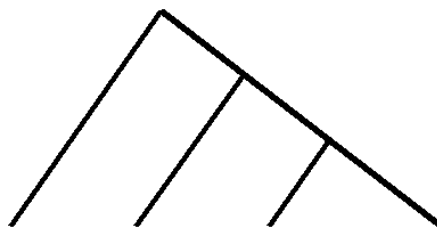
Résumé du chapitre

1. Nous avons utilisé un nouveau paradigme linguistique basé sur la coordination qui présente trois intérêts :

- il force à calculer le sens des phrases ;
- les structures prosodiques et syntaxiques varient de façon orthogonale
- la syntaxe diffère dans les conditions indépendamment des facteurs avec lesquels elle est habituellement confondue (sémantique, mémoire de travail, fréquence des mots ...).

La figure ci-dessous montre un exemple de structures syntaxique et prosodique utilisées.

Arbre à branchement à droite



rouge ou (vert et jaune et bleu)

rouge ou (vert et (jaune ou bleu))

2. Une nouvelle approche d'analyse des données fonctionnelles a été mise au point : la classification. Cette méthode présente les caractéristiques suivantes :

- elle classe les images fonctionnelles en fonction des différentes structures syntaxiques (3 arbres différents différenciés par des indices prosodiques) ;
- elle classe les images fonctionnelles en fonction des différentes structures prosodiques (2 niveaux de profondeurs prosodiques : un type de frontière vs. 2 types de frontières différentes) ;

- elle décode les pattern cérébraux au niveau du groupe ;
 - elle utilise une réduction du nombre de voxels grâce à des parcelles contenant chacune la valeur moyenne d’activation des voxels qui la composent.
3. Le décodage des données fonctionnelles révèle des réseaux cérébraux différents pour les structures syntaxiques et les structures prosodiques.
4. Une analyse classique de soustraction univariée a été menée en parallèle de la nouvelle approche de classification des images. Les résultats de ces deux analyses sont les suivants :
- Réseau Syntaxique : Le gyrus frontal inférieur orbitalis (BA47) gauche, gyrus angulaire droit, précunéus gauche sont les régions mises en évidence dans les deux analyses. Le sillon temporal supérieur antérieur gauche et l’IFG triangularis (BA45) gauche sont révélés dans l’analyse de classification.
 - Réseau Prosodique : Plusieurs régions cérébrales sont communes aux deux analyses : dans l’hémisphère droit le gyrus frontal inférieur orbitalis (BA47), le pôle temporal, le sillon temporal supérieur médian et des régions du cervelet ; dans l’hémisphère gauche : le gyrus précentral, l’insula et le gyrus frontal inférieur orbitalis (BA47).

Structures hiérarchiques en langage et en musique

Sommaire

3.1	Un peu de théorie	101
3.2	Les données de la littérature	107
3.3	La théorie de partage des ressources syntaxiques	108
3.4	Étudier la syntaxe en langage	113

3.1 Un peu de théorie

La combinaison (merge) d'éléments selon des règles récursives pour construire un sens plus élaboré ne semble pas être l'apanage du langage seul. On trouve aussi des structures récursives dans les domaines de la planification de l'action, des mathématiques ou encore de la musique.

Par exemple, pour faire du café avec une cafetière classique à filtre, il faut effectuer une suite de gestes organisés hiérarchiquement (cf. figure 3.1).

De la même manière, les expressions mathématiques et les morceaux de musique peuvent être représentés sous forme de structures hiérarchiques. Depuis l'apparition des théories génératives linguistiques dans les années 1960 (Chomsky, 1957), certains musiciens ont cherché à décrire la musique et sa structure en utilisant les outils de la linguistique générative. Les premiers à formaliser une théorie générative de la musique

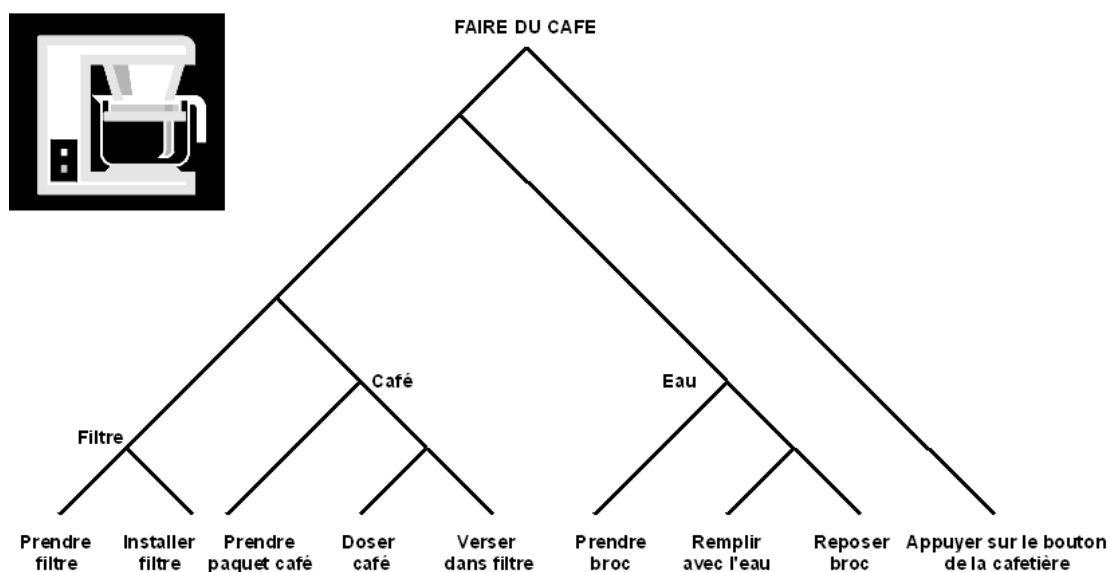


FIGURE 3.1 – Organisation hiérarchique des actions requises pour faire du café

occidentale tonale ont été le compositeur Fred Lerdahl et le linguiste Ray Jackendoff (Lerdahl and Jackendoff, 1983). Leur ouvrage fondateur définit les unités structurales de la musique et fournit une première description de leur combinaison suivant différentes règles : en somme une première grammaire de la musique. Les unités de base sont des notes ou des accords (en langage ce sont les phonèmes). Quatre types de structures sont définies et décrivent l'organisation d'un morceau de musique :

- la **structure de groupe (grouping structure)** est le fait d'organiser les éléments sous forme de séquences plus ou moins longues, de grouper des éléments sous forme de phrase par exemple. Elle est représentée par la figure 3.2 page ci-contre.



FIGURE 3.2 – Exemple de représentation de la structure de groupe du début du scherzo de la sonate de Beethoven opus2 n°2. Les accolades représentent les différents groupes d'éléments et leur hiérarchie. Extrait de (Lerdahl and Jackendoff, 1983) p15.

- la **Structure métrique** est le fait qu'il existe une pulsation ainsi que des temps forts et des temps faibles. Elle est figurée par l'image 3.3.

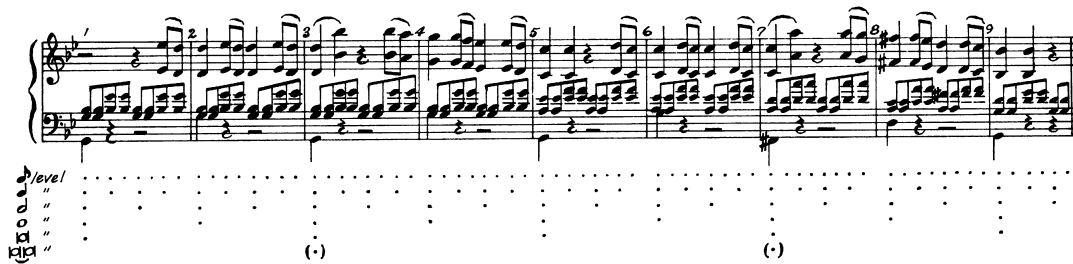


FIGURE 3.3 – Exemple de représentation de la structure métrique du début de la Symphonie en sol mineur de Mozart. Le nombre de points représente l'importance du temps dans la mesure. Les premiers temps de chaque mesure sont des temps forts (au moins 4 points dans la représentation de la structure métrique), en comparaison le 2ème accord croche sol-si à la clé de fa, n'est pas très important. Extrait de (Lerdahl and Jackendoff, 1983) p23.

Ces deux premières structures peuvent être combinées selon la figure 3.4 page suivante.

- **Réduction sur la durée (Time-Span Reduction)** : le fait qu'on puisse simplifier la séquence d'élément tout en conservant le même sens global, la même analyse. Il est possible de créer un arbre représentant les relations des notes ou accords les uns par rapport aux autres en tenant compte de leur hiérarchie, majoritairement harmonique mais aussi temporelle. Plus une branche est longue, plus l'élément est important. Une telle réduction est illustrée par la figure 3.5 page 105.

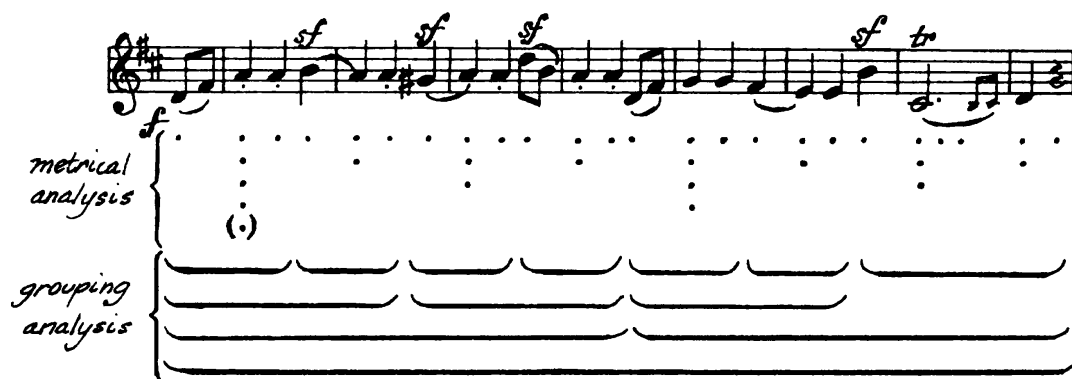


FIGURE 3.4 – Représentation des structures de groupe et métrique de l'Ouverture du Menuet de la Symphonie n° 104 de Haydn. Extrait de (Lerdahl and Jackendoff, 1983) p26. Les 2 types de structures sont plus ou moins corrélées, dans le cas présent, elles sont assez différentes puisque les frontières entre les groupes et les temps fort ne coïncident pas.

- **Réduction prolongationnelle (Prolongational Reduction)** exprime les relations de tension et de relaxation entre les éléments musicaux. La figure 3.6 page 106 symbolise ce type de réduction.

Il n'existe pas de référent externe associé aux éléments musicaux comme il existe un sens pour les mots. Il n'existe donc pas de condition de vérité associée à une phrase en musique ; l'ensemble des structures vont s'organiser différemment selon les morceaux. Ainsi, parmi ces quatre structures, il n'en existe pas une qui soit toujours prédominante ou un même ordre de priorité pour leur application qui définirait un morceau bien construit. La relation qu'entretiennent ces règles les unes par rapport aux autres dépend du morceau et de l'auditeur qui pourra favoriser une analyse particulière de la pièce. Même si tous les auditeurs d'un morceau de musique ne l'interprètent pas de la même façon, le plus souvent il existe une analyse partagée par un grand nombre d'auditeurs qui correspond à une certaine pondération des règles les unes par rapport aux autres et qui conduit à une structure hiérarchique particulière. Une autre différence importante avec le langage provient de l'existence de catégories syntaxiques (nom, verbe, adjectif etc.), qu'on ne retrouve pas en musique.

Malgré ces différences importantes et à mesure que la recherche avance en linguis-

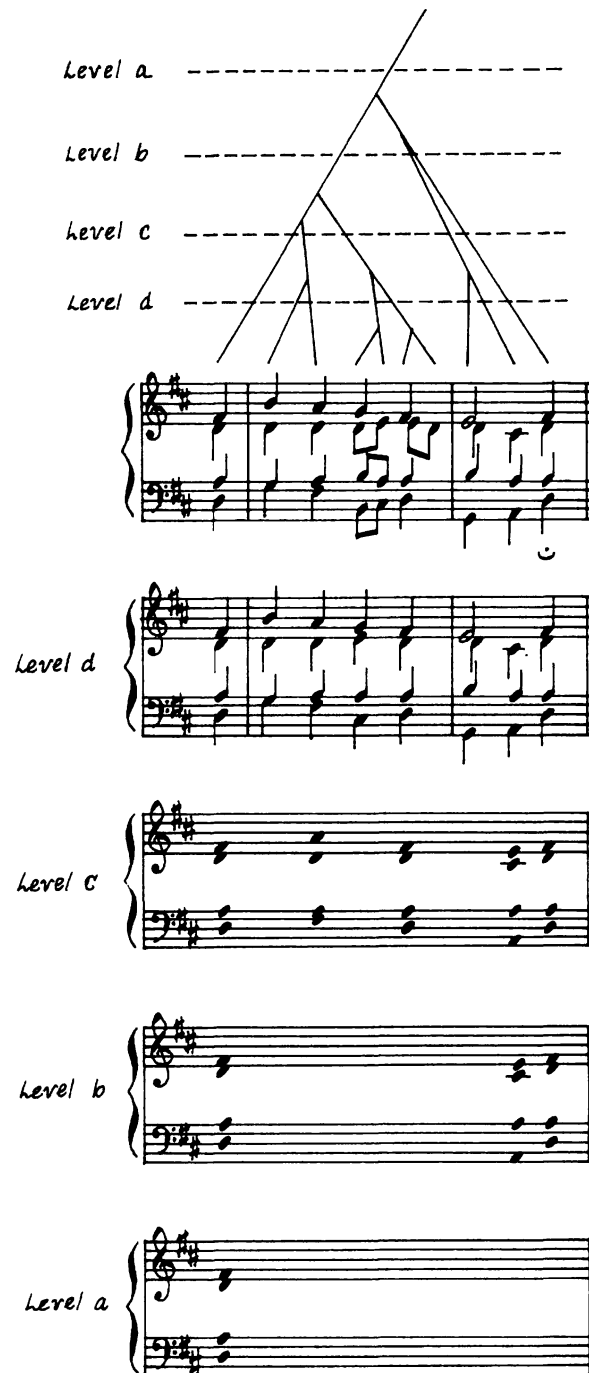


FIGURE 3.5 – Représentation de la time-span reduction de la première phrase de la chorale de Bach 'O Haupt von Blut und Wunden'. Extrait de (Lerdahl and Jackendoff, 1983) p115. Les différents niveaux de structure a, b, c et d correspondent à des simplifications de plus en plus importantes. Les notes les moins importantes peuvent être supprimées de façon hiérarchique.



FIGURE 3.6 – Représentation de la prolongational reduction du début de 'La ci darem la mano' de Don Giovanni de Mozart. Extrait de (Lerdahl and Jackendoff, 1983) p200. Les branchements à droite représentent les éléments de tension et les branchements à gauche les éléments de relaxation.

tique, les chercheurs trouvent de plus en plus de similitudes dans l'organisation structurale de ces deux domaines. Très récemment, deux linguistes ont même proposé la « thèse de l'identité » (Katz and Pesetsky, *tted*). Selon cette thèse, il y aurait identité entre les descriptions syntaxiques du langage et de la musique. Dans leur article, ils appliquent les règles de la syntaxe linguistique aux unités de base de la musique pour en décrire la structure. Ils soutiennent que « Toutes les différences formelles entre le langage et la musique sont la conséquence des différences dans les unités structurales (building blocks) à savoir des paires arbitraires entre son et sens pour le langage et des paires entre classes de hauteurs de notes (pitch) et leur combinaisons pour la musique. Autrement, le langage et la musique sont identiques en tous points ». Dans la même optique, Martin Rohrmeier (Rohrmeier, 2011) a développé une grammaire de type « phrase structure » pour décrire la musique tonale. Toutes ces descriptions formelles semblables pour le langage et la musique permettent de supposer qu'il existerait un système computationnel commun pour traiter les structures hiérarhiques. Cette hypothèse, aujourd'hui supportée par plusieurs théories, n'a cependant pas été la première historiquement émise.

3.2 Les données de la littérature

Depuis une vingtaine d'année des données de neuropsychologie sont présentées en faveur d'une dissociation des deux domaines. En effet, il existe des personnes amusiques qui présentent un déficit spécifique à la musique. Isabelle Peretz et d'autres chercheurs étudient cette population particulière (Peretz, 1993; Peretz et al., 1994). Ce trouble peut être acquis en cas d'une lésion ou congénital (Peretz and Coltheart, 2003), et peut être totalement spécifique à la musique sans aucun trouble linguistique. Par exemple, il existe des amusiques qui présentent un trouble spécifique de la reconnaissance de mélodie sans parole. Ces personnes après leur lésion deviennent incapables de reconnaître une mélodie qui leur était pourtant familière auparavant ; mais ils reconnaissent parfaitement les paroles d'une chanson parlée, les voix familières et les sons de l'environnement (bruits d'animaux, bruits de la ville ou sons vocaux humains). A l'appui de cette dissociation entre langage et musique, il existe à l'inverse des aphasiques qui présentent des troubles spécifiques de la parole sans aucun trouble de perception musicale. Ces données neuropsychologiques ont suggéré l'existence de modules cérébraux distincts pour le langage et la musique.

Cependant, les données obtenues en neuroimagerie sont venues enrichir le débat. Mireille Besson et ses collègues (Besson and Faïta, 1995) ont enregistré à l'aide d'un électroencéphalogramme une réponse tardive positive spécifique environ 600ms après qu'une note déviante c'est-à-dire hors de la mélodie ou de la tonalité, apparaisse à la fin d'une courte mélodie. Ce potentiel cérébral évoqué par la note déviante ressemble à celle provoquée par un mot déviant dans une phrase. Patel (Patel et al., 1998) confirme dans un dessin expérimental intrasujet que cette onde P600 est la même réponse à une incongruité langagière ou musicale¹. De plus, une expérience MEG présentant des séquences d'accords contenant ou non des déviants harmoniques montre l'implication dans la génération d'un composant ERAN (early right anterior negativity) de l'aire de Broca gauche

1. Incongruité = note dans une tonalité distante ou phrase agrammaticale en langage, la complexité syntaxique plus grande = accord dans une tonalité proche et phrase plus complexe avec une ambiguïté locale. La comparaison statistique entre les P600 tient entre la phrase normale et l'agrammaticale et aussi entre la normale et la plus complexe

et son homologue droit, zones présumées être le siège des calculs syntaxiques en langage (Maess et al., 2001). Toujours dans la même veine, plusieurs auteurs (Koelsch et al., 2002; Levitin and Menon, 2003; Tillmann et al., 2003; Schmithorst, 2005) montrent en IRMf l'implication dans le traitement de musique plus ou moins bien construite d'un réseau semblable à celui évoqué par des stimuli langagiers sans toutefois jamais proposer de comparaison directe chez les mêmes sujets entre ces deux domaines². Une expérience en PET de production dans les deux domaines a été réalisée en 2006 (Brown et al., 2006). Les sujets entendaient soit des débuts de phrases en langage soit des débuts de mélodies inconnues et ils devaient inventer la fin de la phrase ou de la mélodie (fredonnée sur la syllabe da). Un certain nombre de régions communes à la production de langage et de musique ont été ainsi mises en évidence, même si, d'autres zones répondaient spécifiquement à chaque domaine. Toutes ces données de neuroimagerie tendent vers l'existence de régions cérébrales communes au traitement de la musique et du langage même si on ne comprend pas encore bien dans quelle mesure.

3.3 La théorie de partage des ressources syntaxiques

Les données de neuropsychologie et de neuroimagerie apparaissent comme contradictoires. La première promeut des régions ou modules cérébraux différents pour le langage et la musique tandis que les observations anatomiques et fonctionnelles nous laissent supposer un partage de ressources. Il n'est cependant pas possible d'exclure l'existence de réseaux fonctionnels séparés pour le langage et la musique mais qui colocaliseraient dans les mêmes régions corticales. La théorie d'Aniruddh Patel de partage de ressources syntaxiques (shared syntactic integration resource hypothesis ou SSIRH) permet de réconcilier ces 2 types d'observations. Patel (Patel, 2003) propose qu'il existerait des représentations spécifiques à chacun des domaines, qui en étant lésés séparément provoqueraient des troubles spécifiques et expliquerait les données neuropsychologiques. Ces représentations propres au domaine seraient ensuite intégrées pour faire sens grâce à un

2. jusqu'à 2011, l'expérience (Abrams et al., 2011) sera détaillée plus loin 5.2 page 199

réseau partagé, au moins en partie. Son atteinte expliquerait les observations de neuroimagerie. Les représentations spécifiques au langage correspondent aux représentations phonologique, sémantique et syntaxique associées aux mots (la forme sonore, le sens et la catégorie syntaxique au moins). Celles de la musique correspondent aux notes ou accords et leurs relations harmoniques. Ces représentations spécifiques impliquent un stockage dans la mémoire à long terme (Tillmann et al., 2000) qui, s'il est lésé, provoque alors un déficit sélectif dans un des domaines. Ces représentations seraient ensuite combinées grâce un module syntaxique distinct qui pourrait être, au moins en partie, commun aux deux domaines (cf schéma 3.7).

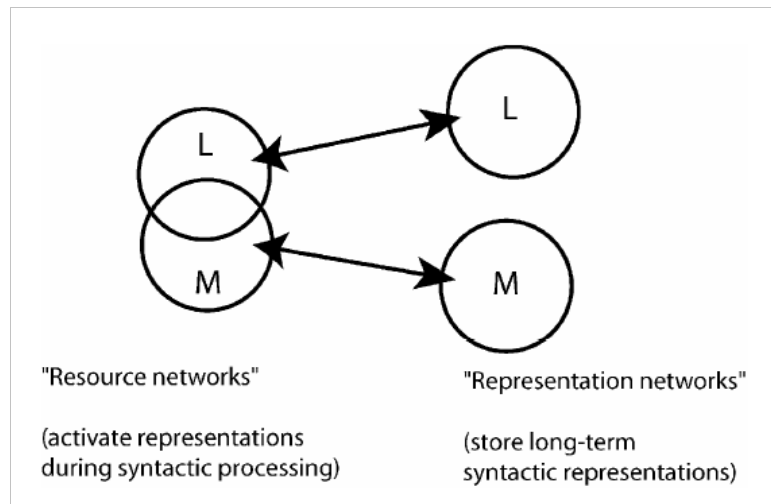


FIGURE 3.7 – Schéma des relations fonctionnelles entre le traitement syntaxique de la musique et du langage. L = langage, M= musique. Les flèches représente des connections fonctionnelles entre les différents réseaux. Il est à noter que les cercles n'implique pas nécessairement de aires cérébrales focales. Par exemple les réseaux des représentations linguistiques ou musicales peuvent s'étendre sur un certain nombre de régions cérébrales ou correspondre à des réseaux fonctionnellement séparés au sein de la même région cérébrale. Extrait de (Patel, 2008) p283.

Le modèle SSIRH émet des prédictions testables. Si les deux domaines partagent effectivement des ressources neuronales, alors augmenter la charge computationnelle dans un domaine devrait interagir et ralentir le traitement simultané dans l'autre domaine. Plusieurs expériences confirment cette hypothèse. Ainsi, Stefan Koelsch et ses collègues (Koelsch et al., 2005) ont présenté visuellement, mot à mot, de courtes phrases pendant

qu'ils enregistraient l'électroencéphalogramme des sujets. A chaque mot était associé un accord musical. Pour augmenter la charge de traitement syntaxique, la stratégie adoptée était que le dernier mot de la phrase pouvait être syntaxiquement incorrect (incorrect en genre), avec le dernier accord lui aussi potentiellement déviant (dans une tonalité distante), voir la gauche de la figure 3.8 page ci-contre pour une représentation visuelle des stimuli. Les sujets ne devaient prêter attention qu'au langage et répondre si la phrase présentée était correcte ou non. Conformément aux attentes, la violation d'attente grammaticale en langage a produit une onde LAN (late anterior negativity) et la violation d'attente grammaticale en musique a évoqué une ERAN (early right anterior negativity) lorsque chacune de ces violations était présentée seule. En revanche, lorsque les phrases musicale et linguistique se terminaient toutes deux par une violation grammaticale, les potentiels évoqués à la surface du scalp n'étaient pas additifs, c'est-à-dire que la réponse à la déviance linguistique était plus faible si elle était présentée simultanément avec un accord déviant que sur un accord cohérent (droite de la figure 3.8). Les auteurs interprètent ces données comme étant le résultat de la compétition des deux stimuli pour les mêmes ressources neurales. Cet effet ne peut pas être expliqué par un effet attentionnel général. En effet, une expérience contrôle de type oddball présentait sur chacun des mots une même note et la dernière note sur le dernier mot, potentiellement déviant, pouvait aussi être déviante. La note déviante dans ce type de paradigme est connue pour évoquer une onde de type MMN (mismatch negativity). La LAN évoquée par le mot déviant en genre n'est pas différente selon que la dernière note soit déviante ou non dans ce paradigme. La réponse à la déviance syntaxique en langage est donc modulée spécifiquement par la déviance harmonique en musique.

Une deuxième étude comportementale (Fedorenko et al., 2009) a utilisé le chant pour présenter simultanément langage et musique. Les chercheurs ont ici choisi de ne pas utiliser de violation syntaxique et ont utilisé des phrases plus ou moins complexes (contenant des propositions relatives sujet et relatives objet) pour augmenter la charge de traitement syntaxique des phrases langagières. Chaque mot était chanté sur une

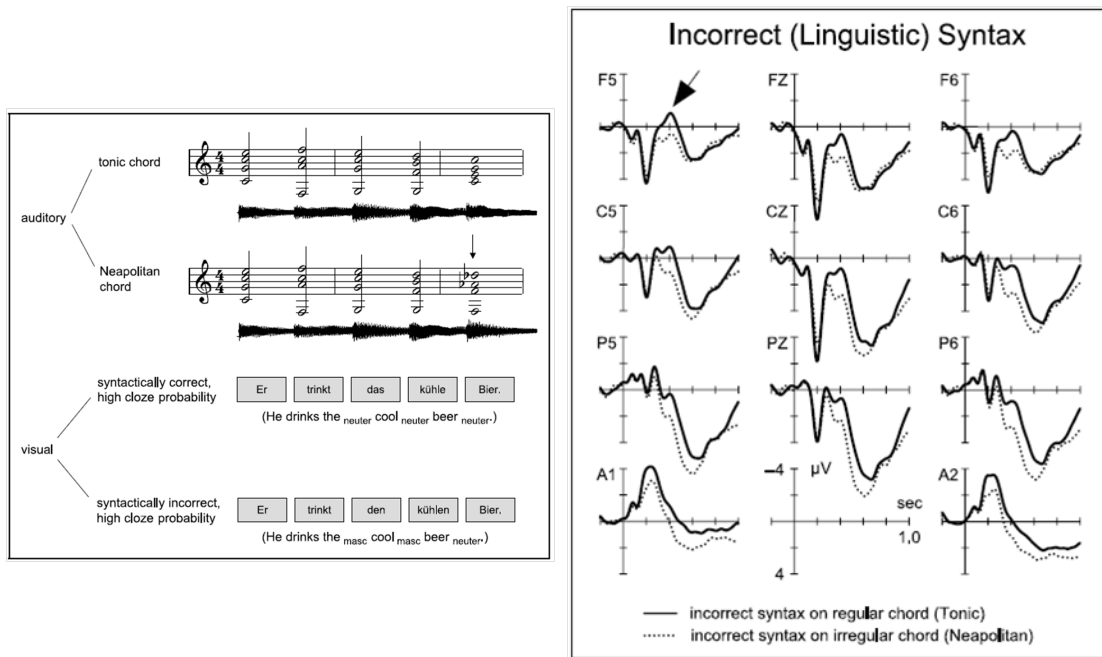


FIGURE 3.8 – Stimuli et résultats de l'expérience de (Koelsch et al., 2005). Une phrase en allemand est présentée visuellement mot à mot. A chaque mot est associé un accord. Le dernier mot de la phrase peut être incorrect syntaxiquement, le dernier accord peut lui aussi être grammatical ou non. A droite, les potentiels évoqués par la violation syntaxique en langage avec et sans violation musicale. La réponse à la violation en langage est plus faible lorsqu'elle est présentée simultanément avec un accord inattendu.

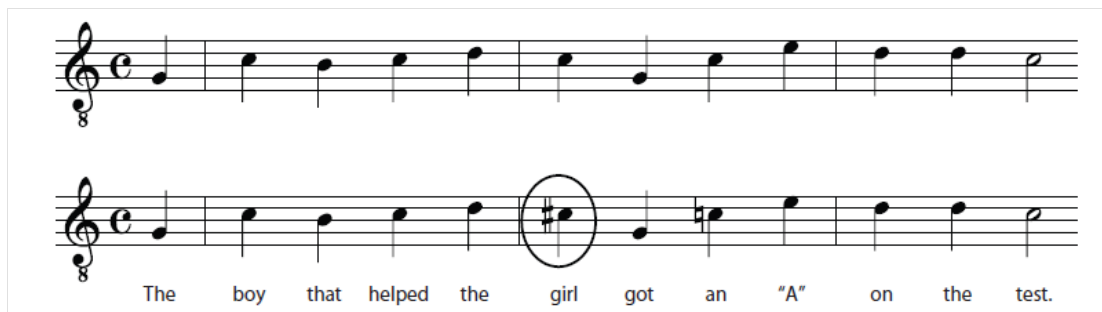


FIGURE 3.9 – Exemple de stimuli utilisé dans l'expérience (Fedorenko et al., 2009)

note. Sur le mot critique³, la note pouvait être en accord avec le début de la mélodie ou être complètement fausse, c'est-à-dire hors de la tonalité installée par le début du morceau. Ici aussi, un contrôle attentionnel avait été ajouté ; le mot critique pouvait être présenté avec un volume plus ou moins fort. A la fin de la phrase, une question testant

3. celui au niveau duquel la charge computationnelle pouvait augmenter

la compréhension était posée et la performance de réponse à cette question mesurait la charge computationnelle. La compréhension des phrases était toujours plus faible pour les phrases plus complexes c'est-à-dire celle contenant une relative-objet par rapport aux phrases relative-sujet. La différence de performance entre ces deux types de phrases était plus grande lorsque le mot était chanté sur une fausse note, mais pas s'il y avait une différence de volume sonore. Les auteurs de cette étude concluent donc que les deux domaines interagissent.

Une troisième étude de lecture progressive (*self paced reading*) a utilisé des phrases localement ambiguës (*gardenpath*) pour augmenter la charge computationnelle⁴. Un mot (ou 2 selon les cas) apparaissait à l'écran. Le sujet devait appuyer sur un bouton pour voir le mot suivant. Cette méthode permet d'enregistrer les temps de réaction associés à chaque appui du bouton et ainsi d'obtenir une approximation du temps de lecture de chaque partie de la phrase. A chaque mot (ou groupe de 2 mots) était associé la présentation auditive d'un accord. Sur le mot critique, là où une réanalyse du sens devait avoir lieu, l'accord pouvait être soit dans la tonalité, soit déviant (hors de la tonalité). Là aussi, un contrôle attentionnel avait été ajouté ; l'accord pouvait être joué par un instrument différent. Les chercheurs montrent que le temps de lecture du mot critique est plus important lorsqu'il est présenté simultanément avec un accord déviant. De plus, aucune différence de temps de lecture n'est observée si l'accord est dans la tonalité mais joué par un autre instrument. Les auteurs expliquent cette différence de temps de réaction comme une interaction des calculs syntaxiques langagiers et musicaux. Il existe donc des preuves comportementales et en neuroimagerie du partage de ressources par les deux domaines.

4. Les phrases de type *gardenpath* sont des phrases contenant une ambiguïté locale qui est résolue par la fin de la phrase. Par exemple, « Le garçon regarde la fille avec des lunettes parce qu'il est myope. ». Dans cette phrase le complément avec des lunettes peut aussi bien se rapporter à la fille qu'au garçon, il faut attendre la fin de la phrase pour que le sens soit clair, les lunettes appartiennent au garçon et non à la fille regardée.

3.4 Étudier la syntaxe en langage

Les 3 études présentées ci-dessus donnent un bon aperçu des différents paradigmes utilisés pour étudier le traitement de la syntaxe en langage à savoir : violations syntaxiques, complexité des relatives objet ou sujet, phrases gardenpath. Cependant, en cas de violation syntaxique, il est possible que le cerveau traite différemment ces stimuli détériorés par rapport à des phrases correctes, traitées, elles, automatiquement. L'utilisation de phrases contenant des relatives sujet par rapport à des relatives objet permet d'utiliser des stimuli naturels et donc de répondre au problème soulevé juste avant. Cependant, il est important de souligner que ces deux types de phrases permettent d'étudier un phénomène linguistique particulier : le mouvement. De plus, la charge en mémoire dans le cas des phrases avec relative objet est plus élevée que dans le cas de relative sujet puisque l'objet doit être conservé en mémoire jusqu'à l'apparition du sujet et du verbe tandis que dans le cas de relative sujet, le sujet est intégré au moment où le verbe apparaît. Les phrases de type gardenpath, elles aussi permettent d'utiliser des stimuli naturels sans introduire de potentiels artéfacts dus à une violation. Cependant, dans ce type de phrase, l'étude porte sur le mécanisme de réanalyse syntaxique puisqu'un premier sens a été assigné, une première structure syntaxique calculée et qu'un certain nombre d'indices nouveaux (syntaxiques, prosodiques) vont à l'encontre de ce sens et impliquent une réanalyse. En comparant (soustrayant les activités cérébrales évoquées par) une phrase sans gardenpath avec une phrase gardenpath, c'est le siège du mécanisme de réanalyse de sens qui va être révélé. Cette réanalyse pourrait utiliser les mêmes réseaux neuronaux que l'analyse initiale ou non. Elle pourrait nécessiter plus de mémoire de travail par exemple pour récupérer le sens des mots précédemment rencontrés. C'est donc bien un mécanisme syntaxique particulier qui est étudié dans ce cas précis également.

Ces phénomènes linguistiques, aussi intéressants qu'ils soient, ne permettent de mener l'investigation que sur un nombre restreint de phrases et de propriétés linguistiques (mouvement pour les phrases relative sujet et relative objet, réanalyse pour les phrases

gardenpath). Or, le but de notre travail est d'étudier les calculs syntaxiques et leur représentation cérébrale dans le cas le plus simple, c'est à dire dans le traitement automatique des phrases et particulièrement le mécanisme d'intégration d'élément (merge) édicté par Chomsky et qui permet d'intégrer de façon successive les éléments pour former une structure syntaxique. Pour étudier ce phénomène, nous nous sommes basés sur un paradigme développé dans une thèse précédente (Devauchelle, 2008) et (Pallier et al., 2011). Anne Dominique Devauchelle a, en effet, mis au point une méthode pour étudier de façon paramétrique l'intégration d'éléments dans une structure syntaxique. Elle a présenté visuellement 12 mots de façon sérielle et rapide (RSVP : rapid serial visual presentation). Ces 12 mots pouvaient soit former une phrase entière, soit des constituants syntaxiques de taille variable. Les 12 mots pouvaient former 2 constituants de 6 mots, 3 constituants de 4 mots, 4 constituants de 3 mots, 6 constituants de 2 mots ou 12 constituants de 1 mot comme sur la figure 3.10. Les stimuli étaient construits de la

Taille des constituants	Exemples
1 mot	chose arbre que signaler elles instruments regardant tendu copie états coeur plus
2 mots	des taches les sources une cabane son chien le client des trous
3 mots	Résoudre un problème réparent le plafond il croit lire achètera une télévision
4 mots	le nord du pays il déteste la couleur elles lisent leur noms
6 mots	la souris qui grignote sa tartine le passant examine le luxueux canapé
12 mots	il reçoit un sac de céréales cultivées dans le nord du pays

FIGURE 3.10 – Exemple de stimuli utilisés dans l'expérience (Devauchelle, 2008; Pallier et al., 2011). 12 mots présentés de façon sérielle visuellement. 5 tailles de constituants différentes : 12 mots, 6, 4, 3, 2 et 1. La condition avec des constituants de 1 mot correspond à une liste de mot dans laquelle aucun des mots ne peut se suivre pour former un plus grand constituant.

façon suivante (figure 3.11). D'abord, des phrases à branchement à droite de 12 mots ont été inventées. La structure syntaxique à branchement à droite permet d'éviter tout phénomène de mouvement et d'étudier spécifiquement le phénomène de merge. Ensuite, des constituants de taille variable ont été sélectionnés dans ces phrases (voir figure 3.11).

Les constituants de même taille provenant de phrases différentes ont ensuite été recombinaés entre eux pour former des stimuli contenant toujours 12 mots contrôlant ainsi la taille totale des stimuli. La théorie linguistique sous-jacente est qu’une seule opération

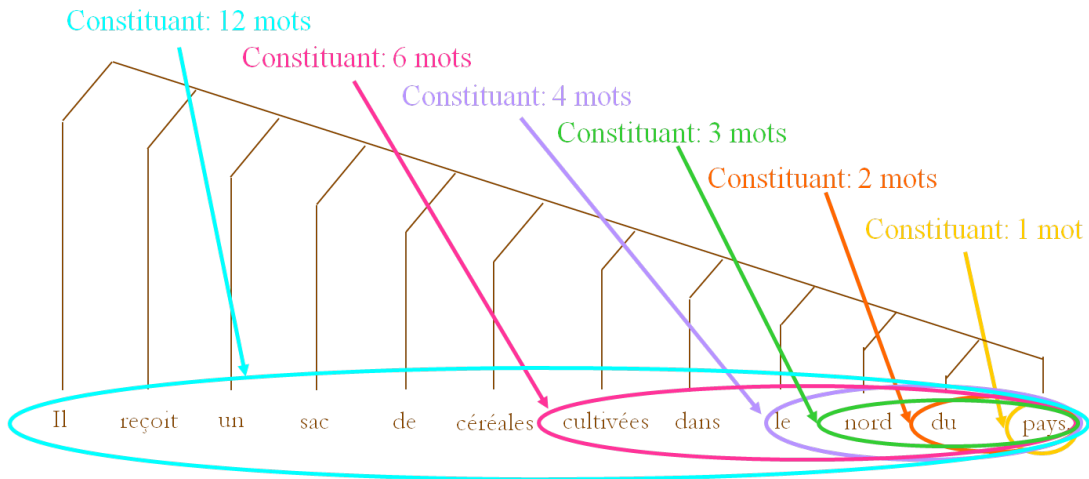


FIGURE 3.11 – Schéma de création des stimuli utilisés dans l’expérience de langage écrit. Un arbre à branchement à droite est créé. Puis des constituants de taille variable sont sélectionnés à partir de cet arbre syntaxique.

d’intégration (merge) est nécessaire pour intégrer un mot dans un constituant à branchement à droite. Par exemple, pour calculer la structure syntaxique du constituant « *du pays* » il faut effectuer une opération pour intégrer *du* et *pays*. De la même façon, il faudra une opération supplémentaire pour intégrer le mot *nord* avec le constituant « *du pays* », on aura donc 2 opérations syntaxiques pour le constituant « *nord du pays* ». En changeant la taille des constituants, on fait ainsi varier le nombre d’opérations merge requises pour calculer la structure syntaxique associée au constituant. Les variations cérébrales d’oxygénation de l’hémoglobine ont été enregistrées grâce à l’IRM fonctionnelle lors de la présentation visuelle de stimuli de ce type. Le paradigme repose sur l’hypothèse qu’une région impliquée dans l’intégration d’éléments syntaxiques devrait s’activer de plus en plus à mesure que le nombre d’opérations requises pour créer une structure augmente (figure 3.12). La figure 5.1 montre les résultats obtenus dans cette expérience : un réseau de l’hémisphère gauche de régions temporales (pôle temporal, aSTS, pSTS et TPJ), frontales inférieures (IFG orb et tri) et le putamen étaient sen-

sibles à la taille des constituants. Afin de s'affranchir du contenu lexical et de contrôler pour les probabilités de transitions entre les mots, la même expérience a été réalisée sur vingt autres sujets en présentant cette fois-ci uniquement des pseudo-mots. Les mêmes constructions syntaxiques ont été réutilisées et les mots de contenu ont été remplacés par des pseudo-mots (Jabberwocky). Trois régions sensibles à la taille des constituants et

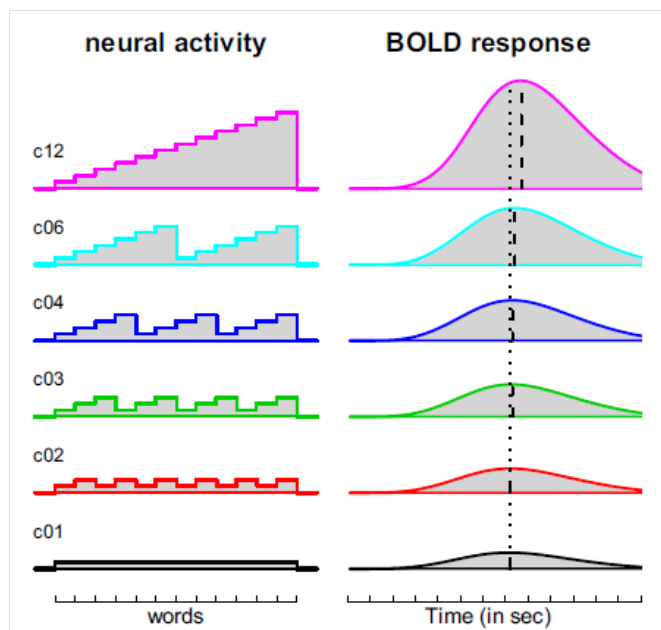


FIGURE 3.12 – Simulation du modèle simple d'activité neurale évoquée par le traitement des constituants de taille variable. Le modèle prédit une augmentation en amplitude du signal BOLD et de la phase d'activation à mesure que la taille des consituant augmente. Extrait de (Pallier et al., 2011).

indépendantes du contenu lexico-sémantique et des probabilités de transition entre les mots ont été mises en évidence : l'IFG orb et tri et le pSTS. L'existence de telles régions sensibles à la structure syntaxique indépendamment de toute référence lexicale, soulève une question importante : sont-elles sensibles à la structure en constituant en général, indépendamment du domaine ? Pourraient-elles être le siège de l'intégration d'éléments dans une structure syntaxique ? Nous avons voulu répondre à ces questions en utilisant chez les mêmes sujets des stimuli linguistiques auditifs et musicaux basés sur le même paradigme.

De nouveaux stimuli langagiers avec des tailles de constituants variables ont été

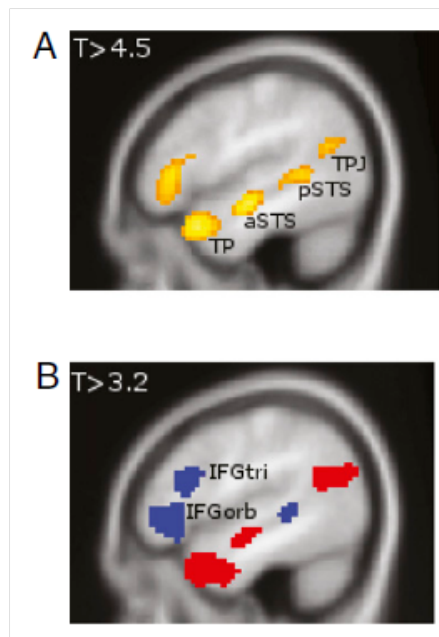


FIGURE 3.13 – Résultats obtenus dans l'expérience linguistique à l'écrit (Pallier et al., 2011). Régions qui montrent une activation à mesure que la taille des constituants augmente. A. Résultats IRMf pour les mots normaux, analyse de groupe $T > 4.5$, $p < 0.05$ FWE, taille de cluster > 10 voxels. B. Les régions en bleu sont sensibles à la taille des constituants aussi bien avec des pseudo-mots (jabberwocky) qu'avec des mots, les régions en rouge sont sensibles uniquement aux vrais mots et ont donc besoin d'un contenu lexical, analyse de groupe, $T > 3.2$, $p < 0.001$ seuil non corrigé

construits. Le même principe de construction a été adapté à des phrases musicales, c'est-à-dire les phrases musicales ont été coupées en plusieurs fragments de longueurs variées. Les stimuli musicaux ont été construits de façon empirique en copiant le procédé utilisé en langage c'est-à-dire que les phrases musicales existantes ont été fragmentées. La taille des fragments variait en fonction du nombre de temps. Les fragments en musique ne sont donc pas des constituants syntaxiques de la même façon qu'en langage puisque plusieurs types de structure ont été cassés par ce processus : harmonique, rythmique ... Dans l'analyse, il sera impossible de dissocier ces différentes structures. Il nous a paru pertinent de détruire tous les types de structure sans se focaliser sur une seule, puisque s'agissant de la première expérience paramétrique sur la structure musicale il fallait maximiser les chances d'obtenir un résultat détectable en IRMf.

Le chapitre suivant sera consacré à la validation de nos stimuli langagiers et musicaux

puisque la méthode de création a été empirique. Il a donc fallu vérifier que les différents niveaux de structure étaient perceptibles en langage auditif et en musique.

Structure hiérarchique en langage et en musique : étude comportementale de validation des stimuli

Sommaire

4.1	Expérience 5 : Structure en Langage et en Musique chez des non-musiciens Français	120
4.2	Perception consciente et traitement automatique de la structure en musique	128
4.2.1	Expérience 6 : Perception consciente, notation des stimuli musicaux	129
4.2.2	Expérience 7 : Traitement automatique, discrimination des stimuli musicaux	138
4.2.3	Discussion générale	146

4.1 Expérience 5 : Perception de la structure hiérarchique en langage et en musique par des sujets non musiciens français

Le but de cette expérience est de valider la construction empirique de nos stimuli musicaux en nous assurant que les différents niveaux de structure sont perceptibles de la même façon dans les deux domaines. Les stimuli musicaux et langagiers ont été créés en suivant des méthodes identiques : en détruisant de façon paramétrique leur structure syntaxique (en musique plusieurs types de structure ont ainsi été détruites : rythmique, harmonique, mélodique ...). Les stimuli ont été présentés de façon auditive et les sujets devaient leur donner une note entre 0 et 5 pour juger de leur cohérence, de leur structure.

Matériel et méthodes

Participants

Dix sujets droitiers, de langue maternelle française (âgés de 19 à 25 ans, en moyenne de 21.8 ; 8 femmes et 2 hommes) ont participé à cette expérience comportementale. Aucun n'avait suivi de cours de musique en plus des cours dispensés en classe de musique au collège, 4 années pendant lesquelles tous les élèves français reçoivent un enseignement musical d'une heure par semaine. Aucun sujet n'a mentionné de déficit auditif. Tous les volontaires ont signé un formulaire de consentement et reçu 8€ de dédommagement pour leur participation.

Stimuli

Les stimuli musicaux ont été créés à partir d'extraits de sonates de Mozart. Seize mélodies de quatre mesures à 4 temps ont été choisies. Onze tonalités majeures sont représentées parmi nos stimuli (Do, Do#, Ré, Mi, Fa, Fa#, Sol, La, Sib, Si). Les extraits ont légèrement été modifiés de sorte qu'ils finissent tous par une cadence parfaite. Cette cadence parfaite (succession des degrés 5 et 1 de la tonalité) donne aux auditeurs

une sensation de fin (Krumhansl and Kessler, 1982), et joue un rôle de ponctuation identique au point en fin de phrase. Le même tempo a été utilisé pour tous les stimuli. Chacune des 16 mélodies a été découpée en 16 intervalles de temps égaux correspondant à un temps musical (figure 4.1).

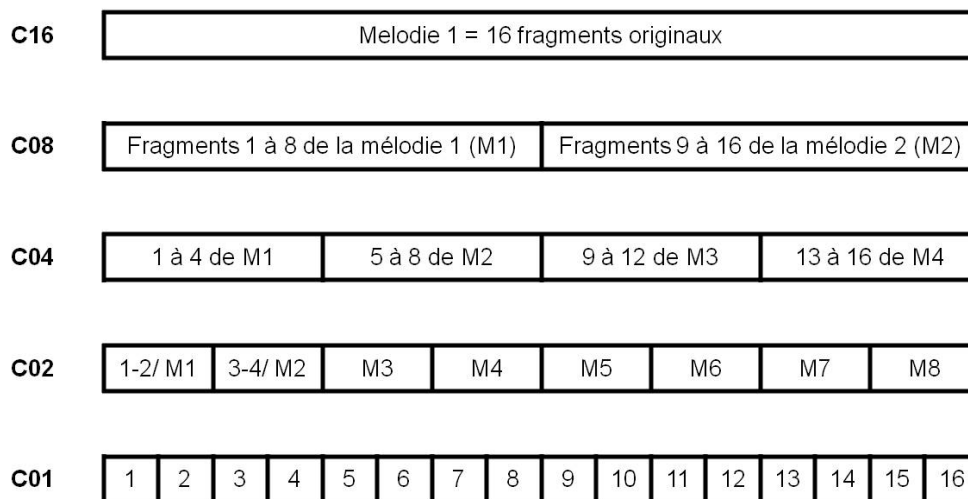


FIGURE 4.1 – Schéma des différents niveaux de structure des stimuli musicaux

256 fragments de même durée ont ainsi été créés. 64 nouvelles mélodies de durée identiques aux mélodies originelles ont été synthétisées en assemblant 16 fragments en suivant le schéma rotatif décrit dans le schéma 4.2 page suivante de sorte qu'un fragment donné apparaît exactement une seule fois par condition. 80 mélodies au total, appartenant à 5 conditions de structure ont été créés : c01 (constituant de taille 1), c02, c04, c08 et c16 comme sur la figure 5.5 page 155 contenant 16 stimuli par condition.

La figure 4.3 page 123 résume les étapes de construction des stimuli. La figure 5.5 page 155 explique le contenu exact de chaque condition. Ainsi, la catégorie c16 contient les 16 mélodies initiales c'est-à-dire un constituant de taille 16 temps. Les mélodies c08 sont obtenues en concaténant 2 séquences de 8 temps issus de 2 mélodies différentes, c'est-à-dire qu'elles contiennent 2 constituants de 8 temps. Par exemple la première mélodie de ce niveau de structure est obtenue en collant la première moitié de la mélodie 1 avec la seconde moitié de la mélodie 2 ; la 2ème mélodie contient la première moitié de la mélodie 3 et la seconde moitié de la mélodie 4 et ainsi de suite pour toutes les mélodies.

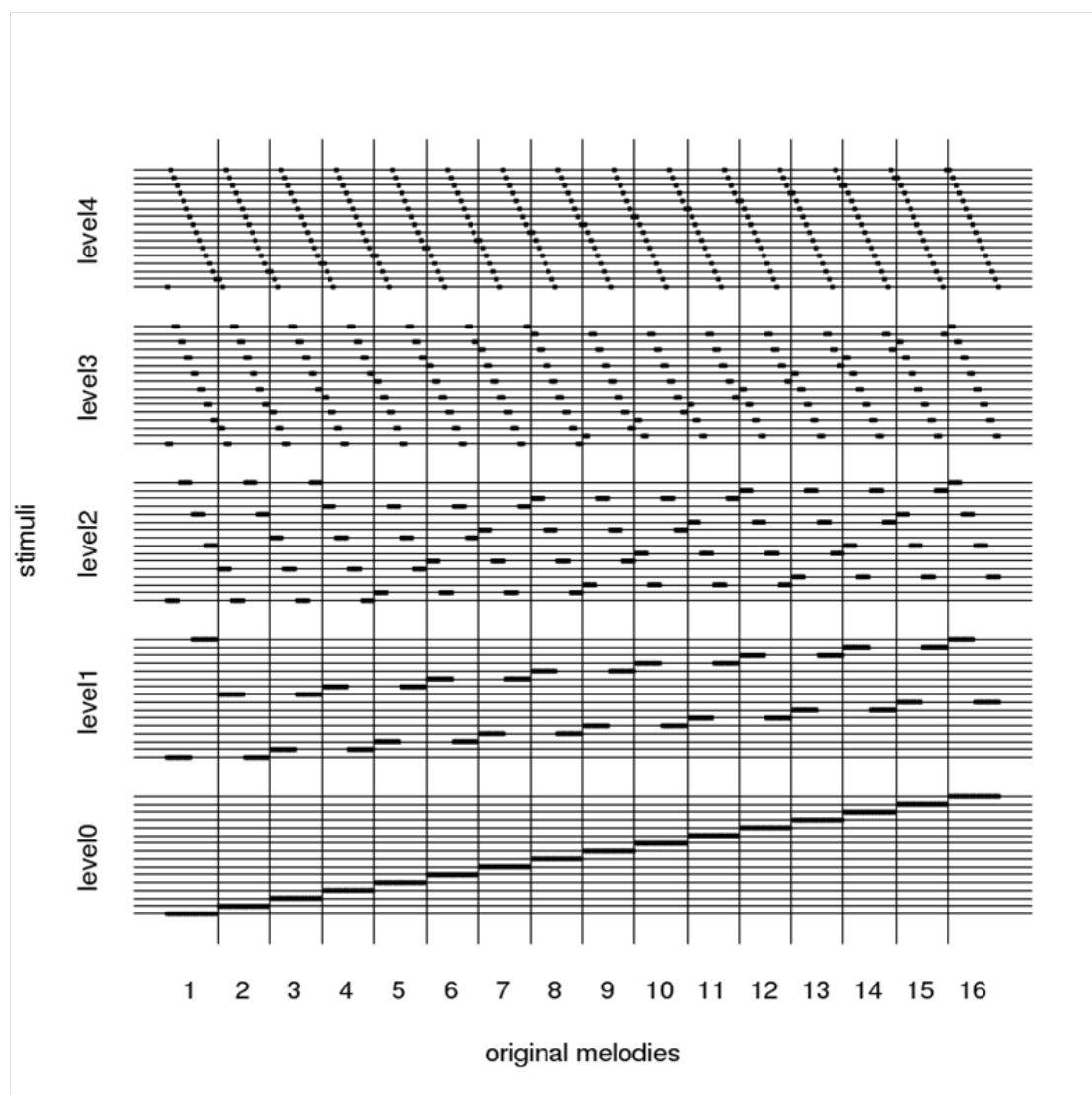


FIGURE 4.2 – Schéma rotatif de concaténation des différents fragments. En abscisse, le numéro des mélodies initiales, la longueur des traits correspond à leur taille en temps, initialement il y a 16 temps. En ordonnée, les différents niveaux de structure : level 0 = niveau initial c16, level 1 = c08 constituant de taille 8, level 2 = c04, level 3 = c02, level 4 = C01. Chaque ligne horizontale représente une mélodie, la ligne la plus basse de chaque niveau est la numéro 1, la plus haute, la numéro 16. Le stimulus 1 du niveau 0 ou c16 contient uniquement la mélodie initiale 1, le stimulus 1 du niveau 1 ou c08 contient la première moitié, soit 8 temps, de la mélodie 1 et les 8 derniers temps de la mélodie 16. Le stimulus 1 du niveau 2 ou C04 contient les 4 premiers temps de la mélodie 1, les 4 temps suivants de la mélodie 16, les 4 temps suivants de la mélodie 12 et les 4 derniers temps de la mélodie 8 et ainsi de suite pour les 5 niveaux et les 16 stimuli de chaque niveau.

Extrait original



Etape 1
Cadence parfaite à la fin de chaque extrait



Etape 2
Découpe en fragments



Etape 3
Recombinaison des fragments des différentes mélodies



FIGURE 4.3 – Schéma des étapes de création des stimuli musicaux

16 mélodies sont ainsi créées pour le niveau c08. Les mélodies de niveau c04 contiennent 4 constituants de 4 temps issus de 4 mélodies initiales différentes. Les mélodies de niveau c02 sont constituées de 8 composants de 2 temps issus de 8 mélodies initiales différentes. Enfin, les mélodies du plus bas niveau de structure, c01 sont créés en collant bout à bout 1 temps provenant de chacune des 16 mélodies initiales. Les positions des temps dans les phrases sont respectées à travers tous les stimuli, c'est-à-dire qu'une note apparaissant au temps numéro 12 d'une mélodie initiale, se retrouvera toujours au temps 12 dans les mélodies des autres niveaux de structure.

Afin de maximiser le sentiment de déstructuration, à l'écoute des différents niveaux de structure, nous avons collé côte à côte des fragments issus de mélodies dans des tonalités aussi éloignées que possible.

Les partitions des mélodies initiales ont été transcrites dans un logiciel de synthèse musicale avec une fonte sonore de piano et sauvegardées au format Midi (Cannam et al.,). Les débuts et fins de chaque fragment ont été marqués. Un programme informatique a extrait les différents fragments des fichiers sons des mélodies initiales et les a recombinaisonnés en nouveaux fichiers midi, puis les a convertis au format wav en utilisant Timidity

C16	Mon père parle souvent de choses très savantes qui fascinent les copains qui viennent chez moi															
C08	Paul mange la tarte sur laquelle se trouvait							qui fascinent les copains qui viennent chez moi								
C04	Il lance la balle			les grimaces des clowns			la biche sort vite			qui sent les fleurs						
C02	Il manque		est tombé	les fleurs		avec soin	très raide		du Brésil							le blé
C01	La	à	sortent	tarte	le	un	grimpe	Luc	les	qui	siège	pains	qui	de	réchauffe	jaunes

TABLE 4.1 – Exemple de phrases pour chaque niveau de structure

(Toivonen and Izumo, 2004).

Pour construire les stimuli langagiers, nous avons d’abord inventé seize phrases contenant chacune 16 mots en respectant les contraintes suivantes (voir l’annexe B page 237 pour la liste complète des phrases inventées.) :

- les mots sélectionnés pour les phrases étaient tous mono ou bi-syllabiques. (Fréquence de mots à calculer pour prouver qu’ils sont fréquents) ;
- il ne devait pas y avoir de liaison entre les mots ;
- les phrases devaient avoir une structure syntaxique « binaire » nous permettant ensuite de les couper en fragments de 2, 4 ou 8 mots ;
- ces fragments devaient être des constituants syntaxiques (voir l’annexe B page 237 pour la liste complète des phrases)

Les constituants ainsi formés ont été concaténés pour donner de nouvelles phrases contenant toujours 16 mots et créant ainsi 5 niveaux de structure selon le même principe que pour les stimuli musicaux. 16 stimuli ont ainsi été créés pour chaque niveau de structure (c01 constituant de 1 mot, c02 constituant de 2 mots, c04 constituants de 4 mots, c08 constituant de 8 mots, c16 phrases initiales ou constituant de 16 mots). Nous avons vérifié dans les nouvelles phrases que les suites de fragments ne pouvaient pas former de nouveaux constituants plus grands que prévu. Par exemple, dans la condition la plus déstructurée c01, deux mots consécutifs ne pouvaient jamais former un constituant syntaxique. La table 4.1 montre un exemple de stimuli langagiers utilisés dans l’expérience.

Les phrases initiales ont ensuite été enregistrées par une locutrice de langue maternelle française. Les phrases ont été prononcées avec des pauses entre les mots mais en gardant le plus possible une prosodie phrasale. Les mots ont ensuite été extraits en fichiers individuels en utilisant le logiciel Praat (Boersma and Weenink, 2010). Ces fichiers

individuels ont alors été concaténés en ajoutant un silence de 40ms entre chaque mot de façon à créer des fichiers sonores pour nos 80 stimuli expérimentaux. Finalement, nous avons égalisé à 6.5 sec la durée des 80 stimuli en utilisant l'algorithme PSOLA implémenté dans Praat. Nous avons également normalisé le niveau sonore des phrases (racine des moyennes au carré de l'intensité).

Procédure

Les participants étaient informés qu'ils allaient entendre des stimuli musicaux et langagiers et qu'ils allaient devoir noter la cohérence de ces stimuli. Plus précisément, on leur a demandé de noter la structure des phrases. Pour cela, ils disposaient d'une échelle de 0 à 5; la note 0 devait être assignée à une structure très mauvaise, la note 5 devait correspondre à une structure qui paraissait correcte. Les sujets entendaient les 80 stimuli musicaux et les 80 langagiers sous forme de 4 blocs de 40 stimuli chacun. Les blocs contenaient uniquement des stimuli d'un domaine : soit langage, soit musique. L'ordre des stimuli dans les blocs a été mélangé de façon aléatoire pour chaque participant. Les stimuli étaient présentés à une intensité sonore confortable par un casque. Le programme de stimulation attendait que le sujet presse une des touches de 0 à 5, enregistrerait la réponse et attendait 2.5 sec avant de jouer le stimulus suivant.

Résultats

La figure 4.4 page suivante présente la moyenne des notes données par sujet en fonction du niveau de structure construit. Chaque ligne représente un sujet. Une Anova à 2 facteurs intra-sujets : le facteur Niveau (c01, c02, c04, c08, c16) et le facteur Domaine (musique ou langage), a été réalisée. De façon cruciale, l'analyse de variance confirme l'effet principal de Niveau $F(4, 36) = 87, p < 10^{-16}$ (visible sur les graphiques) qui reflète le fait que les sujets notent et perçoivent différemment les niveaux de structure. L'effet principal du Domaine $F(1, 9) = 13, p < 0.006$ est significatif et reflète le fait que les sujets répondent différemment au langage et à la musique. L'interaction significative entre les 2

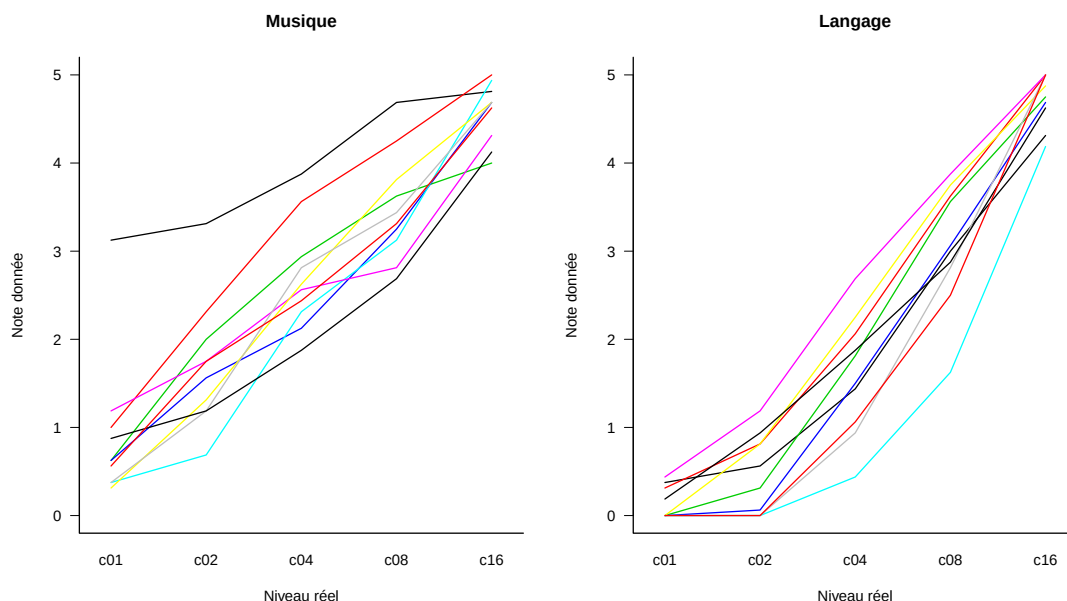


FIGURE 4.4 – Tâche d’estimation du niveau de structure : notes données en fonction des niveaux de structure construits. Chaque courbe représente les notes données pour un sujet.

facteurs $F(4, 36) = 14, p < 10^{-7}$ montre que les comportements de notation des niveaux de structure sont différents pour les deux domaines. Cette différence de comportement est visible en comparant les pentes des différentes courbes. La pente des courbes du langage est plus importante que celle de la musique d’où l’interaction entre les 2 facteurs. Pour aller un peu plus loin, nous avons regardé l’importance de l’effet de structure dans chaque domaine. Tout d’abord en restreignant l’anova au langage, on trouve que l’effet du niveau de structure est toujours très significatif $F(4, 36) = 318, p < 10^{-16}$. De même pour l’analyse restreinte aux conditions musicales, l’effet du niveau de structure sur la note donnée par les sujets est significative $F(4, 36) = 120, p < 10^{-16}$.

Discussion

Les sujets classent les stimuli selon leur niveau de structure, aussi bien en langage qu’en musique. De plus, la bonne corrélation entre les notes données et les valeurs effectives du niveau de structure permettent de conclure que les sujets perçoivent bien les

niveaux de structure et ce dans les deux domaines. L'interaction significative entre le domaine et le niveau de structure reflétée par la différence de pente entre les courbes des deux domaines peut s'expliquer de la manière suivante. La pente plus grande pour le langage montre que la déstructuration des phrases en langage est perçue de façon plus prononcée. Cela revient à dire que pour un même niveau de structure, la note donnée va être plus basse pour le langage que la musique. Ce phénomène est d'autant plus prononcé que les stimuli sont déstructurés. Les sujets sont plus « choqués » par la déstructuration en langage qu'en musique. Les stimuli langagiers portent un sens qui va être perdu assez rapidement avec la déstructuration, la cohérence perçue est influencée par des facteurs sémantiques et pas seulement syntaxique. En musique, la structure vient non seulement de l'organisation temporelle mais aussi du rapport entre les différentes hauteurs des notes ; on ne peut pas assigner un sens aux notes. De plus, il existe de la musique déstructurée par rapport à de la musique plus classique et qui peut s'expliquer par un relâchement des contraintes structurales, voir un abandon de ces contraintes ou une pondération particulière des règles, notamment dans la musique contemporaine. Pourtant cette musique déstructurée reste encore perçue en tant que musique. En s'affranchissant de la composante sémantique grâce à la musique, nous pouvons tester la perception des structures hiérarchiques dans des populations différentes, par exemple chez de jeunes enfants en cours d'acquisition du langage, ou chez des adultes de langues et cultures différentes. Néanmoins avant de pouvoir nous servir de ces stimuli de musique classique occidentale pour étudier la capacité humaine de construction de structures hiérarchiques, il faut s'assurer que cette structure hiérarchique possède effectivement des caractéristiques universelles et que leur perception n'est pas seulement liée à un apprentissage culturel. Tout naturellement, nous avons eu envie de tester la perception de la structure dans nos stimuli musicaux en les soumettant à un groupe d'individus issus de culture non occidentale avec peu (ou pas) de contact avec celle-ci.

4.2 Français et Mundurukus : perception consciente et traitement automatique de la structure en musique occidentale

Nous avons ainsi pu tester des personnes d’une tribu amazonienne, les Mundurukus¹. Les Mundurukus sont une ethnie d’environ 7000 personnes vivant dans des territoires autonomes dans l’état de Par en Amazonie au Brésil (figure 4.5) et ils sont peu ou pas exposés à la musique occidentale. Cette population est particulièrement intéressante parce



FIGURE 4.5 – Carte du territoire des Mundurukus

qu’elle est très isolée. Certains mundurukus ont quelques contacts avec des individus non-indigènes au travers des institutions gouvernementales et des missions. Quelques uns parlent aussi le portugais, et les enfants vont à l’école où ils apprennent des connaissances de base. Leur contact avec la civilisation occidentale est limité et leur exposition à la musique classique occidentale est très restreinte. Les Mundurukus n’ont pas de radio, mais entendent pour certains la musique produite lors de la messe à la mission, pour

1. grâce à une collaboration avec un linguiste Pierre Pica, spécialiste de cette population. La langue Mundurukù, une langue amérindienne, présente la particularité de n’avoir de mots que pour les 5 premiers chiffres : un, deux, trois, quatre, cinq (une main), pour les nombres plus grands il existe 3 mots : deux mains, quelques-pas beaucoup, beaucoup-vraiment beaucoup.

d'autres la musique brésilienne notamment du groupe banda Calypso grâce à des DVD de musique². Dans la langue Mundurukù, il n'y a qu'un seul mot pour musique et chant. Le vocabulaire de la musique est difficile à trouver.

4.2.1 Expérience 6 : Perception consciente, notation des stimuli musicaux

Dans une première expérience, les Mundurukus ont réalisé une tâche de jugement de cohérence des stimuli musicaux. Les consignes et la tâche ont été adaptées aux conditions de passation particulières que représentent les expériences sur le terrain dans la jungle amazonienne et l'analphabétisme des sujets :

- Diminution de la durée des expériences pour que les sessions de test ne dure pas plus de 5 minutes.
- Compression des stimuli pour raccourcir le temps de présentation
- Traduction des explications et les avons adapté à leur compréhension notamment en ajoutant un entraînement en utilisant la langue Mundurukù.
- Comme les Mundurukus ont un concept arithmétique restreint, nous avons utilisé des visages plus ou moins souriant pour qu'ils effectuent la notation.

De plus, dans le but d'étudier l'influence de la tonalité dans la perception de la structure, les stimuli précédents ont été transposés en Do majeur et ajoutés aux précédents.

Matériel et méthodes

Participants

14 Mundurukus, âgés de 12 à 64 ans, moyenne d'âge de 33.0 (sexe des sujets non renseigné) ont participé à cette expérience. Trois d'entre eux ont passé les 2 parties de l'expérience (2 sessions de 80 essais chacune soit 160 essais), les 11 autres sujets n'ont passé qu'une seule des 2 sessions et ont donc entendu 80 stimuli. Nous n'avons exclu aucune donnée. 13 sujets français non musiciens (n'ayant pas plus d'un an de pratique

2. La musique jouée par le groupe banda calypso est une musique pop, mélange de rythmes caribéens et de rythmes régionaux du Par

instrumentale), 12 femmes et 1 homme, âgés de 19 à 28 ans, moyenne d'âge 22.4, ont passé exactement la même expérience comme contrôle.



FIGURE 4.6 – Photo de Mundurucus passant l'expérience dans un village.

Stimuli

Les 80 stimuli musicaux de l'expérience précédente ont été utilisés après avoir été compressés pour les raccourcir : 4.5 secondes en raison de la diminution de la durée de l'expérience. A ces 80 stimuli originaux se sont ajoutés 80 stimuli transposés tous dans la même tonalité : Do majeur, pour étudier l'influence de la tonalité. Pour créer les stimuli transposés, nous avons utilisé Rosegarden ?? page ?? pour rééditer et transposer en Do majeur les stimuli de niveau de structure maximal (c16) issus des tonalités différentes. Puis, nous avons suivi le même processus de construction décrit dans la première expérience de ce chapitre pour les différents niveaux de structure. Au total, 160 stimuli ont été entendus dans l'expérience sous forme de 2 sessions de 80 stimuli. Chaque session contient 40 stimuli originaux - avec des fragments dans des tonalités différentes - et 40 stimuli différents transposés - avec des fragments tous dans la même tonalité. Afin d'expliquer la tâche aux sujets mundurukus, des stimuli linguistiques en Mundurukù ont été construits. En effet, les sujets français qui avaient passé l'expérience précédente ont

signalé, lors de l'entretien à la fin de l'expérience, que les stimuli linguistiques les avaient aidés à mieux comprendre la tâche en musique. Sur cette base, des stimuli en langue Mundurukù ont été créés, avec l'aide du linguiste Pierre Pica, sur le même modèle que les stimuli linguistiques de l'expérience précédente. L'idée est de produire une dégradation progressive de la structure des phrases. Un stimulus pour chacun des 5 niveaux de structure (figure 4.7) a été construit de la façon suivante. Des phrases enregistrées sur place lors d'une mission précédente ont servi de base pour créer tous les stimuli. Deux phrases ont été choisies puis concaténées pour former le niveau le plus structuré, jouant le même rôle que le niveau de structure c16 en musique. Tous les stimuli contenaient toujours tous les mots de ces 2 phrases. Le premier niveau de déstructuration, équivalant à C08, a été produit en changeant un mot de place dans chaque phrase. Pour le deuxième niveau de déstructuration, équivalant à C04, les 2 mots dont la position avait varié ont été échangés. Pour créer le 3^{ème} niveau de déstructuration, équivalent de C02, dans chaque phrase, les mots ont été mélangés et un mot de chaque phrase a été échangé. Le dernier niveau le plus déstructuré, correspondant à c01, a été obtenu comme C02 en mélangeant les mots dans chaque phrase et en échangeant 2 mots de chaque phrase. Cinq phrases d'entraînement ont ainsi été créées. Ces stimuli (figure 4.7) sont téléchargeables sur SITE ELO.

C16	ti akay	chum	pima	oye	pirin	apen	puje	ipi dan	ebutbut	sowat
C08	pima	ti akay	chum	oye	pirin	puje	apen	ipi dan	ebutbut	sowat
C04	puje	ti akay	chum	oye	pirin	pima	apen	ipi dan	ebutbut	sowat
C02	oye	pirin	pima	ti akay	puje	apen	chum	ipi dan	sowat	ebutbut
C01	oye	pirin	pima	ebutbut	puje	apen	chum	ipi dan	sowat	ti akay

FIGURE 4.7 – Stimuli langagiers utilisés en entraînement de la tâche. Les couleurs représentent chacune des phrases initiales.

Procédure

Pour entraîner les participants mundurukus à la tâche, les stimuli linguistiques spécialement conçus pour l'entraînement ont été présentés dans un ordre fixé. Le programme d'entraînement revenait au début tant que la tâche n'était pas comprise. L'expérimentateur s'assurait de la bonne compréhension de la tâche. Comme dans la phase de test, les sujets devaient cliquer sur le visage qui leur semblait correspondre au stimulus. L'expérimentateur a expliqué les différents niveaux de structure en utilisant les adjectifs « très bon », « bon », « un peu bon », « mauvais », « très mauvais ». Les participants écoutaient les stimuli un par un, avec un casque. Après chaque stimulus, une image apparaissait contenant 5 visages plus ou moins souriant (5 niveaux de sourire, figure 4.8). Le programme attendait la réponse du sujet pour passer à l'essai suivant. Le programme de stimulation a été codé en python et recueillait les clics de souris après chaque stimulus. L'entraînement ne contenait que des stimuli langagiers.

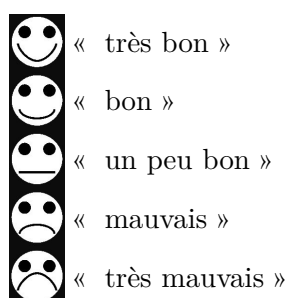


FIGURE 4.8 – Icônes utilisées pour sélectionner la réponse.

Résultats et Discussion

La figure 4.9 page suivante présente les notes moyennes données par les sujets en fonction du niveau de structure. Chaque ligne colorée représente la moyenne des réponses d'un sujet en fonction des différents niveaux de structure.

Une analyse de la variance, ANOVA à 3 facteurs, a été réalisée pour tester comment les différents facteurs et leurs interactions pouvaient expliquer les résultats. Le facteur principal Niveau de structure comprenant 5 valeurs possibles (C01, C02, C04, C08, C16)

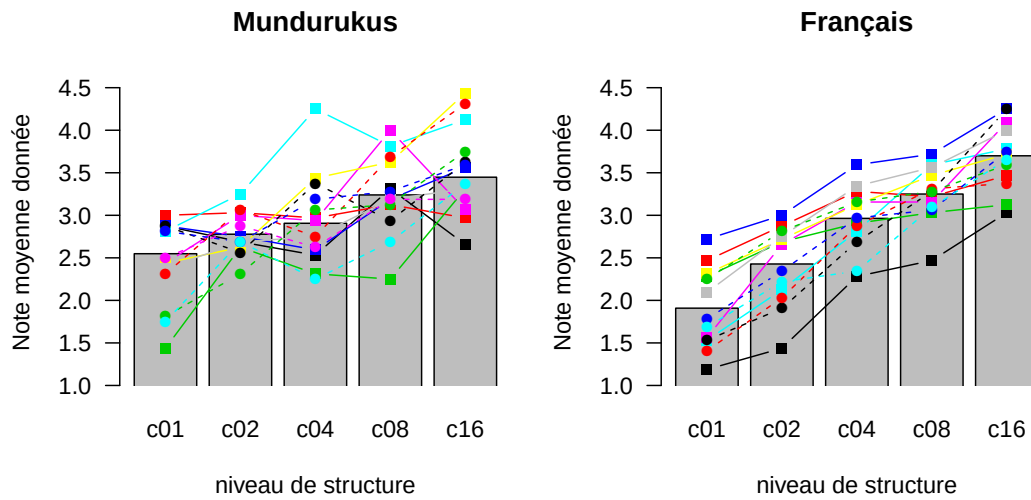


FIGURE 4.9 – Moyenne des notes données par les participants Mundurukus à gauche et français à droite. Les barres grises représentent les moyennes à travers les sujets. Chaque ligne représente la moyenne des notes données par sujet pour les différents niveaux de structure

est très significatif : $F(4, 100) = 89, p < 10^{-16}$. Le facteur principal Groupe (2 niveaux - Mundurukus ou Français) n'est pas significatif $F(1, 25) = 1.2$. Et le facteur tonalité (intrasujet contenant 2 niveaux également - fragments en Do M ou dans toutes les tonalités) est significatif $F(1, 25) = 92, p < 10^{-10}$, les réponses sont données différemment selon que les fragments sont tous dans la même tonalité ou qu'ils sont dans des tonalités différentes. Les interactions avec le niveau de structure se révèlent significatives.

Ainsi, l'interaction entre le facteur structure et le groupe de sujets $F(4, 100) = 9, p < 10^{-6}$, nous montre que les 2 groupes de sujets ne traitent pas la structure de la même façon : la relation entre le niveau de structure et le jugement des sujets est plus forte pour les participants français.

Les réponses données par les Mundurukus sont plus dispersées que celles des français. La variabilité est plus importante chez les Mundurukus que chez les français pour qui les moyennes individuelles (les lignes) ressemblent plus à des fonctions linéaires croissantes.

L'interaction significative entre le facteur structure et le facteur tonalité $F(4, 100) = 15, p < 10^{-10}$, montre que la tonalité a eu un impact sur la notation et donc sur la

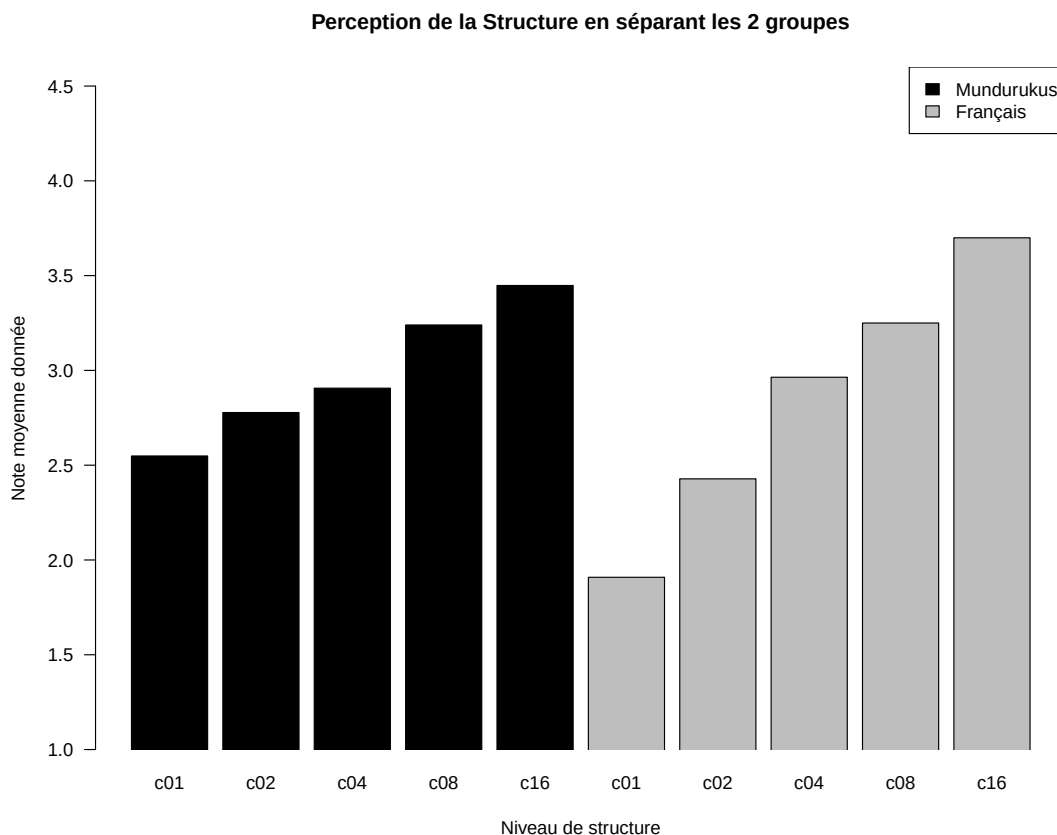


FIGURE 4.10 – Comparaison des 2 groupes de sujet sur la perception de la structure.

perception du niveau de structure (figure 4.11 page ci-contre). A part pour le niveau le plus bas, les notes données pour les stimuli issus de tonalités différentes sont meilleures que celles toutes issus de la même tonalité. Il semble que l'évaluation ait été rendue plus difficile par l'absence de changement de tonalité.

L'interaction triple (niveau de structure, groupe, tonalité) n'est pas significative ($F(4, 100) = 2, p < 0.09$), les 2 groupes de sujets ont subi essentiellement le même impact de la tonalité sur la notation. On regardera donc cette interaction sur le groupe entier de sujets (mundurucus et français).

Il est tout à fait possible que les participants aient repéré les différents niveaux de structure en s'aidant des changements de tonalité existant à la frontière entre les fragments, l'effet de tonalité étant fortement significatif. Il est donc crucial de véri-

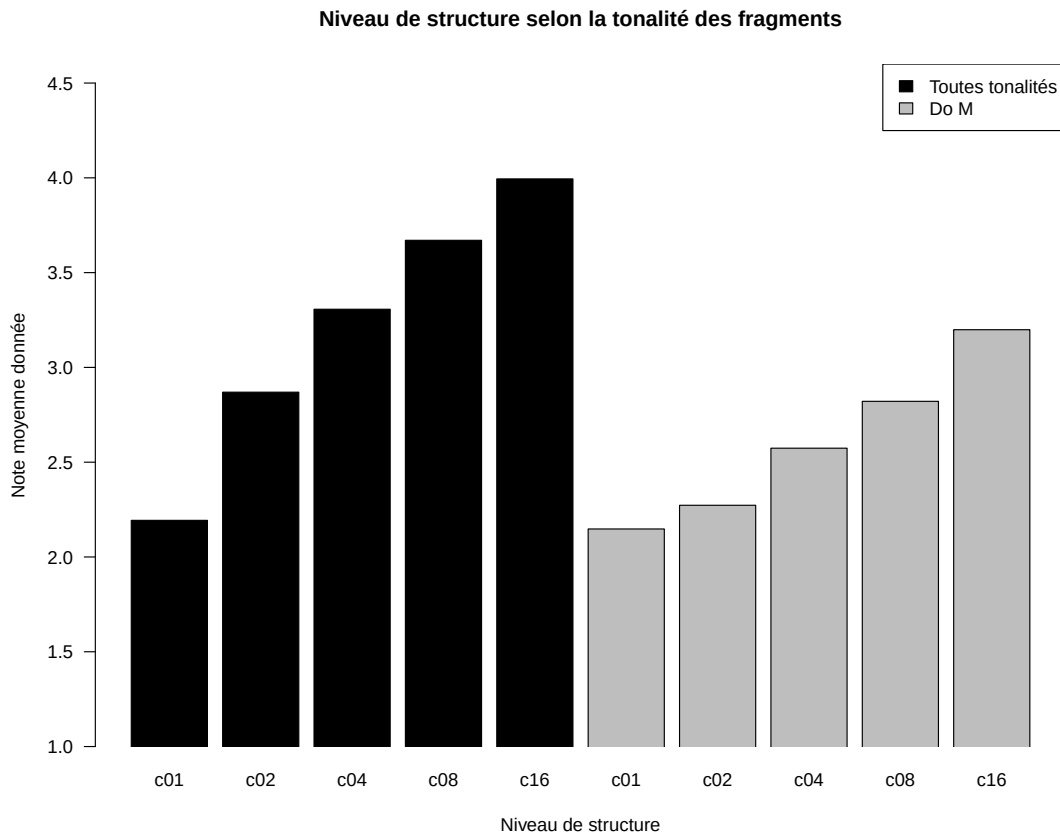


FIGURE 4.11 – Effet de la tonalité sur la perception de la structure.

fié que l'effet de structure est également présent alors qu'il n'y a pas de changement de tonalité (condition Do Majeur) c'est-à-dire dans la moitié des stimuli entendus (résultats figure 4.12 page suivante). En restreignant l'analyse statistique aux seules données collectées pour les stimuli en Do M, le facteur structure est toujours significatif, $F(4, 100) = 51, p < 10^{-16}$, ainsi que l'interaction avec le facteur groupe ($F(4, 100) = 7, p < 0.0001$). La déstructuration progressive des stimuli est bien perçue même si on enlève les informations de tonalité qui rendaient saillantes les frontières entre les fragments et ce aussi bien chez les sujets français que chez les mundurukus (figure 4.2.1 page 137).

En comparant les notes données aux stimuli en Do M aux réponses données aux stimuli avec un changement de tonalité entre les fragments, nous pouvons déduire que

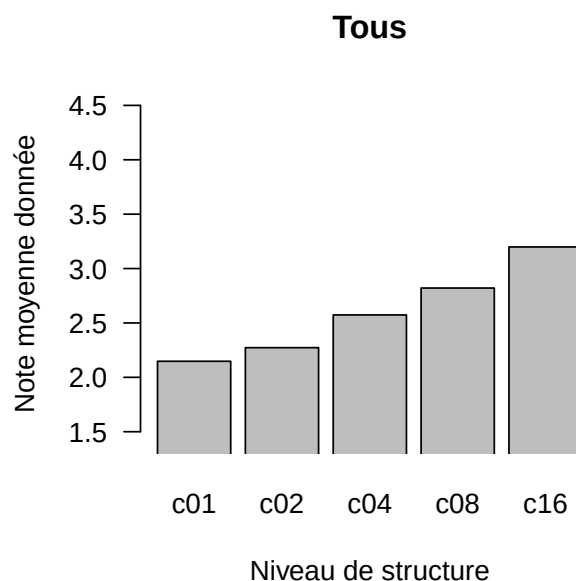


FIGURE 4.12 – Moyenne des notes données par tous les participants aux seuls stimuli en Do majeur.

le changement de tonalité a aidé à la perception de structure mais qu'elle n'est pas la seule responsable puisque l'effet est toujours visible en ayant ôté ce paramètre.

Les Mundurukus, comme les français, sont capables de noter le degré de structure d'un morceau de musique classique occidentale. Les notes données corrélaient avec le niveau de structure effectif. La perception de la structure n'est pas seulement due à la différence de tonalité aux frontières entre les fragments puisque l'effet est toujours présent, même si moins fort, lors de la présentation de stimuli issus tous de la même tonalité. Les 2 groupes de sujets, français et mundurukus ne traitent pas la structure exactement de la même façon : les effets sont plus robustes pour les français, habitués à ce type de musique. A ce niveau de l'étude, nous ne pouvons pas savoir si notre effet est effectivement dû à la manipulation structurelle ou à des effets plus bas niveau. Les Mundurukus, comme les français pourraient avoir traité la structure dans le sens où nous l'avons manipulé, mais ils pourraient aussi avoir noté des différences acoustiques de plus

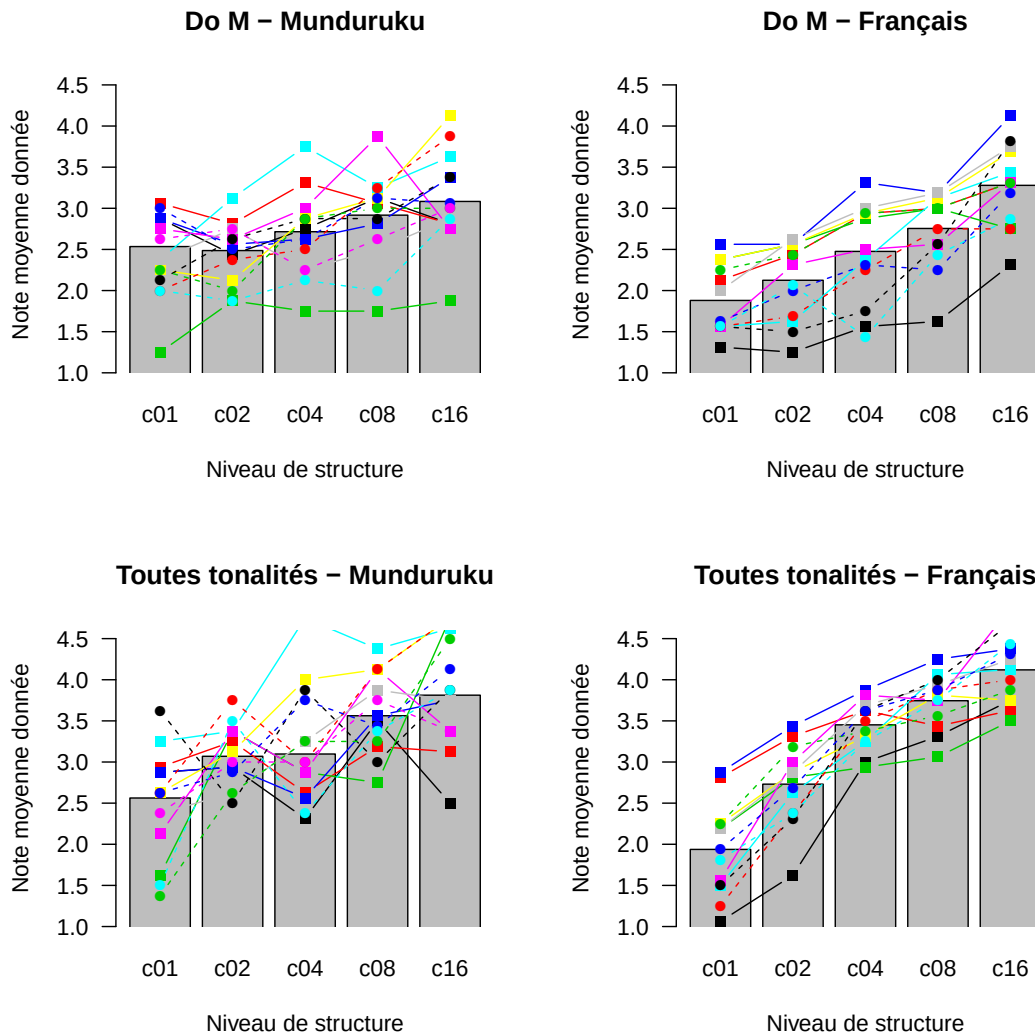


FIGURE 4.13 – Moyennes des notes données selon les groupes et la tonalité des fragments

bas niveau rendant les frontières entre fragments perceptibles ; ils auraient alors évalué le nombre de frontières qui corrèle avec le niveau de structure. Cette perception et notation du nombre de frontières, malgré une absence de différence de tempo des fragments, une prise en compte la tonalité, ne peut pas être exclue, même si l'entretien après la passation des sujets français montre qu'ils (et encore pas tous les sujets) n'ont eu conscience que de 3 niveaux : un morceau entier, 2 moitiés de morceaux et des morceaux moins cohérents. Les sujets sont capables de noter explicitement la structure des stimuli de musique classique occidentale. Cependant, rien ne prouve que ce traitement de la structure ne

soit pas induit par la tâche en attirant l'attention sur celle-ci. Il est envisageable que la musique n'entraîne pas de calculs syntaxiques de façon automatique même s'il est possible de les effectuer en attirant l'attention de l'auditeur. Une tâche plus indirecte, n'attirant pas l'attention du sujet sur la structure, nous permettrait de vérifier que le traitement de la structure syntaxique en musique est bien automatique et effectué en temps réel par les sujets.

4.2.2 Expérience 7 : Traitement automatique, discrimination des stimuli musicaux

L'expérience précédente demandait explicitement de noter la structure de la musique. Dans cette expérience, nous examinons si le niveau de structure a un impact sur la capacité de mémorisation des mélodies. Nous demandons aux sujets de comparer 2 mélodies successives qui peuvent être soit identiques soit différentes mais toujours du même niveau de structure. Nous nous attendons à ce que les mélodies les plus structurées soient plus faciles à encoder (mémoriser) que les mélodies les moins structurées ?? page ?? . Cette expérience fournit un test implicite de la perception de la structure.

Matériel et méthodes

Participants

15 Mundurucus, âgés de 12 à 64 ans, moyenne d'âge de 33.0 (sexe des sujets non renseigné) ont passé l'expérience. Trois d'entre eux ont passé 2 fois l'expérience (160 essais), 1 sujet n'a passé qu'une seule partie de l'expérience (40 essais) et 12 ont passé l'expérience en entier, 2 sessions soit 80 essais. Nous n'avons exclu aucune donnée. La plupart des sujets de cette expérience avait passé précédemment la tâche de notation précédente. Les mêmes 13 sujets français non musiciens ont servi pour tester implicitement la perception de la structure chez les français.

Stimuli

Les mêmes stimuli que l'expérience précédente de notation ont été utilisées dans cette expérience. L'expérience a été découpée en 2 parties pour que les sujets Mundurukus puissent la passer entièrement, en faisant une pause au milieu. Les stimuli ont été appariés par niveau de structure c'est-à-dire que 2 stimuli présentés successivement pouvait être soit être identiques soit différents et dans ce cas ils appartenaient au même niveau de structure.

Procédure

Les sujets Mundurukus ont d'abord été entraînés jusqu'à ce qu'ils comprennent la tâche de comparaison : les 2 extraits musicaux sont-ils semblables ou différents. L'entraînement consistait, comme le test, en une présentation successive de 2 stimuli musicaux appartenant au même niveau de structure. Le programme attendait que le sujet réponde si les 2 extraits étaient identiques ou différents. C'est l'expérimentateur sur place qui pressait la touche correspondant à la réponse donnée par le sujet et qui s'assurait de la compréhension de la tâche. Pendant l'entraînement, le programme donnait un retour sur la réponse fournie et un sourire s'affichait si la réponse était correcte, ou un visage triste si la réponse était incorrecte. Pendant le test, les sujets ne recevaient plus d'information sur la justesse de leur réponse. L'entraînement contenait 6 paires d'extraits musicaux et revenait au début tant que la tâche n'était pas comprise.

Résultats et Discussion

La figure [4.14 page suivante](#) montre les pourcentages de réponse correcte et les *d-prime* (ou indice de sensibilité de la mesure, voir annexe [C page 239](#) pour une explication plus détaillée de ce qu'est le *d-prime*) associés en séparant les français des mundurukus. La mesure des pourcentages de réponse correcte nous montre que la tâche a correctement

été effectuée.^{3 4}

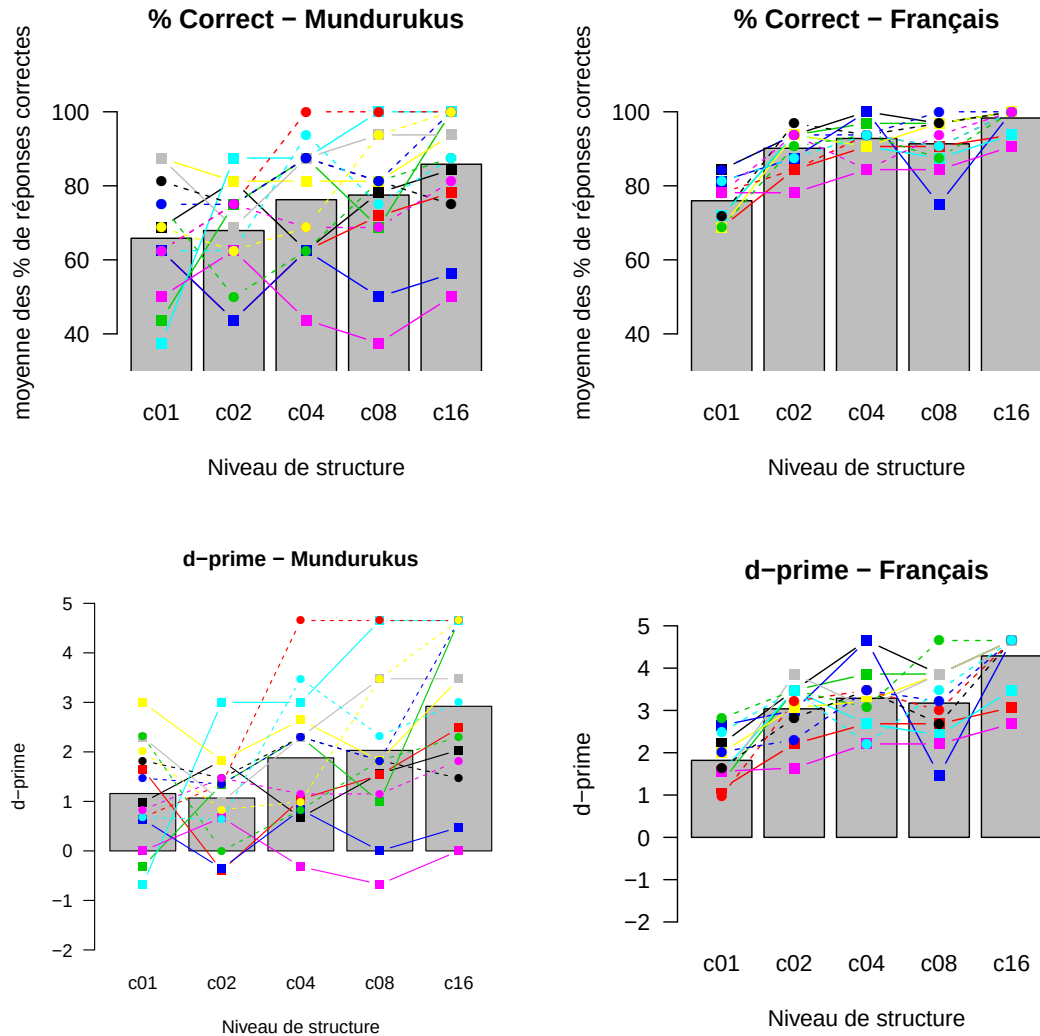


FIGURE 4.14 – Moyenne des réponses correctes (en haut) et d’prime associés (en bas) en séparant français et mundurukus. Les lignes correspondent aux moyennes individuelles.

Plus le niveau de structure est élevé, plus le taux de bonnes réponses est important. Plus les stimuli sont structurés plus la valeur du d-prime est importante, donc plus l’effet recherché est grand. Plus la réponse de nos sujets est claire. Une analyse de variance a

3. Si un individu répond au hasard, il a 50% de chance de donner une réponse correcte.

4. En regardant les données on s’aperçoit que 2 des sujets mundurukus (Akay Samuel et Paygo Aginaldo) répondent au hasard quel que soit le niveau de structure. On gardera tout de même ces sujets dans l’analyse dans la mesure où celle-ci n’est quasiment pas modifiée par ces 2 sujets supplémentaires. La tâche a donc globalement été bien comprise.

été appliquée au d-prime en déclarant comme facteur intrasujet la structure (5 niveaux) et la tonalité (fragments en Do M ou fragments dans des tonalités différentes), et comme facteur intersujet le groupe (mundurukus ou français).

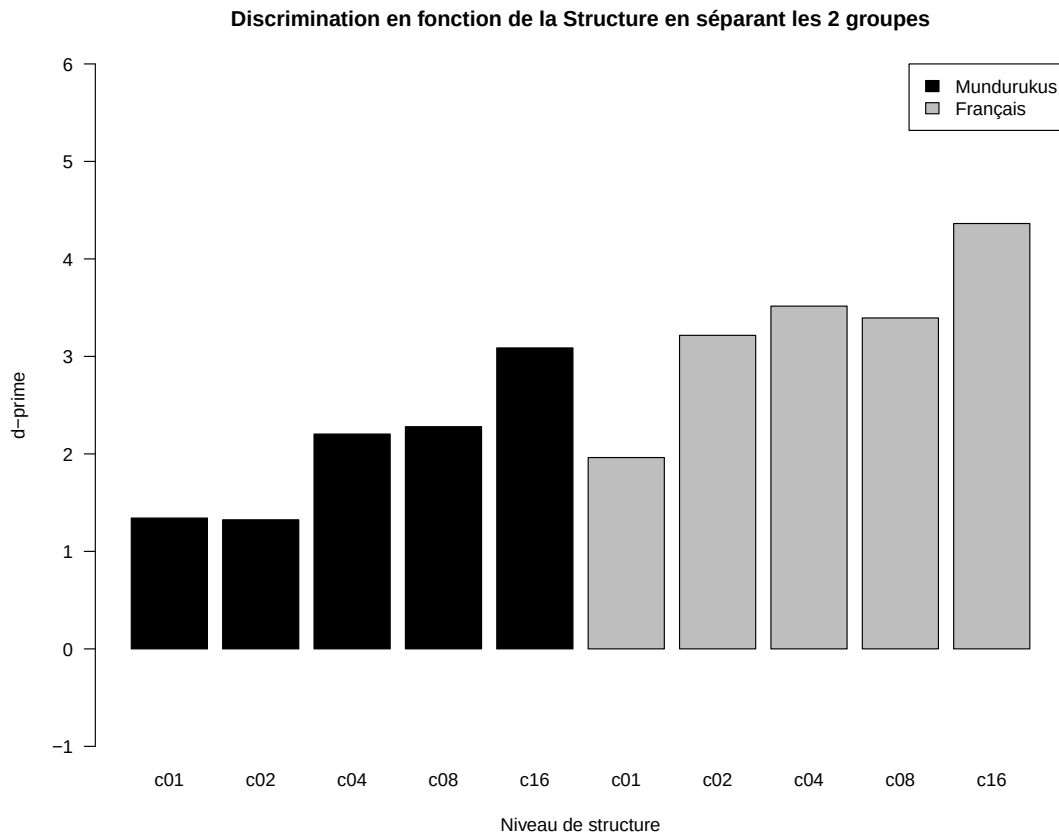


FIGURE 4.15 – Effet du groupe sur la discrimination.

Tous les facteurs principaux sont significatifs. L'effet de structure est le plus fort $F(4, 104) = 20, p < 10^{-12}$, les sujets répondent donc différemment aux différents niveau de structure. L'effet significatif de groupe, $F(1, 26) = 17, p < 3.10^{-4}$, nous apprend que les 2 groupes de sujet ont réalisé de façon différente la tâche, ce qui est visible sur les graphes, les mundurukus ayant fait plus d'erreur que les français (figure 4.15). L'effet de tonalité, $F(1, 26) = 8, p < 0.01$ nous apprend que les sujets n'ont pas traité de la même façon les stimuli qu'ils soient tous en Do M ou contenant des fragments de tonalités différentes (figure 4.16 page suivante). De façon intéressante, aucune des interactions de

facteurs n'est significative, notre facteur structure a l'air d'être traité de la même façon par les 2 groupes de sujet.

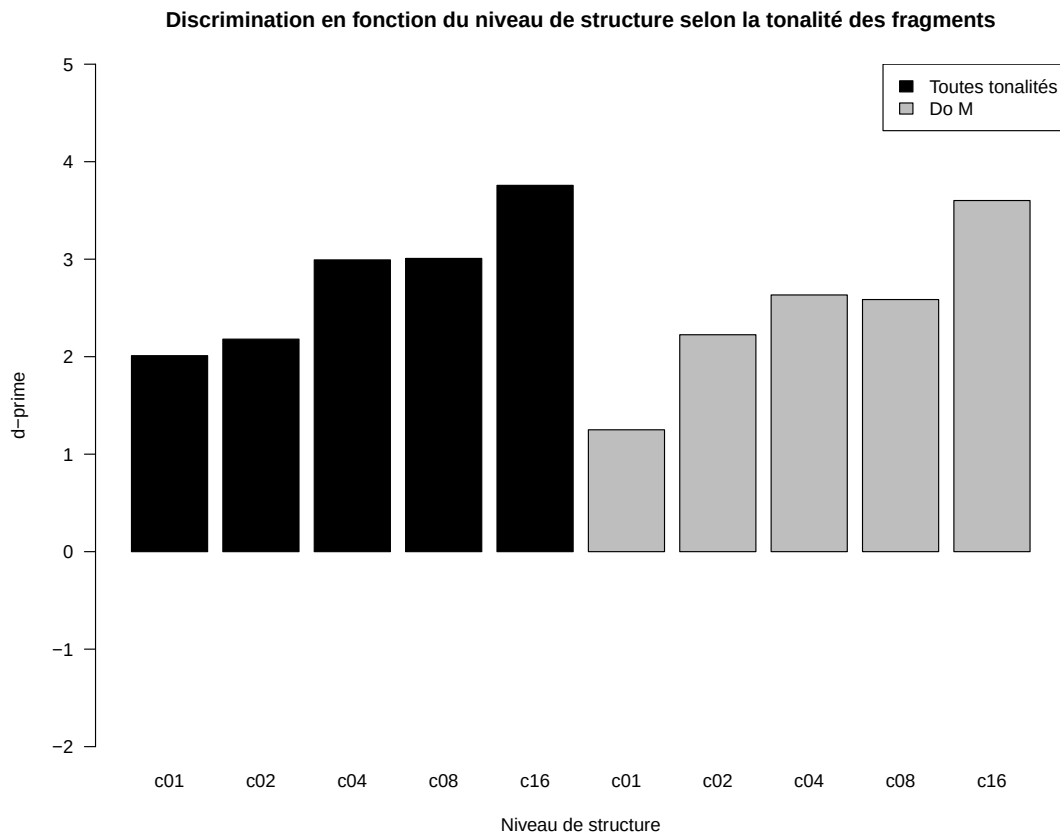


FIGURE 4.16 – Effet de la tonalité sur la discrimination.

Comme l'effet de groupe est significatif, nous allons étudier l'effet de la structure ainsi que l'influence de la tonalité dans chacun de nos groupes séparément (figure 4.2.2 page 144).

L'effet de structure est très fort chez chacune des 2 populations testées, pour les Mundurucus $F(4, 56) = 7, p < 0.0001$ et pour les français $F(4, 48) = 29, p < 10^{-12}$. Les 2 effets vont dans le même sens, plus les stimuli sont structurés, plus le d-prime est grand, plus les réponses sont bonnes. L'effet de tonalité, quant à lui est marginal dans chaque groupe, Mundurucus : $F(1, 14) = 3.8, p < 0.07$ et Français : $F(1, 12) = 4.1, p < 0.06$. Visuellement, on aurait tendance à croire que les changements de tonalité entre les

fragments ont aidé à la discrimination des morceaux. De nouveau, l'interaction entre les facteurs structure et tonalité n'est significative dans aucun des groupes. On voit moins l'influence du changement de tonalité sur la perception de la structure dans cette tâche plus indirecte.

Grâce aux temps de réaction récoltés pour les sujets français, il est possible de regarder la temporalité du traitement syntaxique. La structure a aussi un effet sur les temps de réaction (figure 4.18 page 145) ; une anova a été réalisée sur le log des temps de réaction en déclarant les facteurs intrasujet structure et tonalité, (effet de structure : $F(4, 40) = 35, p < 10^{-12}$). Il est plus rapide de comparer 2 éléments structurés que 2 éléments non structurés. La tonalité n'a pas d'effet significatif sur les temps de réaction ⁵. Par contre, il est intéressant de remarquer que même pour les 3 niveaux intermédiaires c08, c04 et C02, là où les différences sont le moins net dans les performances, l'effet de structure est très significatif sur les temps de réaction ($F(2, 19) = 16, p < 10^{-5}$).

La tâche de discrimination a donc été bien réalisée dans l'ensemble y compris les Mundurucus (seuls 2 participants de ce groupe sont au niveau du hasard). Elle nous permet de mettre en évidence le traitement automatique de la structure musicale de nos stimuli puisque ce paramètre a un effet dans une tâche qui n'attire pas l'attention sur la structure. Comme nous nous y attendions, plus les stimuli sont structurés plus il est plus facile de les encoder en mémoire en vue de les comparer. L'absence d'indices de tonalité aux frontières entre les fragments semble avoir eu un léger impact sur cette tâche, rendant un peu plus difficile la comparaison des niveaux les moins structurés puisque cet effet est marginal chez chacun des groupes.

5. $F(1, 8) = 2.3, p < 0.1$

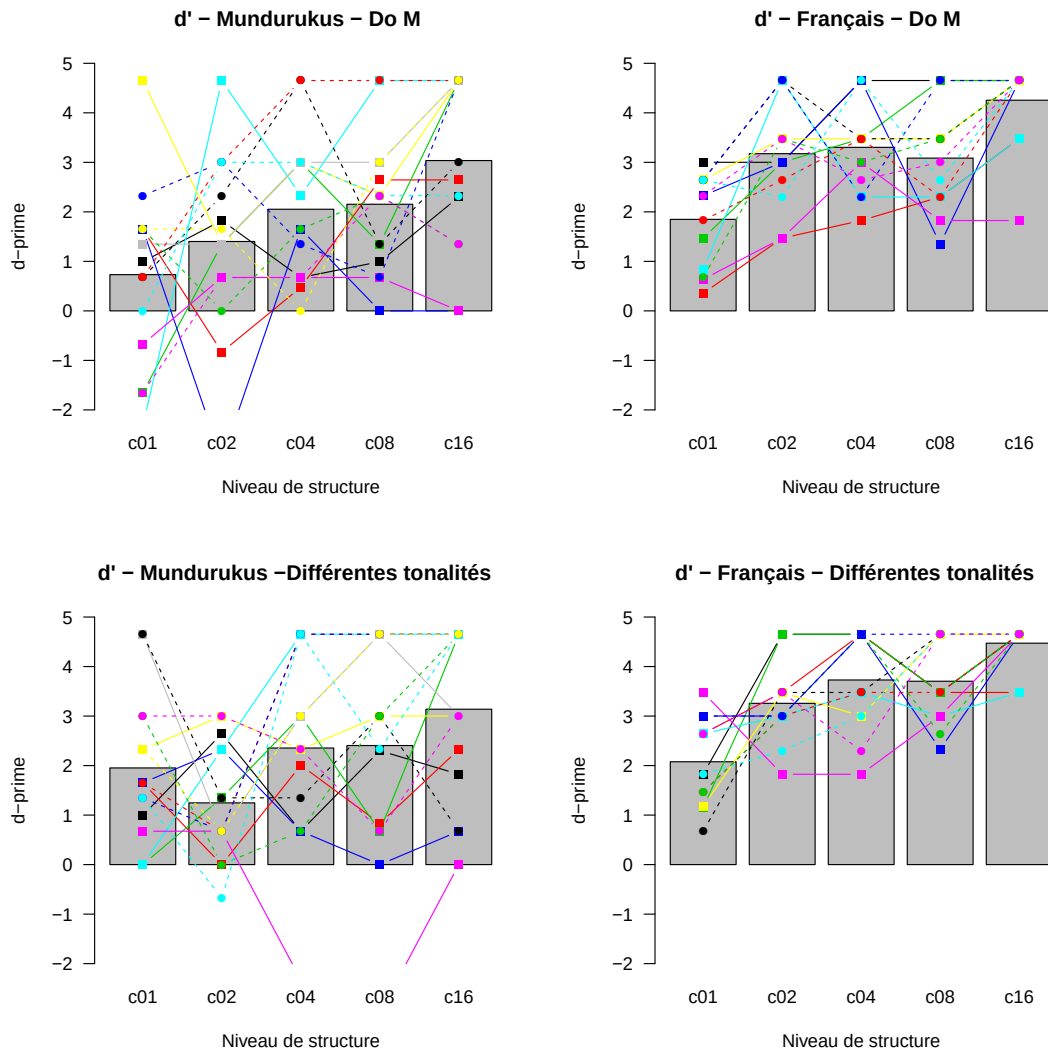


FIGURE 4.17 – Moyennes des d-prime calculés en séparant par groupes et par tonalité des fragments

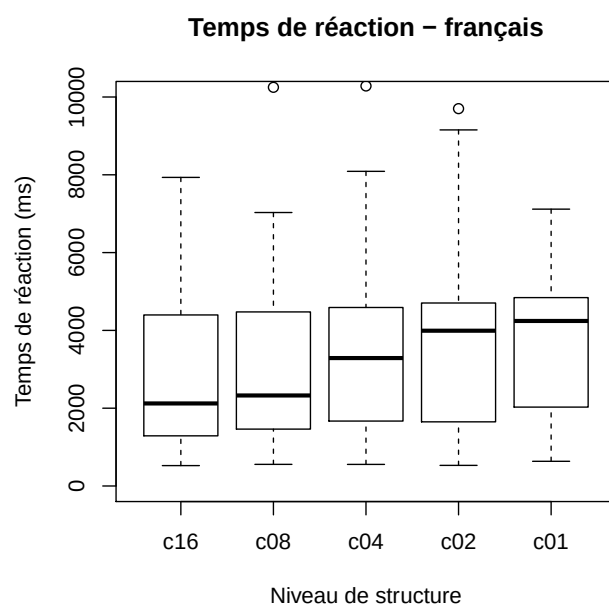


FIGURE 4.18 – Effet de la structure sur les temps de réaction

4.2.3 Discussion générale

Ces 2 expériences inter-culturelles présentaient des difficultés importantes. D'une part, il a fallu trouver des paradigmes réalisables dans les conditions particulières de village dans la jungle. D'autre part, étant donné le rôle de la musique dans la vie quotidienne des Mundurukus, il était envisageable qu'ils ne reconnaissent pas la musique classique occidentale comme de la musique, qu'ils la trouvent désagréable, ne comprennent pas l'intérêt des tests et refusent de les passer. Finalement, même si les participants ont trouvé Mozart très beau et l'ont comparé au chant des oiseaux, ils n'ont aucune envie d'écouter cette musique en tant que loisir. Malgré cette différence de la place de la musique dans la vie quotidienne entre les occidentaux et les Mundurukus, les 2 tâches ont été réalisées correctement par les 2 groupes de sujets. Les données recueillies pour les mundurukus sont plus bruitées que celles des français. Ceci est à mettre en relation avec :

- les conditions de passation de l'expérience (dans une salle commune dans la jungle et non seul dans une cabine de test insonorisée) ;
- l'effet de culture puisque les mundurukus trouvent tout de même étrange ce genre de tests ;
- l'effet de langue puisque les instructions ont dues être traduites par un locuteur non natif et adaptées au cas par cas pour être sûr que les sujets mundurukus aient bien compris.

Ces 2 expériences nous permettent de montrer que les sujets, qu'ils soient de culture occidentale ou non, sont capables de traiter automatiquement (expérience de discrimination) et consciemment (expérience de notation) la structure de stimuli musicaux issus de musique classique occidentale. Plus le niveau de structure est élevé, plus il est facile de mettre en mémoire et de comparer les extraits musicaux, et plus la note donnée pour évaluer la structure est élevée. Notre manipulation des stimuli musicaux est effectivement perçue, même si elle est n'est pas spécifique d'un type de structure syntaxique en musique puisque nous les avons presque toutes détruites (structure de grouping, struc-

ture mélodie, harmonique ...) ⁶. La perception et le traitement de la structure ne sont pas seulement dus à la présence d'un changement de tonalité entre les fragments, même si ce paramètre semble aider. Cependant, nous ne pouvons pas, grâce à ces expériences, préciser quel(s) aspect(s) de la structure sont responsables du changement de perception et de traitement de nos stimuli. Il est possible d'imaginer des stimuli musicaux dont la structure ne varierait que sur un seul ou 2 paramètres orthogonalement. Cela nous permettrait de tester l'influence et l'innéité de ce paramètre sur la construction d'une structure syntaxique en musique. L'avantage de tels stimuli musicaux est qu'ils peuvent être utilisés sur des populations diverses (enfants, adultes de culture différentes ou même des animaux) pour en apprendre plus sur le traitement de la syntaxe en général.

6. Il faudrait également tester s'il existe des différences acoustiques de bas niveau, pour être sûr que notre effet ne s'explique pas seulement par ces différences.

Résumé du chapitre

- Nous avons montré que des sujets français non musiciens réussissaient aussi bien à noter la structure de nos stimuli musicaux et langagiers (notation corrélée au niveau de structure). Nous avons donc validé la construction empirique de nos stimuli musicaux. Nous pourrions donc les utiliser dans l'expérience suivante en IRMf.
- Nous avons également mis en évidence que les français et les mundurukus, tribus amazonienne peu voir pas confronté à de la musique classique occidentale percevaient également la structure de nos stimuli de façon consciente et automatique grâce à 2 expériences comportementales :
- Perception explicite grâce à une notation de la cohérence des stimuli en présentant un visage plus ou moins souriant comme méthode de notation.
- Perception implicite et donc traitement en temps réel de la structure en musique grâce à une expérience de comparaison de 2 éléments musicaux partageant le degré de structure. La mise en mémoire est différente selon les niveaux de structure. Il est plus difficile d'encoder et de comparer des éléments plus déstructurés que des éléments structurés.
- Enfin, nous avons montré que cette perception de la structure en musique n'était pas seulement lié à la tonalité dans nos 2 populations. Il existe donc d'autres indices qui sont responsables de la perception de la structure en musique.

CorrélatS cérébraux du traitement des structures hiérarchiques en langage et en musique

Sommaire

5.1	Etat de l'art en IRMf	149
5.2	Expérience 8 : Bases neurales du traitement syntaxique en lan- gage et en musique chez les musiciens et non musiciens	154
5.3	Conclusion	203

5.1 Etat de l'art en IRMf

L'expérience en lecture ([Pallier et al., 2011](#)) a isolé un réseau qui réagit à la taille des constituants en langage. La figure [5.1 page suivante](#) montre que ce réseau peut être décomposé en 2 types d'aires cérébrales :

- celles qui répondent à la structure indépendamment du sens (en bleu sur la figure) : IFG tri, IFG orb et pSTS, actives en langage normal et en Jabberwocky (pas de représentation sémantique de la phrase possible) ;
- celles dont l'activité est modulée par la taille des constituants uniquement en

présence de sémantique (langage normal, en rouge sur la figure) : TP, aSTS, TpJ, putamen.

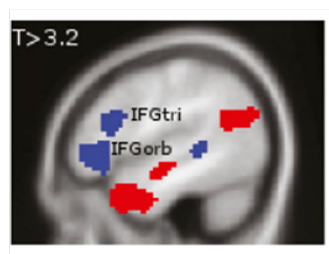


FIGURE 5.1 – Régions qui s’activent de plus en plus à mesure que la taille des constituants augmente, expérience à l’écrit. Les régions en bleu détectent la taille des constituants aussi bien avec des pseudo-mots (jabberwocky) qu’avec des mots, les régions en rouge réagissent uniquement aux vrais mots et ont donc besoin d’un contenu lexical, analyse de groupe, $T > 3.2$, $p < 0.001$ seuil non corrigé. Extrait de (Pallier et al., 2011).

L’existence de zones qui réagissent à la structure indépendamment du sens suggère que ces aires cérébrales pourraient effectuer des calculs syntaxiques de façon non spécifique au langage. Récemment, Rogalsky et ses collègues (Rogalsky et al., 2011) ont comparé le traitement du langage délexicalisé (jabberwocky) et de la musique chez les mêmes sujets. Alors qu’ils étaient dans l’IRM, les sujets entendaient soit des phrases en Jabberwocky préservant la structure syntaxique, soit des phrases mélangées en Jabberwocky (sans structure syntaxique), soit des mélodies inventées. La figure 5.2 page suivante compare les activités cérébrales évoquées par les phrases structurées avec celles des phrases sans structure syntaxique dans lesquelles l’ordre des mots avait été altéré. Cette carte cérébrale montre que le réseau comprenant pour l’hémisphère gauche, deux clusters du lobe temporal antérieur, une région postérieure du gyrus médian temporal, une partie du gyrus frontal inférieur ; pour l’hémisphère droit, une partie du lobe temporal antérieur, du gyrus frontal inférieur et du gyrus postcentral, voit son activité modulée par la dégradation de la structure syntaxique.

Ce réseau n’est pas exactement identique à celui mis en évidence par (Pallier et al., 2011) mais présente des similitudes comme par exemple la zone inféro-frontale gauche et les régions temporelles antérieures. Les différences entre ces deux études peuvent s’ex-

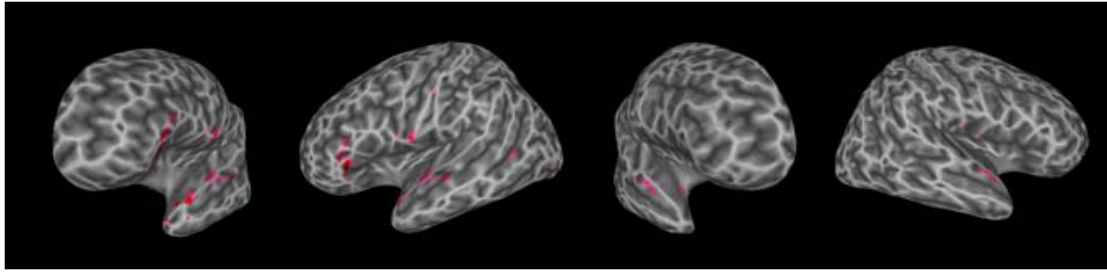


FIGURE 5.2 – Voxels plus activés pour les phrases bien construites par rapport aux phrases mélangées, $p < 0.005$. (extrait du matériel supplémentaire de (Rogalsky et al., 2011)).

pliquer d'une part par l'absence de tâche dans (Rogalsky et al., 2011) qui ne forçait pas l'attention des sujets sur les stimuli au contraire de (Pallier et al., 2011) où les sujets devaient appuyer sur un bouton lorsque la phrase lue le leur demandait¹. La différence d'attention peut se traduire par une diminution des activités cérébrales associées au traitement des stimuli. D'autre part, le contraste observé dans l'expérience de (Pallier et al., 2011) est plus fin que celui présenté dans le Rogalsky puisque la taille des constituants varie de façon paramétrique. L'analyse du traitement des stimuli musicaux dans (Rogalsky et al., 2011) n'est présentée que dans l'IFG et le pôle temporal gauche (figure 5.3) sans rapporter les résultats dans les autres régions cérébrales. L'analyse pour

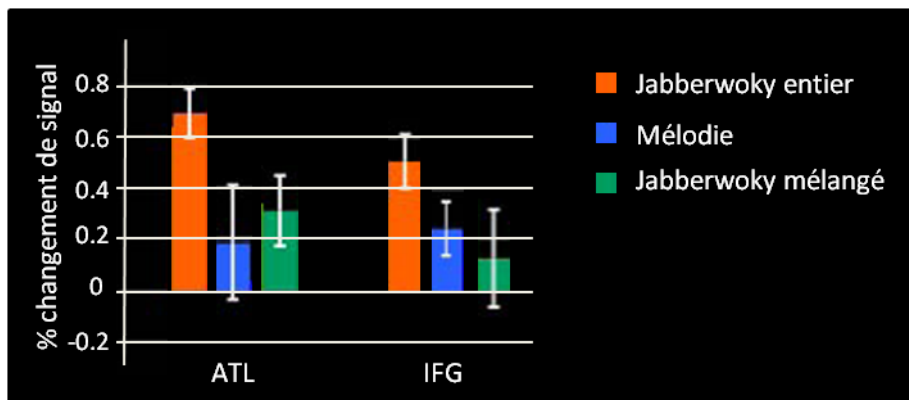


FIGURE 5.3 – Amplitude moyenne pour chaque type de stimulus pour la région d'intérêt du pôle temporal gauche (ATL) et du gyrus fronto-latéral inférieur gauche (IFG). Adapté de (Rogalsky et al., 2011).

1. cette tâche permettait de s'assurer que les sujets restaient attentif tout au long de l'expérience

tester le partage des ressources neurales par les deux domaines a été effectuée dans les régions où les activations pour la musique (par rapport au repos) et pour le langage bien construit (par rapport au repos) se chevauchaient. Dans ces zones, une analyse par classification de pattern multivariés (multivariate pattern classification analysis) montre que les deux domaines provoquaient des activations distinctes puisqu'il est possible de classer, de façon intrasujet, les essais provenant de l'écoute de langage par rapport à ceux de la musique. Les auteurs concluent de leur étude que le traitement du langage et de la musique sont effectués par des réseaux distincts. Cependant, dans cette étude, l'effet de la structure en musique n'est pas testé clairement puisque seules des mélodies intactes ont été présentées. Les activations cérébrales provoquées par l'écoute de mélodie reflètent un ensemble de traitements computationnels (analyse des hauteurs des notes, de leurs relations, du tempo, du rythme, du groupement des notes, émotions etc.), elles ne traduisent donc pas directement le traitement de la structure en musique. Dans une autre étude en IRMf, Levitin et Menon ([Levitin and Menon, 2003](#)) ont manipulé directement la structure en musique en modifiant la cohérence temporelle d'un morceau. Ils ont enregistré les réponses cérébrales à l'écoute passive de morceaux de musique avec une structure temporelle cohérente ou incohérente (fragments de 250ms à 300ms recombinaison et concaténés pour former de nouveaux morceaux déstructurés). L'utilisation de tels stimuli permet de ne détecter que des régions sensibles à la structure musicale en contrôlant les indices acoustiques de bas niveau (volume sonore, hauteur de note, timbre). Un des problèmes de ces stimuli est que la méthode de création des stimuli provoque des artefacts auditifs (qui sont corrigés dans une expérience suivante ([Abrams et al., 2011](#)) qui sera détaillée dans la discussion). La figure 5.4 page suivante montre le réseau réagissant à la structure en musique. Parmi les structures cérébrales isolées par cette étude, on trouve la partie orbitalis du cortex inféro-frontal (BA47) à gauche ainsi que l'insula antérieure adjacente et leurs homologues de l'hémisphère droit ; mais aussi, le cortex cingulaire antérieur, le noyau accumbens, le tronc cérébral et le vermis postérieur. Les auteurs proposent que l'aire de Brodmann 47 et l'insula adjacente soient

le siège de l'organisation des unités structurales indépendamment de la modalité étant donné que ces structures ont été retrouvées dans le traitement de la structure en langage dans d'autres études (Snijders et al., 2009; Pallier et al., 2011; Rogalsky et al., 2011). Cependant, cette étude ne teste pas chez les mêmes sujets des stimuli comparables en langage et en musique qui seraient plus ou moins structurés, contrairement à notre étude.

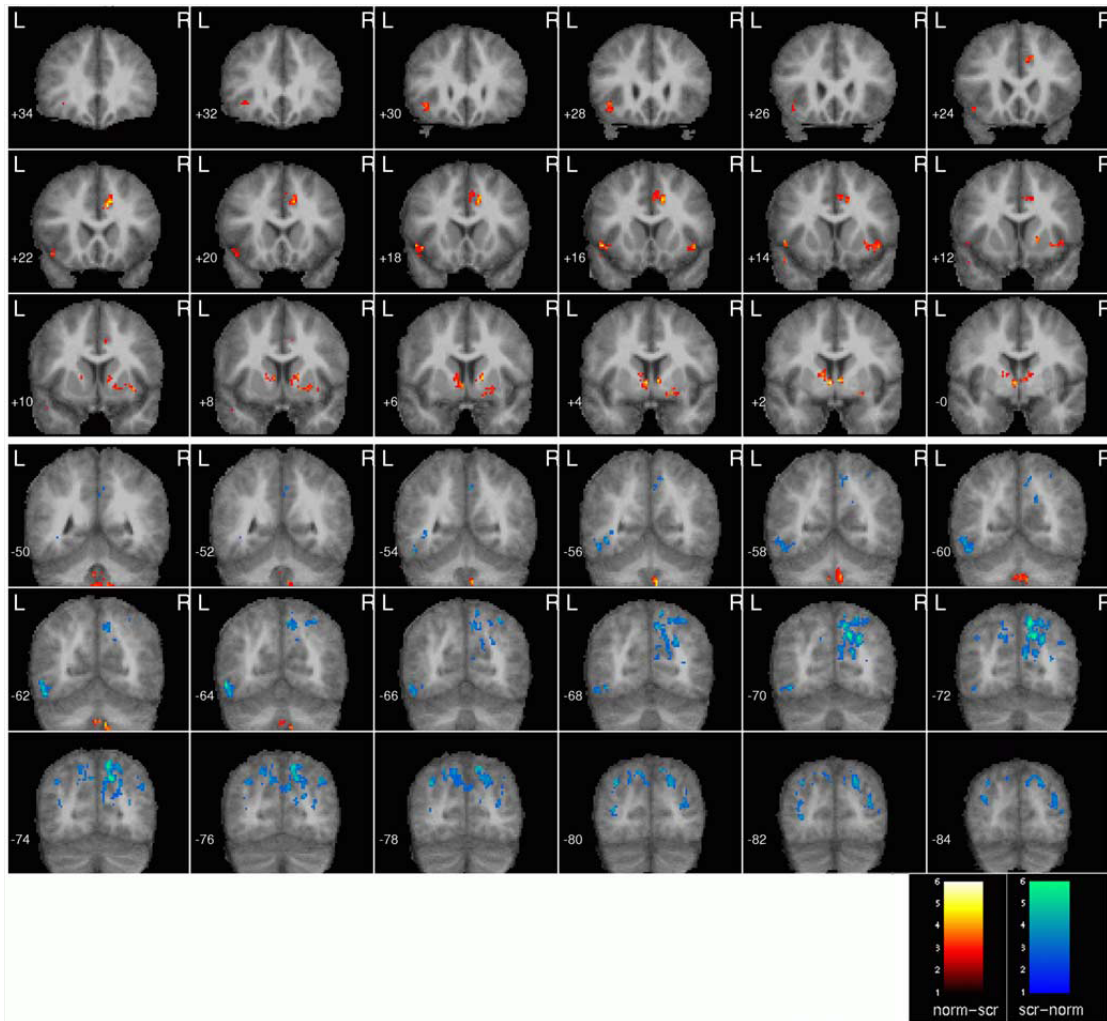


FIGURE 5.4 – Sections coronales montrant le signal BOLD, activations et désactivations, comparant la musique bien construite et la musique mélangée. Dans chaque hémisphère, l'activation de la région orbitalis du cortex inféro-frontal BA47 a été observée même si les activations dans l'hémisphère gauche sont plus larges. Extrait de (Levitin and Menon, 2003).

5.2 Expérience 8 : Bases neurales du traitement syntaxique en langage et en musique chez les musiciens et non musiciens

Nous proposons une expérience en IRMf comparant de façon intra-sujets (musiciens et non-musiciens) la structure en langage oral et en musique. Notre but est de déterminer la spécificité des zones cérébrales réagissant à la structure syntaxique linguistique, particulièrement celles effectuant des calculs syntaxiques de combinaison d'éléments (merge) trouvées dans (Pallier et al., 2011). Notre paradigme présente l'avantage d'étudier de façon paramétrique ces calculs en augmentant le nombre d'opérations nécessaires à la construction de la structure syntaxique. Grâce aux expériences comportementales précédentes, nous avons validé nos stimuli langagiers et musicaux créés de façon empirique puisque les sujets sont sensibles à leur déstructuration progressive.

Matériel et méthodes

Participants

Trente-cinq sujets sains droitiers de langue maternelle française (âgé de 18 à 30 ans, moyenne = 22.79, 17 femmes et 18 hommes) parmi lesquels 18 non-musiciens (âgés de 18 à 29 ans, moyenne de 22.17, 6 femmes et 12 hommes) et 17 musiciens (âgés de 18 à 30 ans, moyenne de 23.59, 11 femmes et 6 hommes). Pour être inclus dans le groupe musicien, les sujets devaient avoir commencé la musique avant 7 ans, ne jamais avoir arrêté de pratiquer et jouer leur instrument au moins cinq heures par semaine. Pour le groupe des non-musiciens, les sujets ne devaient pas avoir reçu d'autres cours de musique que ceux dispensés au collège entre la classe de 6^{ème} et de 3^{ème}. Aucun des participants n'a signalé de déficit auditif, ni de trouble neurologique. Tous ont signé un consentement éclairé et reçu 80 euros en échange de leur participation au protocole. L'expérience a été approuvée par le comité éthique du Commissariat à l'Energie Atomique.

Stimuli

Les mêmes stimuli que l’expérience comportementale ont été utilisés (pour leur construction voir 4.1 page 120). La figure 5.5 et la table 5.1 donnent un exemple des stimuli des cinq niveaux de structure dans chaque domaine.

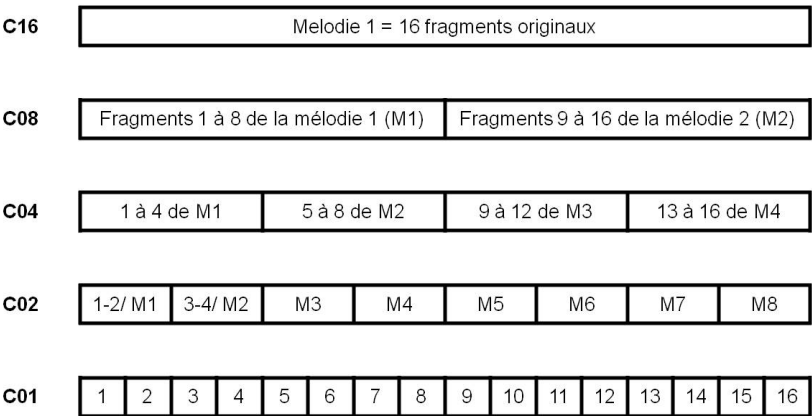


FIGURE 5.5 – Schéma des différents niveaux de structure

C16	Mon père parle souvent de choses très savantes qui fascinent les copains qui viennent chez moi															
C08	Paul mange la tarte sur laquelle se trouvait								qui fascinent les copains qui viennent chez moi							
C04	Il lance la balle				les grimaces des clowns				la biche sort vite				qui sent les fleurs			
C02	Il manque		est tombé		les fleurs		avec soin		très raide		du Brésil		le blé		les femmes	
C01	La	à	sortent	tarte	le	un	grimpe	Luc	les	qui	siècle	pains	qui	de	réchauffe	jaunes

TABLE 5.1 – Exemple de phrases pour chaque niveau de structure

En sus de ceux-ci, vingt autres stimuli ont été construits pour constituer les cibles de la tâche à réaliser dans le scanner. Une phrase correspondant à chaque niveau de structure pour chaque domaine a ainsi été construite. Dans les cibles musicales, un fragment de un temps était joué au clavecin au lieu du piano. Pour les cibles linguistiques, un mot était prononcé par une voix masculine au lieu de la voix féminine habituelle. Pour l’entraînement, réalisé hors du scanner, six autres phrases en français ont été créées suivant le même schéma de construction que les autres stimuli pour donner six stimuli de quatre niveaux de structure différents. Six stimuli musicaux destinés à l’entraînement ont été choisis parmi l’ensemble des 80 stimuli créés pour l’expérience appartenant à 4 niveaux de structure différents.

Procédure

Les participants ont été entraînés avec 12 stimuli (6 linguistiques et 6 musicaux) hors du scanner afin d'assimiler la tâche. Ils devaient détecter un changement d'instrument (clavecin à la place de piano) pour les stimuli musicaux et un changement de voix (masculine à la place de féminine) pour le langage. Dans le scanner, les sujets entendaient 180 stimuli la moitié musicaux et l'autre moitié linguistique répartis en 4 blocs de 45 essais organisés en deux blocs de stimuli musicaux et deux blocs linguistiques. Chaque bloc contenait 8 stimuli de chaque niveau de structure plus 5 essais cibles insérés de façon aléatoire. Un bloc de musique était suivi par un bloc de langage et vice versa. La moitié des participants ont d'abord entendu un bloc de musique et l'autre moitié un bloc de langage. Chaque stimulus durait 6.5s et la durée entre chaque début de stimulus était de 16s. Nous avons utilisé un paradigme événementiel lent pour avoir accès au décours temporel de la réponse à chaque stimulus. Le temps de réponse des sujets et leur performance ont été enregistrés par le programme de stimulation. Les essais cibles ont été écartés de l'analyse.

Acquisition des données

Les acquisitions ont été réalisées sur un scanner IRM 3 Tesla Siemens Trio. Pour chaque participant, une image anatomique a d'abord été acquise en utilisant une séquence écho gradient 3D (TI = 900 msec, TE = 2.98 msec, TR = 2.3 sec; Taille de Voxel = 1x1x1.1 mm, champ de vue = 256, Matrice = 256x256, angle de bascule=9°, 160 coupes). Ensuite, un total de 1232 scans fonctionnels ont été acquis en 4 sessions de 308 scans chacune, en utilisant une séquence EPI 2D bold (TR = 2.4 sec, TE = 30 msec, angle de bascule = 81°, Matrice = 64x64; Champs de vue = 192; Taille de Voxel = 3x3x3 mm; 40 coupes axiales ont été acquises selon une séquence ascendante pour couvrir le cerveau entier).

Analyses des données

L'analyse des données a été réalisée avec le logiciel SPM5 (Wellcome Department of Cognitive Neurology, www.fil.ion.ucl.ac.uk/spm). Les images anatomiques ont été normalisées en utilisant les paramètres par défaut (transformation non linéaire) sur le template standard défini par le standard neurologique du Montreal Neurological Institute réalisé en moyennant 152 cerveaux. Les volumes fonctionnels ont été corrigés pour les différences de déphasage lors de l'acquisition des coupes par rapport à la première coupe, puis réalignés pour corriger le mouvement, normalisés spatialement en utilisant les paramètres de normalisation issus de l'anatomie, lissés au niveau individuel en utilisant un noyau gaussien de 5*5*5mm.

Les effets expérimentaux à chaque voxel ont été estimés en utilisant un modèle linéaire multi-session modélisant 6 conditions (c01, c02, c04, c08, c16, cible), en spécifiant 2 domaines (musique, langage) et les 6 paramètres de mouvement calculés pendant l'étape de réalignement. Chaque stimulus a été modélisé en tant qu'évènement de durée 6.5 secondes correspondant à la durée de présentation des stimuli. Les régresseurs ont été créés en convoluant ces durées (epochs) par la fonction idéale hémodynamique (fourni par SPM5) et sa dérivée temporelle, un 'filtre' de 128 secondes a été appliqué.

Ces estimations des effets individuels pour chaque condition (10 cartes de beta au total : une par condition de structure par domaine) ont été entrées dans un modèle d'analyse de variance intra-sujet avec un régresseur par groupe et par condition expérimentale ainsi qu'un régresseur par sujet pour s'affranchir des effets principaux liés aux sujets. Les effets principaux ont été modélisés dans une analyse à effet fixe en tenant compte de la variance des sujets.

Pour chercher les régions dans lesquelles les activations augmentent avec les conditions c01, c02, c04, c08 à c16, nous avons appliqué un contraste linéaire (poids = (-2 -1 0 1 2)) pour chaque domaine (musique ou langage), dans chaque groupe de sujet (musicien ou non-musiciens).

Le seuil des résultats présentés ci-dessous correspond à un seuil corrigé FDR $p < 0.05$

sauf quand un autre seuil est précisé. Pour les graphiques issus des régions d'intérêts (ROI), les données ont été extraites en utilisant la toolbox Marsbar ((Brett et al., 2002) ; marsbar.sourceforge.net) qui moyenne les amplitudes sur tous les voxels de la ROI. Les ROI ont été définies comme l'intersection de sphères de 10mm de rayon avec les clusters identifiés par le contraste linéaire du domaine langage à un seuil non corrigé de $p < 0.001$. Nous avons par ailleurs utilisé les 7 ROIs identifiées dans (Pallier et al., 2011) auxquelles a été rajouté la région BA44 définie anatomo-fonctionnellement par (Amunts et al., 2004)² pour mener nos analyses sur les décours temporels, amplitudes et phases.

Résultats

Réseaux Globaux

La figure 5.6 page suivante montre les réseaux activés par l'écoute du langage et de la musique par rapport au repos, sans tenir compte des différents niveaux de structure.

Les régions auditives classiques (régions temporales supérieures bilatérales) sont activées dans ces contrastes d'effets principaux. En langage, le réseau correspond à celui classiquement observé (Mazoyer et al., 1993) auquel s'ajoute des zones temporales, le gyrus frontal inférieur bilatéral, une partie du thalamus droit et l'aire supplémentaire motrice de ce même hémisphère. Les conditions musicales activent une région volumineuse englobant le gyrus frontal inférieur droit, l'aire supplémentaire motrice droite ainsi que le thalamus de façon bilatéral.

2. notamment pour répondre à d'autres études impliquant cette région dans les calculs syntaxiques (par exemple (Makuuchi et al., 2009))

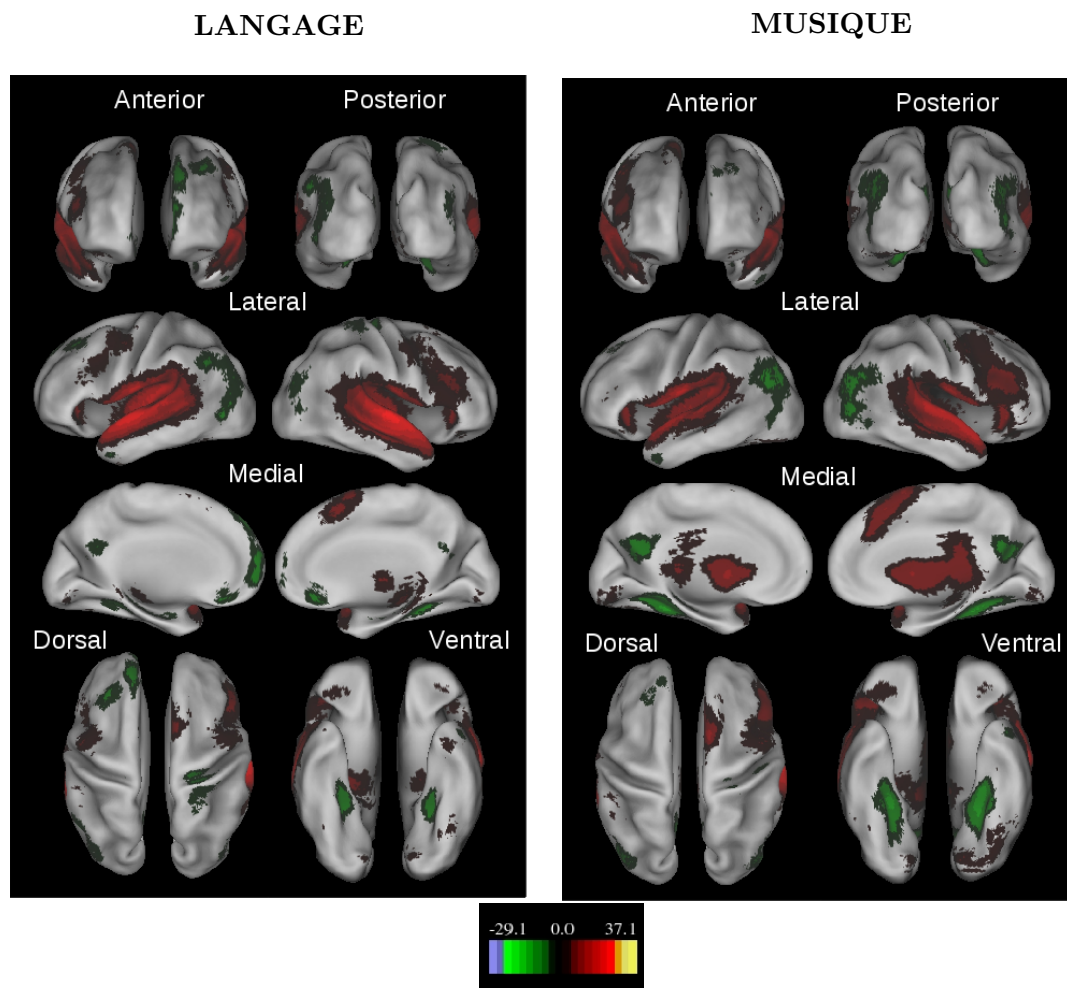


FIGURE 5.6 – Carte statistique associée aux activations globales au langage à gauche et à la musique à droite, toutes conditions expérimentales additionnées. Seuil corrigé fwe $p < 0.05$. En rouge, les régions plus activées par les stimulations auditives par rapport au repos et en vert les régions répondant à l'inverse : plus actives pour le repos que pour les stimulations auditives.

Différence entre langage et musique

La figure 5.7 page suivante montre les zones cérébrales plus activées par le langage par rapport à la musique (à gauche) et inversement celles plus activées par la musique que le langage (à droite) sans distinguer les niveaux de structure au sein des domaines.

Les régions temporales supérieures bilatérales ainsi que le gyrus frontal inférieur gauche sont plus activés par le langage que la musique. Le réseau plus activé par la

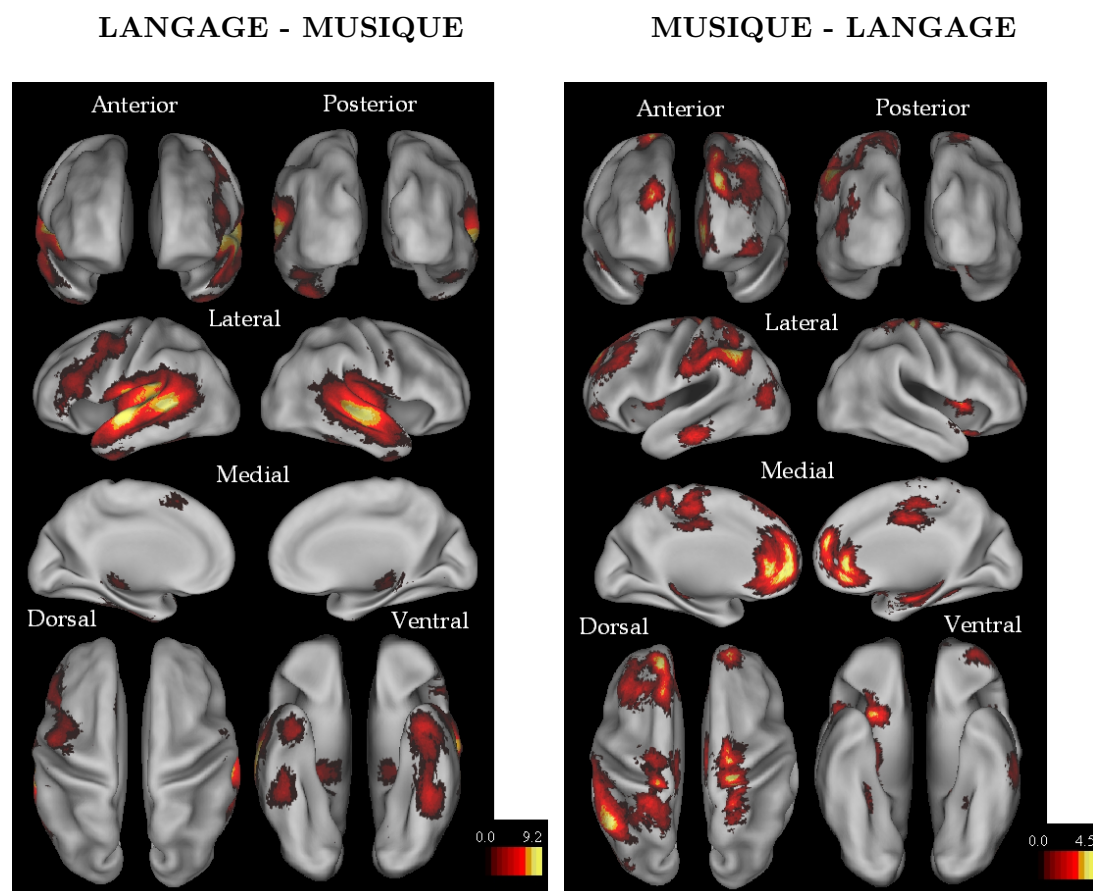


FIGURE 5.7 – Carte d’activation cérébrale étudiant les différences d’activation entre le langage et la musique. Seuil corrigé $FDR < 0.05$. Modèle d’analyse à effet fixe en tenant compte de la variance des sujets.

musique contient entre autres, les aires frontales antérieures médiales bilatérales, des aires pariétales gauches, des régions frontales supérieures bilatérales avec une prévalence de l’hémisphère gauche.

Ces cartes ne montrent pas la part d’activation due à un domaine par rapport à l’autre. La figure 5.8 page 162 complète les cartes 5.7 puisqu’elle sépare les activations dues à chaque domaine. En effet, les zones qui sont plus activées pour le langage par rapport à la musique (contraste langage - musique) peuvent refléter des activations pour le langage ou des désactivations pour la musique. De façon inverse, les zones plus activées pour la musique (contraste musique - langage) reflète aussi bien des activations pour la musique que des désactivations pour le langage. La table 5.2 page suivante précise ces

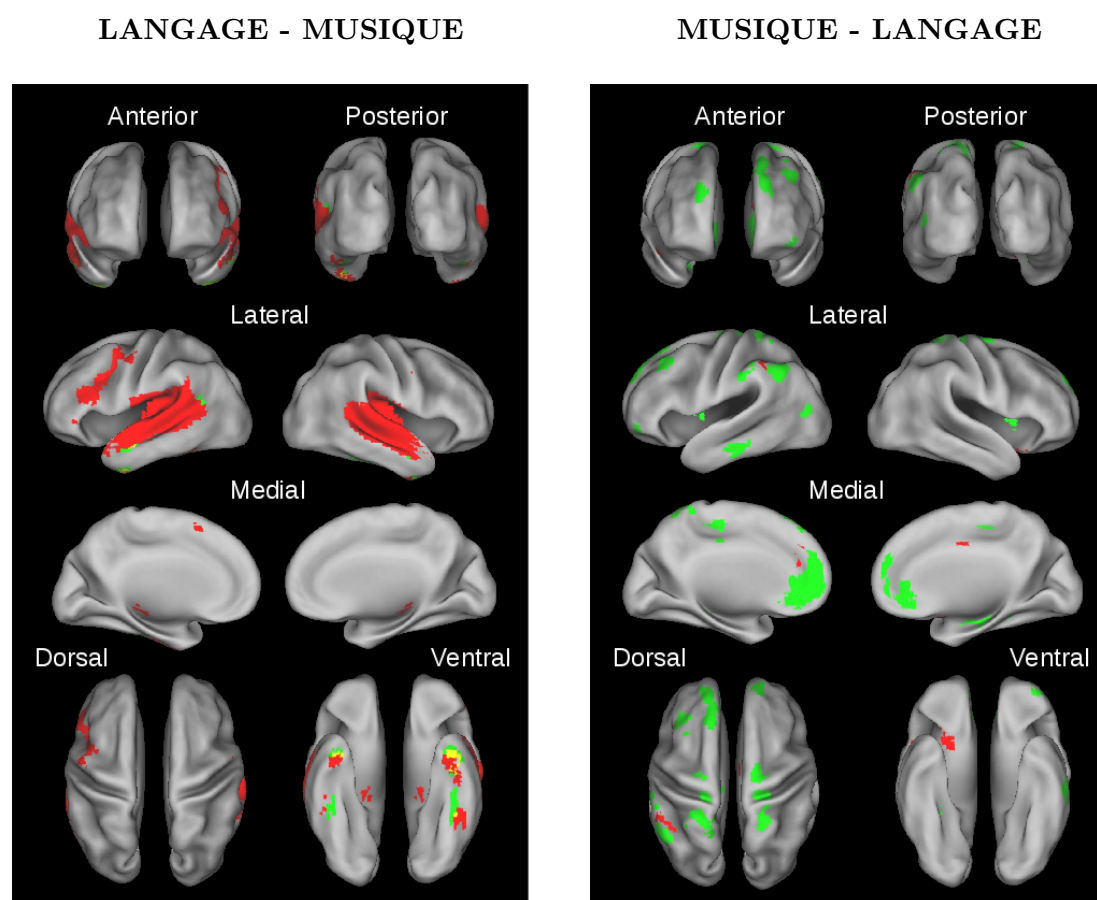
différents cas.

Lang - Mus > 0			Mus - Lang > 0		
Lang > 0		Lang < 0	Mus > 0		Mus < 0
Mus > 0	Mus < 0	Mus < 0	Lang > 0	Lang < 0	Lang < 0
Activation Lang		Déactivation Musique	Activation Mus		Déactivation Lang

TABLE 5.2 – Différents cas de figure pour les différences d’activation entre le langage et la musique. Les couleurs des cellules sont celles visibles sur la figure 5.8 page suivante

La comparaison entre le langage et la musique montre une réponse plus importante au langage. Sur la carte langage - musique, le réseau d’activation du langage (en rouge) est majoritaire dans la comparaison, seules les parties polaires et médiales du sulcus inférieur temporal de façon bilatéral sont le siège d’une déactivation du langage. De la même manière, les déactivations dues au langage (en vert) sont majoritaires sur la carte musique - langage, seules de petites zones du cortex pariétal inférieur gauche, de l’amygdale et insula droites sont activées par la musique .

Globalement le réseau du langage semble être plus activé que le réseau répondant à la musique.



En rouge, les activations dues au langage, en jaune, les zones activées par le langage et déactivées par la musique, en vert les zones déactivées par la musique.

En rouge, les activations dues à la musique, en jaune, les zones activées par la musique et déactivées par le langage, en vert les zones déactivées par le langage.

FIGURE 5.8 – Carte d'activation cérébrale étudiant les différences d'activation entre le langage et la musique. Seuil corrigé $FDR < 0.05$.

Comparaison entre les musiciens et les non-musiciens pour les contrastes globaux

Les analyses suivantes étudient les différences d'activations entre les musiciens et les non-musiciens. En regroupant tous les niveaux de structure des conditions langagières, la comparaison entre les musiciens et les non-musiciens n'est pas significative ni à un seuil corrigé FDR $p < 0.05$, ni à un seuil non corrigé et $p < 0.001$. Les musiciens et les non-musiciens traitent donc globalement nos stimuli langagiers de la même façon.

Lorsque toutes les conditions musicales sont regroupées, la comparaison entre les musiciens et les non-musiciens n'est significative qu'à un seuil non corrigé, $p < 0.001$. La carte 5.9 illustre ces différences.

MUSICIENS - NON-MUSICIENS

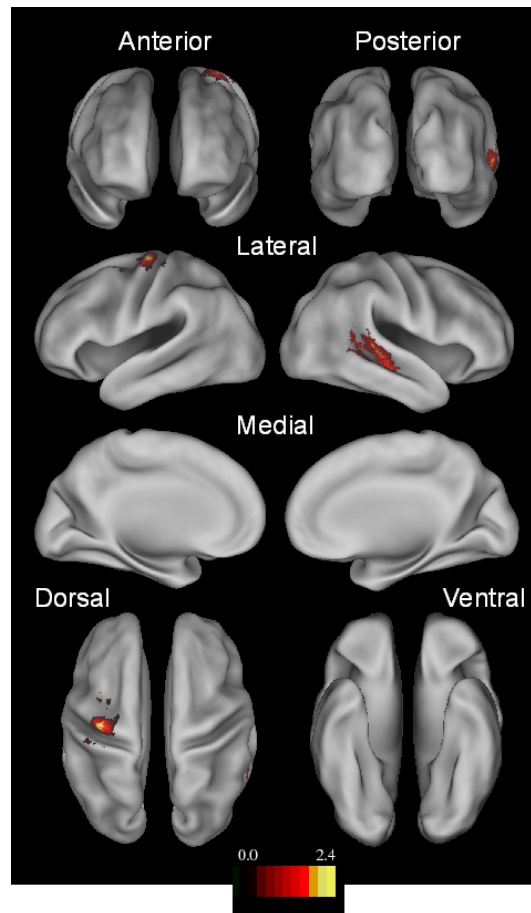


FIGURE 5.9 – Carte statistique relative aux activations plus importantes pour les musiciens que les non-musiciens pour toutes les conditions musicales. Seuil non corrigé, $p < 0.001$

Les musiciens ont des activations supplémentaires par rapport aux non-musiciens, notamment dans le sillon temporal supérieur moyen/postérieur droit, le gyrus precentral gauche (cortex moteur de la main³).

3. Cortex moteur de la main (Sastre-Janer et al., 1998; Caulo et al., 2007). En fait la zone que nous mettons en évidence est à cheval entre BA4 - cortex moteur - et BA6 - l'aire prémotrice -, notre pic d'activation est à -36 -21 63 en MNI et celui du cortex de la main -34 -39 50. On se trouve donc un peu en antérieur et un peu plus bas que les coordonnées précises de l'étude. Cependant il y a une variation assez importante dans la localisation de la zone.

Effet de structure

L'analyse suivante a étudié les régions dont le profil d'activation augmentait ou diminuait avec le degré de structure. L'analyse a été effectuée séparément pour le langage et la musique sans distinguer les musiciens des non-musiciens.

Structure syntaxique en Langage La figure 5.10 page 166 et la table associée 5.3 page ci-contre illustrent les zones cérébrales dont l'activité est modulée par la taille des constituants syntaxiques pour les stimuli linguistiques. L'activation des régions suivantes augmente à mesure que la taille du plus grand constituant augmente (en rouge) : à gauche le pôle temporal, le sillon temporal supérieur antérieur, le gyrus frontal inférieur dans sa partie orbitalis et triangularis (et pas opercularis), la jonction temporo-parietal (gyrus angulaire) et le putamen, également l'hippocampe bilatéral, le pôle temporal droit et le gyrus frontal inférieur orbitalis droit, (IFG droit qui n'a pas été mis en évidence dans l'expérience à l'écrit), des aires frontales médiales orbitales bilatérales et une aire frontale plus médiane supérieure à gauche. Nous répliquons essentiellement les résultats obtenus dans l'expérience en langage écrite de (Pallier et al., 2011) (figure 5.24 page 190).

Notre paradigme révèle un réseau dont l'activité corrèle avec une diminution de la taille des constituants (en vert) qui ressemble essentiellement à celui trouvé dans l'expérience en langage écrit (Pallier et al., 2011). On notera ainsi des aires frontales médiales bilatérales, des régions bilatérales cingulaires antérieures et médianes, le precuneus et une région pariétale inférieure à gauche. En outre, nous mettons en évidence une partie du cervelet, le gyrus frontal inférieur opercularis droit, une région supramarginale gauche.

La figure 5.11 page 166 présente le même contraste d'activation corrélée avec l'augmentation ou la diminution de la taille des constituants suite à une correction statistique par erreur de type 1 sur la famille de comparaisons (family wise error), $p < 0.05$ sur laquelle 5% des voxels dont l'activité était corrélée à la structure n'ont pas été détecté. Dans l'hémisphère gauche, on retrouve le pôle temporal, le gyrus frontal inférieur dans

Zones dont l'activité est modulée par la structure en langage
Analyse au seuil corrigé FDR, $p < 0.05$

Augmentation d'activité avec la taille de la structure	Taille du cluster	T (df=)	x	y	z
Frontal inferior triangularis and Temporal pole left	1227	8.09	-45	15	-27
Temporal mid		7.85	-51	-6	-18
Frontal inferior orbitalis		6.24	-39	30	-15
Temporal Pole right	118	5.58	45	18	-27
Frontal inf orbitalis right	66	4.94	36	33	-15
Temporal mid left and Angular	239	5.11	-48	-57	18
Hippocampus left and Amygdala	92	4.91	-21	-12	-15
Frontal inferior and mid Orbitalis Right	85	4.74	36	33	-15
Temporal mid and sup Right	37	4.45	51	-9	-18
Parahippocampal left and fusiform	36	4.43	-24	-36	-15
Rectus left and Frontal med Orbitalis bilateral	58	4.27	-6	54	-21
Cerebellum right and left	25	3.91	6	-51	-42
Hippocampus right	17	3.8	21	-12	-18
Putamen left	22	3.69	-18	6	9
Temporal mid left	29	3.65	-54	-42	-3
Frontal sup medial left	42	3.60	-9	60	30
Precentral left	27	3.52	-39	3	45
Cerebellum right	11	3.45	18	-69	-30
Precuneus	9	3.26	-6	-51	39
Rectus left	1	2.98	-3	30	-24
Frontal inferior tri Right	1	2.95	57	33	15
Diminution de l'activité avec l'augmentation de la taille des constituants					
Precuneus bilateral	510	4.92	-15	-69	33
Occipital superior Right		3.46	21	-72	45
Cingulum anterior bilateral	507	4.91	12	48	0
Frontal mid Left	194	4.76	-39	30	42
Supramarginal right and Temporal Superior right	329	4.21	60	-33	18
Parietal inferior right		3.91	54	-51	39
Parietal inferior left	119	4.21	-54	-57	45
Cingulum mid bilateral	189	4.17	0	-27	45
Front superior left	67	4.02	-18	15	60
Cerebellum right	25	4.02	48	-48	-42
Frontal mid right	101	3.7	36	30	48
Cerebellum left	31	3.51	-48	-60	-42
Frontal inferior opercularis right	59	3.49	54	12	3
insula right		3.04	36	21	6
Supramarginal left	37	3.36	-60	-30	24
Frontal mid left	9	3.32	-39	54	18
Occipital mid left	14	3.24	-30	-78	9
Frontal sup and mid right	32	3.21	24	45	36
Frontal sup right	11	3.21	24	57	12
Fusiform right	7	3.16	27	-69	-12
Cerebellum left	6	3.10	-36	-45	-45
Temporal superior left	7	3.09	-57	-3	3
Occipital mid right	5	3.02	36	-81	6
Parietal inferior right	2	2.96	39	-48	51
Temporal inf left	2	2.93	-57	-21	-27
Occipital sup right	1	2.93	27	-78	27
Insula left	1	2.92	-39	-12	0

TABLE 5.3 – Table statistique des zones dont l'activité est corrélée avec la taille des constituants langage. Moitié haute, augmentation de l'activité cérébrale avec l'augmentation de la taille des constituants. Moitié basse, diminution de l'activité à mesure que la taille des constituants augmente.

Langage oral

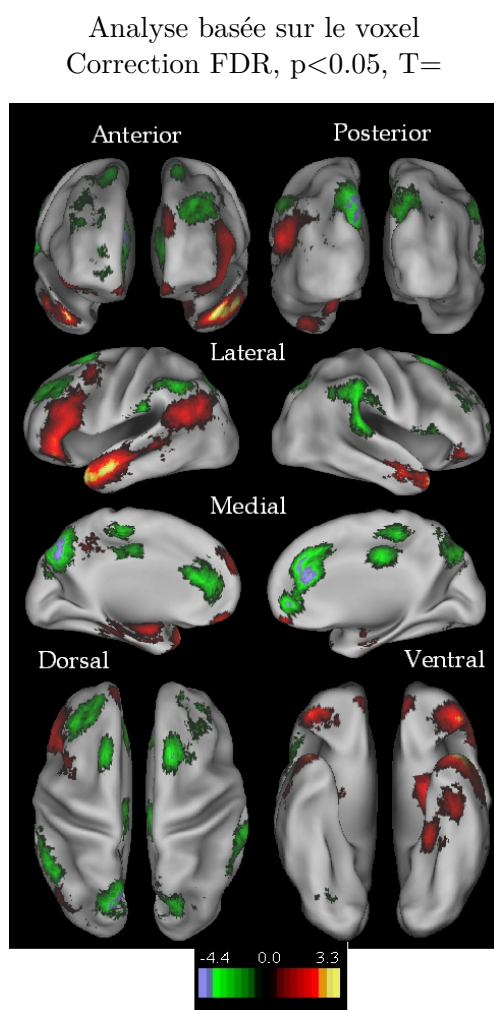


FIGURE 5.10 – Carte statistique représentant les zones sensibles à la structure en langage. Seuil corrigé FDR, $p < 0.05$. **En rouge**, les régions qui s'activent de plus en plus à mesure que la taille du plus grand constituant augmente. **En vert**, les régions qui répondent inversement, c'est-à-dire qui sont de moins en moins en activités à mesure que la taille du plus grand constituant augmente.

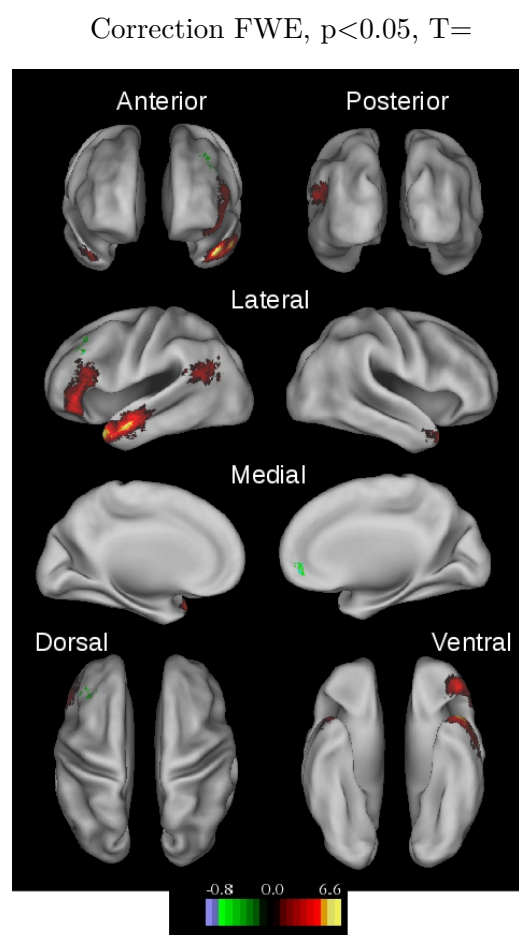


FIGURE 5.11 – Régions sensibles à la structure en langage. Carte statistique correction fwe, seuil $p < 0.05$.

ses parties orbitalis et triangularis, la jonction temporo-parietale, et le pôle temporal de l'hémisphère droit. Dans les zones dont l'activation corrèle avec une diminution de

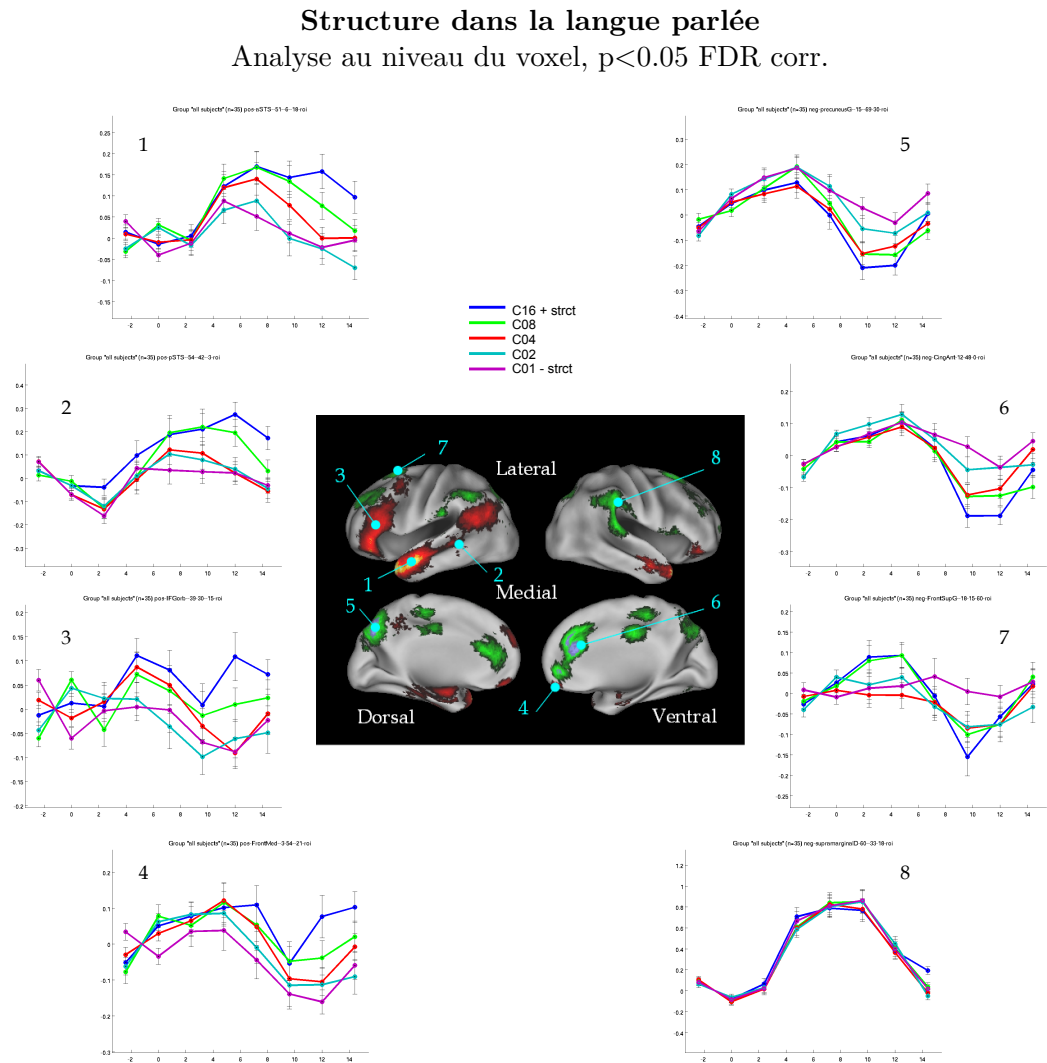


FIGURE 5.12 – Carte statistique représentant les zones sensibles à la structure en langage avec les décours temporels correspondant aux différentes conditions. 1 à 4 exemples de décours temporel pour les régions sensible de façon positive à la structure ; 1 : aSTS gauche, 2 : pSTS gauche, 3 : IFGorb gauche, 4 : Frontal supérieur médian. De 5 à 8 exemples de décours temporel pour les régions sensibles négativement à la structure ; 5 : precuneus, 6 : cingulaire antérieur, 7 : frontal supérieur, 8 : planum temporele droit. Les couleurs des courbes codent pour les différentes conditions de structure. $p < 0.05$ FDR corrigé.

la taille des constituants, il ne reste que 2 petites zones : frontale supérieure médiane droite, frontale médial droite.

La figure 5.12 montre le décours temporel des réponses hémodynamiques en fonction de la taille des constituants syntaxiques dans différentes régions caractéristiques. Le

aSTS et pSTS gauche, IFGorb gauche et la région frontale médiane (régions de 1 à 4) correspondent à des régions dont l'activité augmente avec la taille des constituants. Le précuneus, le cortex cingulaire antérieur, l'aire frontale supérieure et le planum temporel droit voient leur activité corrélée de façon négative avec l'augmentation de la taille des constituants. De plus, les activités des zones frontales supérieures médianes, du précuneus, du cingulaire antérieur et du frontal supérieur sont des déactivations c'est-à-dire que ces zones sont moins activées lors de l'écoute de nos stimuli par rapport à une ligne de base au repos.

Structure en Musique La figure 5.13 page ci-contre présente les régions dont l'activité était modulée par la taille des constituants en musique. L'analyse à un seuil non corrigé $p < 0.001$ met en évidence des zones répondant à la variation de structure syntaxique en musique. Aucune réponse n'est observée à un seuil plus strict utilisant une correction pour les comparaisons multiples (FDR ou FWE et $p < 0.05$). Le pôle temporal gauche, le gyrus précentral gauche, une petite partie du gyrus frontal inférieur droit à cheval sur la pars triangularis et opercularis voient leur activité augmentée à mesure que l'extrait musical est de plus en plus structuré. Les régions frontales médianes bilatérales répondent de façon inverse en augmentant leur activité à mesure que la taille des fragments musicaux diminue.

Des régions d'intérêt (intersection des activations de la carte statistique précédente $p < 0.001$ et de sphères de 10mm de rayon centrées sur les pics d'activation) ont été dessinées pour étudier l'amplitude des réponses hémodynamiques en langage et en musique (figure 5.14 page 170). L'amplitude de la réponse hémodynamique varie de la même façon pour les 2 domaines en fonction de la taille des fragments dans le pôle temporal gauche (réponse positive) et dans la région frontale médiale (réponse négative). En revanche, la réponse positive à la taille des constituants semble spécifique à la musique dans les régions du gyrus frontal inférieur droit et du STS médian droit.

Etant donné le peu d'activations visibles au seuil non corrigé $p < 0.001$, nous avons exploré la même carte statistique d'activation à un seuil plus tolérant, $p < 0.01$ (droite de

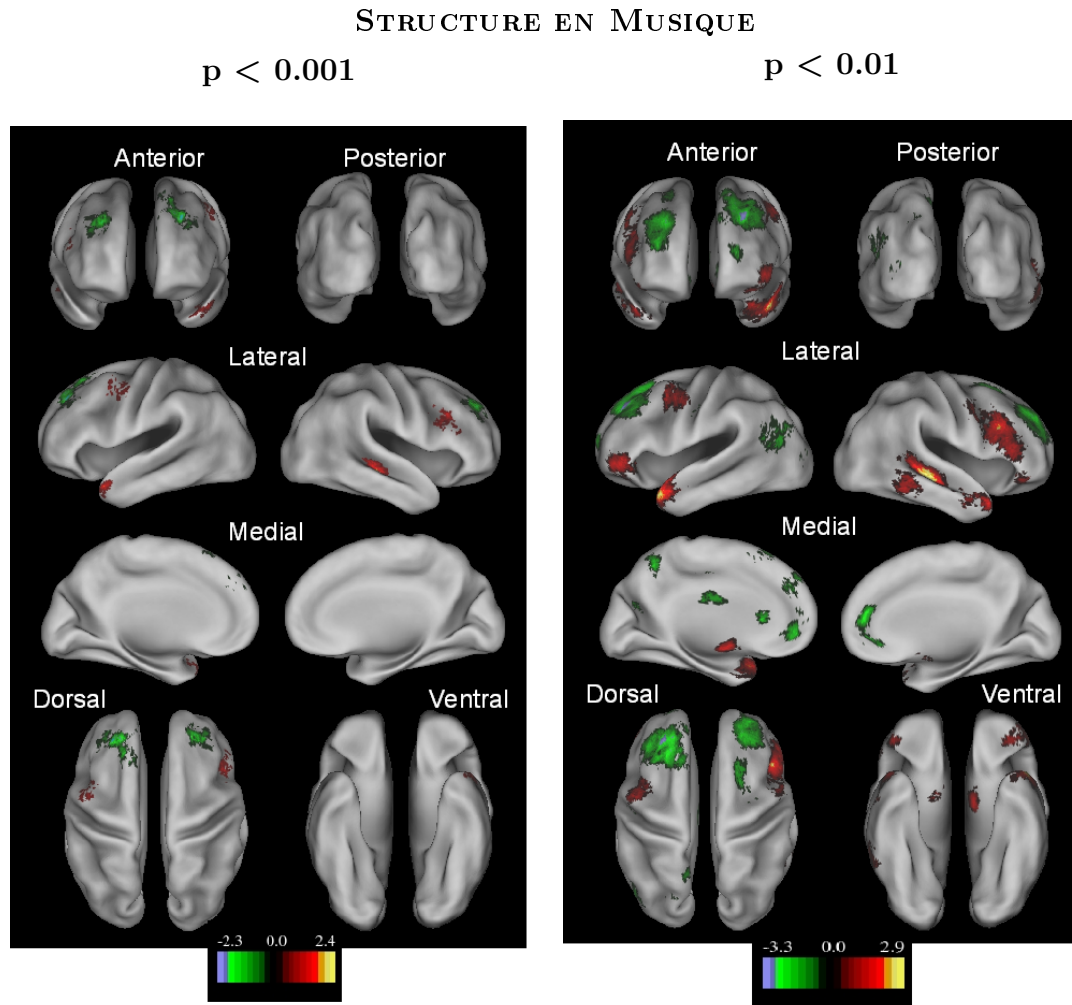


FIGURE 5.13 – Carte statistique : zones sensibles à la structure en musique à différents seuils statistiques, analyse en cerveau entier. **En rouge**, les zones dont l'activité augmente à mesure que les fragments sont de plus en plus longs. **En vert**, les zones dont l'activité diminue à mesure que les fragments sont de plus en plus longs.

la figure 5.13). En plus des activations décrites précédemment, on observe des activations du pôle temporal droit. La région inférieure frontale droite est plus large et la partie orbitalis du gyrus frontal inférieur gauche apparaît ainsi que le sulcus temporal inférieur dans sa partie postérieure. On remarque qu'un sous-ensemble de ces régions est commun au traitement de la structure en langage comme le pôle temporal gauche et le gyrus frontal inférieur orbitalis gauche.

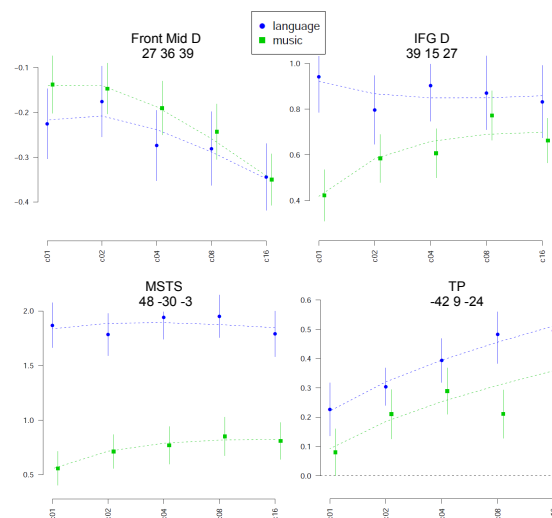


FIGURE 5.14 – Amplitude des réponses hémodynamiques en fonction de la taille des constituants syntaxiques. Les régions sélectionnées correspondent aux clusters sensibles à la structure en musique au seuil de $p < 0.001$ non corrigé intersectés avec des sphères de 10mm de rayon ayant pour centre le pic du cluster.

Réseau commun pour la structure en langage et en musique ? Dans le but de savoir si les mêmes zones étaient sensibles à la structure en langage et en musique, nous avons calculé l'interaction entre le degré de structure et le domaine (langage vs musique). Cette comparaison des aires sensibles à la structure en langage avec celles sensibles à la structure en musique révèle des interactions dans un seul sens à savoir entre la structure et (langage-musique) (figure 5.15 page suivante et table 5.4), c'est-à-dire des régions qui répondent plus fortement à la structure en langage qu'en musique.

TABLE 5.4 – Régions dont l'activité augmente plus avec la taille des constituants en langage par rapport à la musique

	Taille du Cluster	T (df=289)	x	y	z
Analyse corrigé à $p < .05$ FDR					
left Temporal mid and Angular	153	5.04	-51	-63	18
left Temporal Pole mid and sup	35	4.67	-45	15	-30
left Frontal superior medial	12	4.08	-9	48	39
left Temporal mid	8	3.87	-51	-9	-15

Les régions de l'hémisphère gauche sont plus sensibles à la structure en langage qu'en musique : le pôle temporal gauche, l'aSTS et le TPJ. Il est à noter que ce réseau de 3

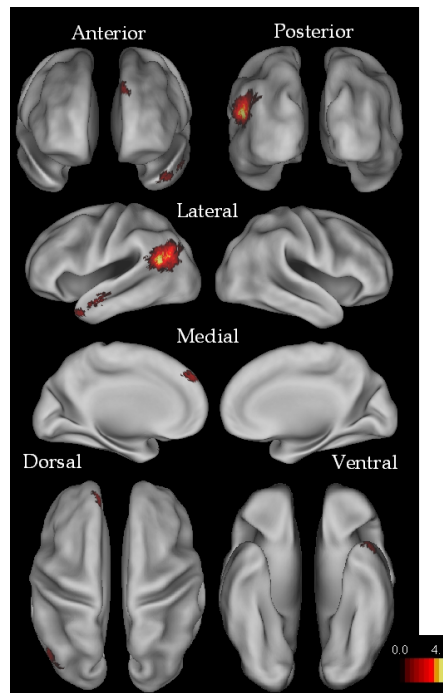


FIGURE 5.15 – Carte statistique correction fdr, seuil $p < 0.05$. Interaction entre la structure et le domaine (langage - musique). Régions plus sensibles à la structure en langage qu'en musique.

régions correspond aux zones qui répondaient spécifiquement à la structure en langage en présence de sémantique dans l'expérience en langage écrit (Pallier et al., 2011), c'est-à-dire dans les conditions avec de vrais mots et pas en Jabberwocky. Cette analyse confirme que ces 3 zones sont sensibles à la structure associée à un sens. Cette analyse suggère donc qu'il n'existe une différence de traitement de la structure entre langage et musique au profit du langage uniquement dans les zones mentionnées ci-dessus. Elle nous renseigne en outre sur les différences de traitement de la structure entre langage et musique mais pas sur leurs similitudes.

Une analyse de l'effet de structure en musique a été effectuée en restreignant l'analyse statistique aux zones répondant à la syntaxe linguistique (ou small volume correction). Pour cela, nous avons utilisé un masque ne contenant que les voxels des régions d'intérêt c'est-à-dire les voxels répondant à la syntaxe dans des expériences indépendantes (Pallier et al., 2011; Amunts et al., 2004). Notre masque comprenait les 7 régions d'intérêt

délimitées par l'expérience à l'écrit ((Pallier et al., 2011)). L'aire BA44 définie anatomiquement et fonctionnellement (Amunts et al., 1999; Amunts et al., 2004) a été ajoutée à ce masque parce que plusieurs expériences (par exemple (Friederici et al., 2010; Brauer and Friederici, 2007)) la mettent en évidence comme siège du traitement de la syntaxe. La figure 5.16 représente graphiquement les huit ROIs utilisées par la suite. Un second masque a été utilisé pour la recherche d'effet décroissant avec la structure et contient 6 régions d'intérêt provenant uniquement de l'expérience à l'écrit (figure 5.17 page ci-contre).

RÉGIONS D'INTÉRÊT POUR LES RÉPONSES POSITIVES À LA STRUCTURE EN LANGAGE

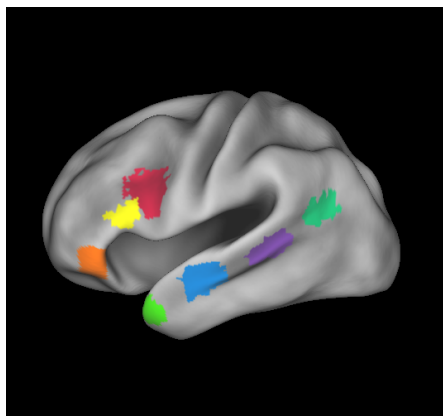


FIGURE 5.16 – Les 8 Régions d'intérêts utilisées dans les analyses suivantes. En frontal : en orange - gyrus frontal inférieur orbitalis (IFG orb), en jaune - IFG triangularis (IFG tri), en rose - BA 44 (IFG opercularis). En temporal : en vert pomme - Pôle temporal (TP), en bleu - la partie antérieure du sillon temporal supérieur (aSTS), en violet - la partie postérieure du STS (pSTS), en vert plus pastel - la jonction temporo pariétal ou gyrus angulaire (TPJ). Le putamen n'est pas visible sur cette carte bien que cette région ait été entrée dans les analyses.

L'intérêt de ce type d'analyse restreignant la recherche d'un effet significatif à un petit ensemble de voxels que l'on suppose intéressants, est d'obtenir des tests statistiques plus sensibles. En effet, l'optique ce genre d'analyse est un peu différente de celle en cerveau entier, on part désormais d'a priori que l'on cherche à vérifier ou à infirmer en sélectionnant les données.

Parmi les 8 régions d'intérêt répondant de façon positive à la structure en langage,

RÉGIONS D'INTÉRÊT POUR LES RÉPONSES NÉGATIVES À LA STRUCTURE EN LANGAGE

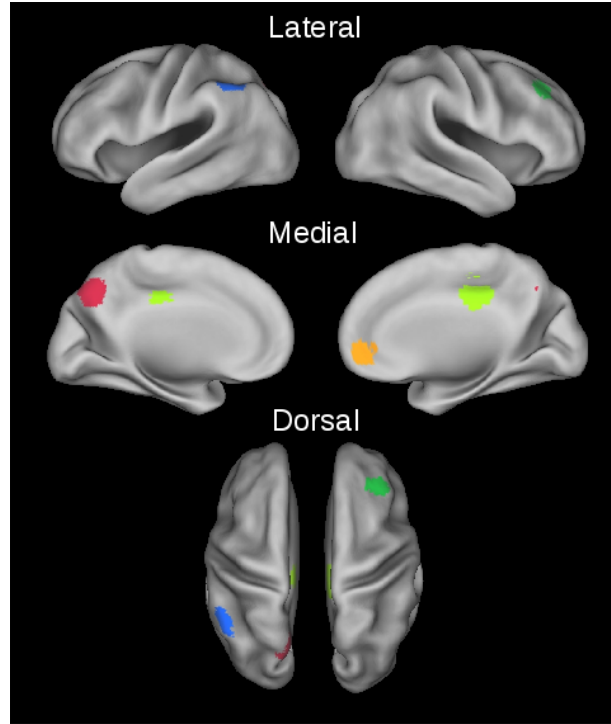


FIGURE 5.17 – Les Régions d'intérêt répondant de façon négative à la structure en langage dans l'expérience à l'écrit (Pallier et al., 2011). En bleu - le gyrus inférieur pariétal gauche, en vert - région médiane frontale droite, en rose - le précunéus, en vert anis - le cingulaire médian, en orange - cingulaire antérieur. Sur cette figure, le région d'intérêt occipitale n'est pas représentée (pour des raisons techniques) même si elle a été utilisée dans les analyses suivantes.

seule une partie du pôle temporal gauche est sensible à la taille des constituants en musique de façon significative (correction fwe effectuée au niveau du voxel - $p < 0.05$ - et au niveau du cluster - $p < 0.08$).

Parmi les régions répondant de façon négative à la structure en langage, seule une partie du cortex cingulaire antérieur (frontal médian) voit son activité corrélée à la diminution de structure en musique (correction fwe $p < 0.05$ effectuée au niveau du voxel et au niveau du cluster - $p < 0.07$, 8 voxels)

Notre paradigme, développé pour des stimuli linguistiques, ne permet pas de mettre en évidence de la même manière les zones sensibles à la structure en musique. Cependant,

même s'il est moins sensible pour la musique, les cartes obtenues pour la structure en musique à un seuil moins strict de $p < 0.01$ non corrigé, laisse penser à un manque de sensibilité au niveau du groupe⁴.

L'analyse de groupe est sensible à la localisation de la réponse hémodynamique évoquée par un stimulus. Elle permet si sa localisation est stable à travers les sujets de la généraliser au niveau de la population. Même si l'amplitude de cette réponse est faible et que la localisation est stable alors l'analyse au niveau du groupe sera significative et informative au niveau de la population. En revanche, si la localisation de la réponse diffère entre les sujets même si son amplitude est forte, l'analyse de groupe ne sera pas significative même si une réponse hémodynamique est retrouvée systématiquement chez tous les sujets. L'analyse de groupe ne peut donc pas répondre à toutes les questions et il est possible de ne pas trouver de résultat au niveau de la population alors qu'au niveau individuel, l'effet sera relativement stable mais sa localisation légèrement différente. C'est pourquoi, nous nous sommes intéressés ensuite à la réponse à la structure en musique au niveau individuel. Il s'agit de savoir si chaque sujet active les mêmes zones cérébrales pour traiter la structure en constituant en langage et en musique. Nous avons utilisé l'approche proposée par Fedorenko (Fedorenko et al., 2010). Cette approche fonctionne de la façon suivante : les régions d'intérêt définies comme précédemment (8 pour la sensibilité positive à la structure et 6 pour la négative) servent de masque. Au sein de ce masque, pour chaque sujet sont sélectionnés les voxels spécifiquement activés par le langage en intersectant le masque avec un localisateur correspondant au contraste contenant toutes les conditions langagières réunies (réponses positives ou négatives selon le cas, les cartes sont seuillées à un seuil statistique de $p < 0.001$ non corrigé). Ainsi un certain nombre de voxels, différents pour chaque sujet, sont sélectionnés comme répondant et au langage dans notre expérience et potentiellement sensibles à la taille de constituants en langage (masque de régions d'intérêt). Ensuite l'activité dans ces voxels

4. Une autre explication plausible pour expliquer le peu de zones répondant à la complexité positive en musique serait que les zones activées le soient trop peu (en intensité). Ce manque de sensibilité pourrait être lié à la construction de nos stimuli musicaux puisque nous avons déstructuré plusieurs éléments responsables de la structure, comme la structure rythmique, structure harmonique, structure mélodique etc.

est analysée en fonction de la taille des fragments en musique pour éventuellement trouver une corrélation entre l'amplitude de la réponse hémodynamique et la structure syntaxique. Une analyse statistique en modèle mixte est effectuée en rassemblant les données individuelles, pour tester si l'effet est significatif dans la population.

Dans un premier temps, la figure 5.18 page suivante montre que le masque de région d'intérêt provenant de l'expérience en langue écrite reflète effectivement les activations à la structure en langage dans notre expérience auditive. Ce graphique montre la taille de l'effet linéaire de structure en langage dans les zones sélectionnées par l'analyse : voxels répondant au langage et appartenant aux régions d'intérêt de l'expérience indépendante en langue écrite.

La sensibilité à la structure dans les voxels sélectionnés est très significative dans presque toutes les régions d'intérêt, qu'elles répondent de façon positive ou négative à la structure. Notre masque de régions d'intérêts définit de façon correcte les régions dont l'activation augmente ou diminue à mesure que la taille des constituants croît. Le putamen est la seule région d'intérêt issue de l'expérience en langue écrite dont l'activité ne se semble pas modulée par la taille des constituants en langage. Elle était définie par 9 voxels dans l'expérience en langue écrite et parmi les sujets de notre expérience on en retrouve en moyenne zero. Cette zone n'est sans doute pas assez grande et assez stable chez les sujets pour être reproduite au niveau individuel dans notre expérience. A l'appui de cette explication, la zone de putamen que nous mettons en évidence dans notre analyse de groupe n'est pas située exactement au même endroit que la région d'intérêt définie dans l'expérience en langue écrite. La région d'intérêt occipitale de l'expérience en langue écrite n'est pas sensible à la structure en langage oral ce qui est plutôt logique étant donné qu'il s'agit d'une zone répondant aux stimuli visuels.

Il est donc valide d'utiliser les régions d'intérêt définies dans l'expérience en langue écrite comme masque des régions répondant de façon linéaire à la taille des constituants en langage orale dans notre expérience.⁵

5. Il aurait été plus logique de choisir comme régions d'intérêt celles qui répondaient à la structure en langage au niveau individuel, et de regarder dans ces régions définies individuellement si elles réagissaient à la structure musicale. Cela n'a pas été possible. En effet, au niveau individuel, les réponses à la structure

TAILLE DE L'EFFET DE STRUCTURE EN LANGAGE PAR ROIS INDIVIDUELLES (APPROCHE PROPOSÉE PAR (FEDORENKO ET AL., 2010))

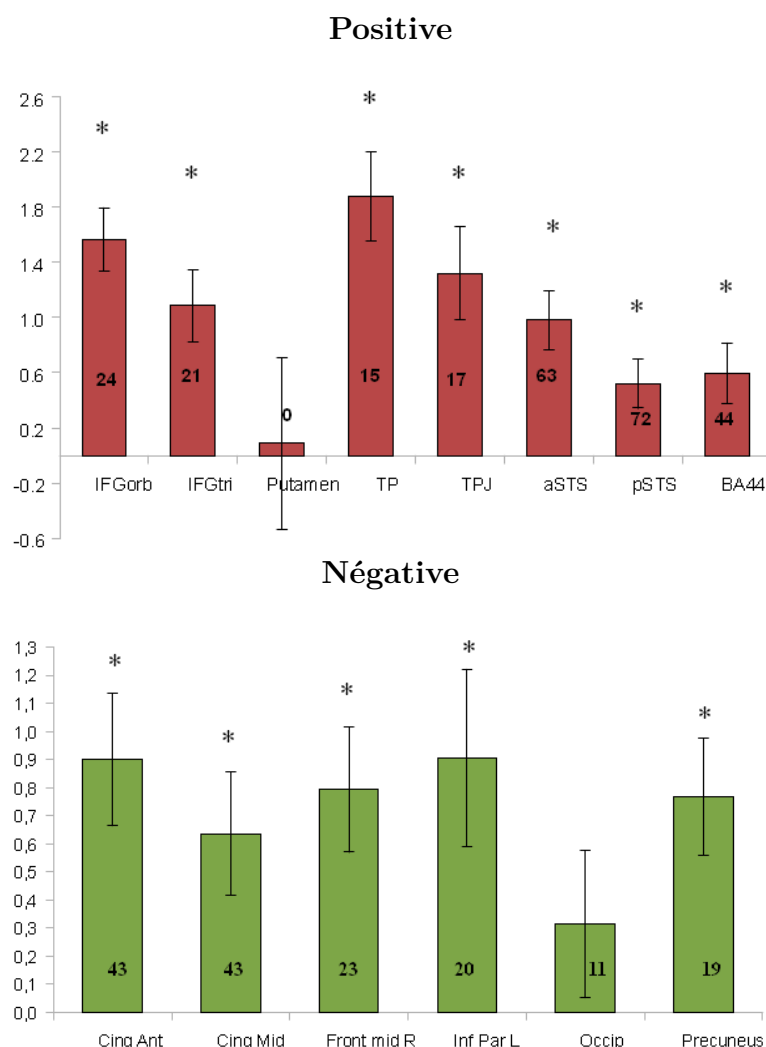


FIGURE 5.18 – Sensibilité à la structure en langage oral des voxels sélectionnés au niveau individuel par la méthode Fedorenko : ces voxels répondent au langage (localiser toutes les conditions langagières) et appartiennent aux régions d'intérêt définies dans l'expérience en langue écrite + BA44. Le nombre inscrit à l'intérieur des barres correspond au nombre moyen de voxels sélectionnés par sujet. Les étoiles représentent les régions dont la réponse à la structure est significative à un seuil corrigé $FDR < 0.01$. En abscisse, l'axe représente la valeur de la pente de la courbe qui représente le mieux l'activation linéaire à la taille des constituants en musique.

en langage ne sont pas assez robustes pour pouvoir servir de masque à de nouvelles analyses : il n'y aurait eu en effet que très peu de voxels, voir pas du tout selon les sujets. Dans ces conditions d'absence de réponse robuste à la structure en langage au niveau individuel malgré la force des réponses au niveau

Nous pouvons désormais essayer de répondre à la question qui nous préoccupe : les régions sensibles à la structure en constituant le sont-elles dans tous les domaines et particulièrement en musique ?⁶

Cette analyse prouve que les réponses hémodynamiques dans le gyrus frontal inférieur gauche (orbitalis et triangularis), le pôle temporal, la partie postérieure du gyrus temporal supérieur sont corrélées de façon positive avec l'augmentation de la taille des fragments en musique. Le recouvrement intersujets des voxels sélectionnés est bon (plus de 75% des voxels sélectionnés sont communs) et le nombre de sujets présentant des voxels dans les zones est relativement important.

du groupe, nous avons choisi un masque de régions sensibles la structure en constituant en langage provenant d'une expérience indépendante, pour éviter toute circularité dans l'analyse.

6. Les effets en musique étant plus restreints en nombre de voxels impliqués et plus faible en amplitude par rapport à ceux du langage, nous avons d'abord cherché le seuil statistique optimal pour le localisateur du langage. En effet, plus le seuil sera conservateur, plus le nombre de voxels sélectionnés sera petit. Il sera donc plus difficile de regarder parmi ce petit jeu de voxels l'effet de la taille des constituants en musique. Nous avons testé 2 seuils statistiques non corrigés pour ce localisateur : $p < 0.01$ et $p < 0.001$. La significativité des effets n'est pas très différente entre ces 2 seuils, même si le nombre de voxels varie. Pour des raisons de logique de comparaison avec les effets en langage, nous choisissons d'utiliser le seuil statistique non corrigé $p < 0.001$ pour les contrastes linguistiques servant de localisateur comme dans le test précédent.

**TAILLE DE L'EFFET DE STRUCTURE EN MUSIQUE PAR ROIS
INDIVIDUELLES (APPROCHE PROPOSÉE PAR (FEDORENKO ET AL.,
2010))**

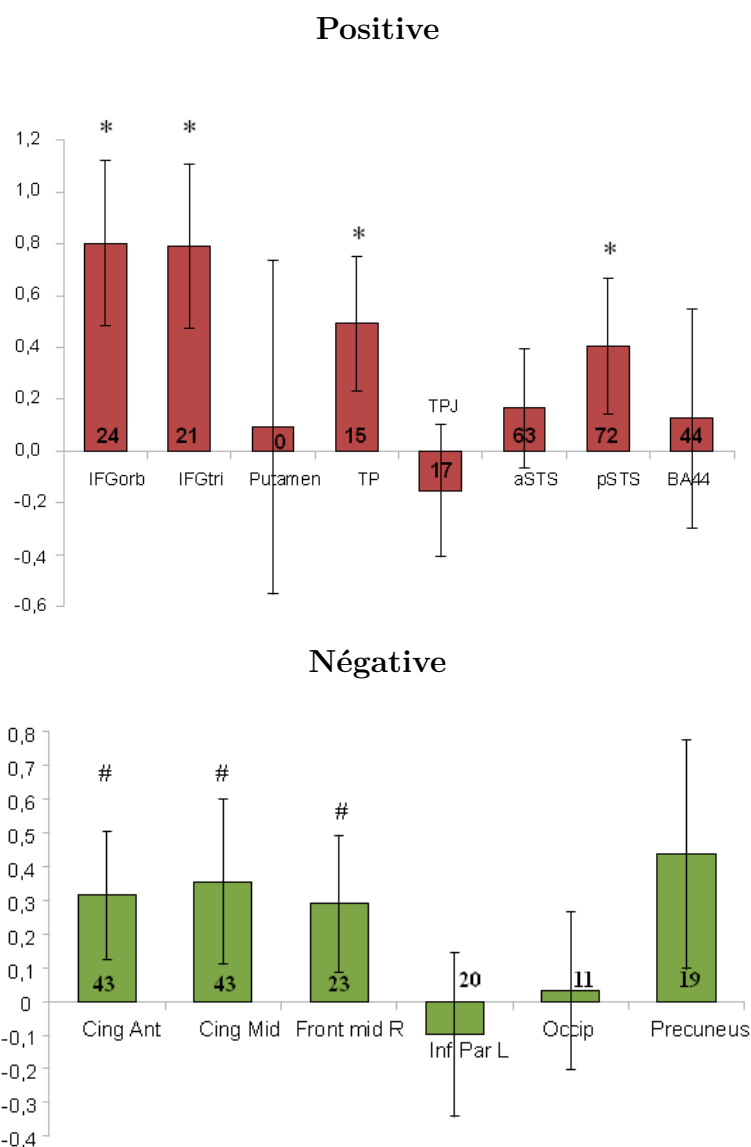


FIGURE 5.19 – Sensibilité à la structure en musique des voxels sélectionnés au niveau individuel par la méthode Fedorenko : ces voxels répondent au langage (localiser toutes les conditions langagières) et appartiennent aux régions d'intérêt définies dans l'expérience en langue écrite + BA44. Le nombre inscrit à l'intérieur des barres correspond au nombre moyen de voxels sélectionnés par sujet. Les étoiles * représentent les régions dont la réponse à la structure est significative à un seuil corrigé FDR < 0.05 et les dièses # représente des régions dont la réponse est marginale, $0.05 < p < 0.09$ non corrigé.

L'analyse des régions dont l'activité décroît à mesure que la taille des fragments

augmente en musique montre un effet marginal dans les régions cingulaires et la région frontale médiane droite. Il est à noter que ces régions d'intérêt mises en évidence dans l'expérience en langue écrite (Pallier et al., 2011) sont légèrement décalées par rapport à celles détectées dans notre expérience auditive. Cette différence de localisation pourrait être à l'origine de l'absence de réponse significative à la variation de structure dans ces régions.

Différence de traitement de la structure entre musiciens et non-musiciens

Concernant le langage, aucune différence dans le traitement de la structure syntaxique n'a été trouvée au niveau de l'analyse de groupe entre les musiciens et les non-musiciens, ni lors de l'augmentation de la taille des constituants, ni lors de leur diminution.

Les résultats de la partie musicale séparant les musiciens des non-musiciens sont présentés dans la figure 5.20 page suivante à un seuil non corrigé $p < 0.001$. Les cartes d'activation sont différentes pour nos 2 groupes. Les non-musiciens ne présentent que des zones dont l'activité augmente avec la taille des constituants : le gyrus frontal inférieur gauche dans sa pars orbitalis, une région pariétale inférieure bilatérale. Les musiciens, eux, ne présentent pas d'aires cérébrales répondant de façon linéaire croissante avec la taille des constituants. Par contre, plusieurs zones relativement volumineuses répondent de façon linéaire à mesure que la taille des constituants diminue, comme la jonction tempor-pariétale gauche, le précuneus bilatéralement, les régions frontales médianes bilatérales et les régions médiales supérieures bilatérales (cingulaire antérieur). Ces zones à l'exception de la jonction temporo-pariétale sont communes au réseau dont l'activité décroît à mesure que la taille des constituants augmente en langage.

Nous avons ensuite cherché les régions qui répondaient différemment à la taille des constituants en musique selon l'expertise musicale des sujets. La carte obtenue grâce au contraste [(musiciens - non-musiciens) * complexité positive en musique], seuillée à $p < 0.05$ FDR ne montre pas de région répondant plus chez les non-musiciens que chez

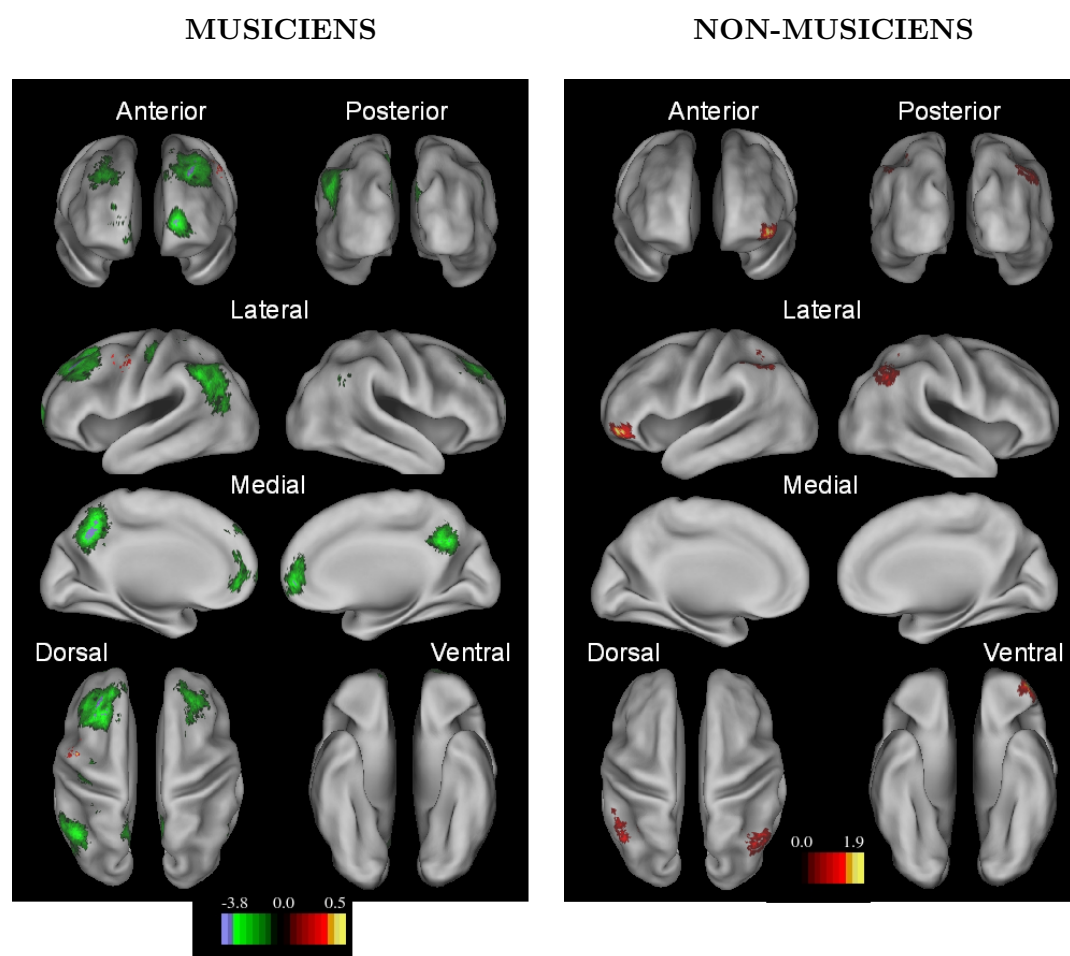


FIGURE 5.20 – Carte statistique relative au traitement de la structure en musique en séparant musiciens (à gauche) et non-musiciens (à droite). En rouge, les régions sont sensibles de façon positive à la structure et en vert, les régions le sont de façon négative. Seuil non corrigé $p < 0.001$

les musiciens pour la taille des constituants. Cette carte nous aurait révélée des régions plus actives chez les musiciens par rapport aux non-musiciens au niveau de la structure positive et des régions plus actives chez les non-musiciens répondant à la structure de façon négative. En revanche, l'interaction opposée [(non-musiciens - musiciens) * complexité positive en musique] $p < 0.05$ FDR corrigée est montrée sur la figure 5.21 page suivante.

Intéraction Compétence musicale et structure en musique

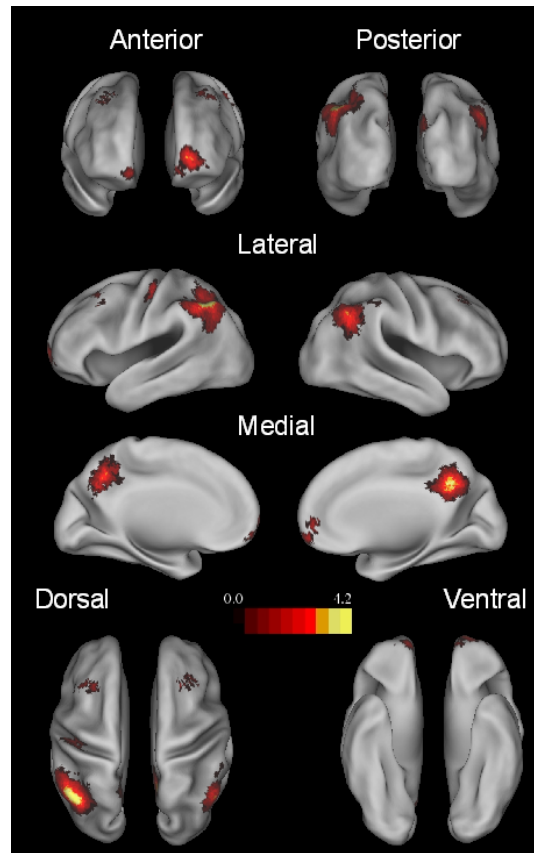


FIGURE 5.21 – Carte statistique étudiant l'impact de la compétence musicale sur le traitement de la structure en musique. Seuil corrigé FDR $p < 0.05$

La jonction temporo-pariétale gauche présente un comportement opposé pour les musiciens et les non-musiciens. Elle est sensible à la structure de façon positive pour les non-musiciens et de façon négative pour les non-musiciens. Cette carte 5.21 résume majoritairement les réponses à la structure de façon négative pour les musiciens avec une différence d'activation opposée pour le TPJ. Il existe une différence entre les musiciens et les non-musiciens dans le traitement de la taille des constituants particulièrement visible lorsque la structure diminue. Afin d'explorer les activations au niveau individuel, nous avons appliqué la même méthode que précédemment (Fedorenko et al., 2010) chez les deux groupes de sujets.

Le graphique 5.22 page 183 montre une implication des régions de l'hémisphère

gauche détectant la structure en langage dans le traitement de la structure en musique au niveau individuel. Chez les musiciens, la taille des fragments en musique module la réponse hémodynamique dans l'IFG orbitalis et triangularis, le pôle temporal et dans le pSTS⁷. Chez les non-musiciens, aucune zone ne voit son activité modulé par la taille des fragments musicaux de façon significative, la seule qui le soit marginalement (mais sans correction FDR) est l'IFG tri ($p < 0.07$). Dans les ROIs dont l'activité diminuait à mesure que la taille des constituants syntaxiques linguistiques augmentaient, les résultats sont faibles pour la variation de taille des fragments musicaux puisqu'on observe une réponse marginale chez les musiciens dans les zones cingulaires antérieure et médiane et frontale médiane droit. Aucune réponse n'est significative chez les non-musiciens.

Les 2 analyses de groupe utilisant les statistiques individuelles, concordent sur le fait que les musiciens et les non-musiciens ont des profils d'activation différents en localisation et en amplitude pour traiter la structure en musique.

L'analyse de groupe, qui n'est pas très sensible, met en évidence un réseau dont les réponses hémodynamiques décroissent à mesure que la taille de la structure syntaxique augmente chez les musiciens. Ce réseau semble différent du réseau utilisé pour le langage comme le suggère les analyses individuelles sur les ROIs négatives. Le réseau mis en évidence pour les non-musiciens dans l'analyse de groupe n'est pas très étendu mais contient une région commune au langage à savoir l'IFG orbitalis, l'analyse individuelle montre une tendance marginale de cette zone pour le traitement de la structure en musique.

7. De nouveau, on remarque une absence de données pour le putamen pour effectuer l'analyse statistique sans doute en raison de la taille restreinte de la région d'intérêt initiale et/ou de l'absence de sensibilité de la zone pour le langage à travers les sujets.

TAILLE DE L'EFFET DE STRUCTURE SYNTAXIQUE EN MUSIQUE PAR ROIS INDIVIDUELLES EN SÉPARANT MUSICIENS ET NON-MUSICIENS

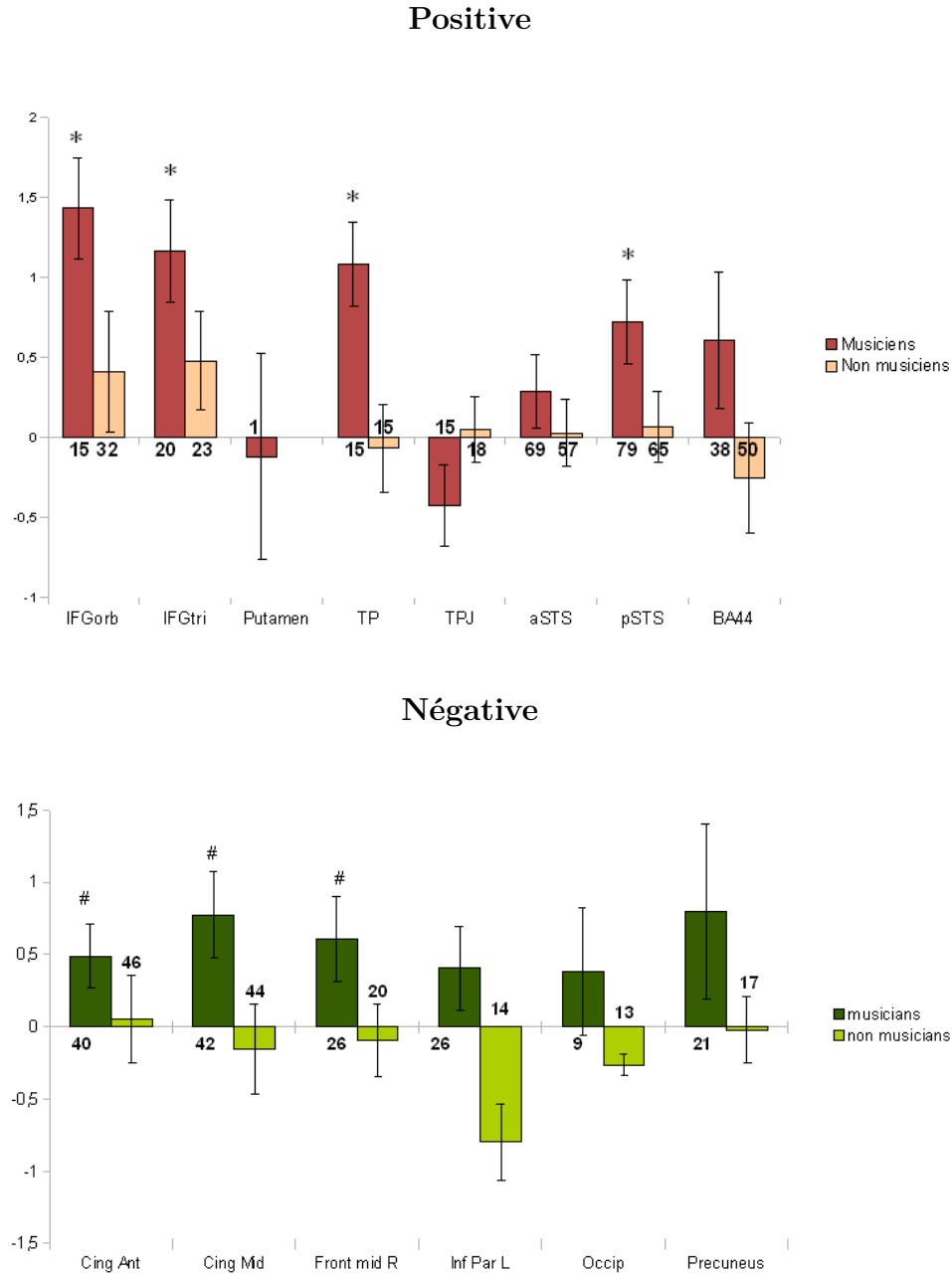


FIGURE 5.22 – Taille de l'effet de structure en musique dans les ROIs individuelles chez les musiciens et les non-musiciens : ces voxels répondent au langage (localisateur avec toutes les conditions langagières) et appartiennent aux régions d'intérêt définies dans l'expérience en langue écrite + l'aire BA44. Le nombre inscrit à l'intérieur des barres correspond au nombre moyen de voxels sélectionnés par sujet. Les étoiles * représentent les régions dont la réponse à la structure est significative à un seuil corrigé FDR $p < 0.01$. Les dièses # représentent les régions pour lesquelles la réponse à la structure en musique est marginalement significative à un seuil corrigé FDR $p < 0.07$

Discussion

Des musiciens et des non-musiciens ont entendu des stimuli langagiers et musicaux dans lesquels nous avons fait varier la quantité de structure syntaxique de façon symétrique.

Différences des activations globales pour le langage et la musique

Les régions temporales supérieures bilatérales correspondant aux aires auditives et le gyrus frontal inférieur bilatéral sont activés par nos stimuli linguistiques et musicaux. Des activations plus fortes pour le langage ont été observées dans les aires auditives. Une explication possible pour cette différence pourrait être une différence de niveau sonore entre les deux types de stimuli. L'analyse du questionnaire rempli par les sujets après avoir passé le test dans le scanner ne révèle pas de différence d'intensité dans un sens ou dans l'autre de la perception du niveau sonore⁸. Une analyse post-hoc du niveau d'intensité des stimuli par domaine a été réalisée, voir les tables 5.5 page ci-contre et 5.6 page suivante. La valeur de la pression acoustique (en Pascal)⁹ révèle que les stimuli langagiers ont une intensité sonore plus grande que les stimuli musicaux. La mesure du volume sonore perçu (en sone)¹⁰ montre que la perception des stimuli des 2 domaines est du même ordre de grandeur, la perception n'est donc pas très différente. Il est donc peu probable que les différences d'activation dans les aires auditives soient liées à des différences d'intensité.

Une autre explication pour ces écarts d'intensité d'activation des aires cérébrales pourrait être liée à des différences d'attention entre les deux types de stimuli engendrée par la tâche. Il est possible que la tâche soit plus facile pour la musique par rapport au

8. Certains sujets (Nombre) ne perçoivent pas de différence d'intensité, N ont trouvé la musique plus forte que le langage et N ont trouvé le langage plus fort que la musique

9. La pression acoustique et l'intensité sonore sont liées par une relation linéaire lorsque l'amplitude est petite. En fait, l'intensité sonore locale instantanée est le produit de la pression acoustique et de la vitesse des particules acoustiques. Cette vitesse et la pression acoustique sont liées par une relation linéaire en cas de faible amplitude. (cf l'article Sound Pressure de la wikipedia)

10. Le sone est une unité de mesure du volume sonore perçu. 1 sone correspond au niveau sonore d'un son pur de 1kHz à 40dB. 2 sonos correspondent à l'intensité sonore d'un son perçu 2 fois plus fort qu'un sone.

Pression Acoustique

RMS (en Pascal)	Langage	Musique
Min.	0.173	0.055
1 ^{er} quartile	0.182	0.057
Médiane	0.186	0.057
Moyenne	0.186	0.057
3 ^{ème}	0.189	0.058
Max.	0.198	0.059

TABLE 5.5 – Pression acoustique des stimuli des différents domaines exprimé par leur RMS : root mean square (ou valeur efficace/ moyenne quadratique) en Pascal.

**Volume Sonore perçu
(en sone)**

Langage	Musique
666.49	632.75

TABLE 5.6 – Volume sonore en Sone

langage du fait de la plus grande saillance acoustique des cibles. Même si les sujets ont rapporté que la tâche était facile, ils ont remarqué que la détection des cibles musicales était plus facile que les cibles linguistiques. En effet, la différence de timbre entre les instruments est importante : piano vs. clavecin alors que la différence de voix entre les voix féminine et masculine était moins frappante notamment du fait de la prosodie de phrase utilisée pour l'enregistrement. De plus, il est possible d'envisager le fait que les stimuli langagiers étaient plus intéressants que les stimuli musicaux puisqu'il comportait de la sémantique. Notre interprétation de la différence d'activation des aires auditives serait due à des différences d'attention dues aux stimuli ou à la tâche ou à une interaction des deux.

Différences dans les activations globales entre les musiciens et les non-musiciens

La musique provoque des activations plus importantes de la zone prémotrice et motrice gauche chez les musiciens par rapport aux non-musiciens. Cette dissimilitude pourrait être expliquée par le fait que les musiciens pourraient être en train de s'imaginer jouer ce qu'ils entendent ; ils recruteraient alors le cortex moteur gauche représentant donc la main droite, main qui joue la mélodie dans nos extraits. Quant à la région postérieure du sillon temporal supérieur droit, il correspond à l'analogue de la zone pSTS

que nous mettons en évidence dans le traitement du langage. Cette région a été liée avec le traitement de la hauteur des notes en musique ((Warren et al., 2003) ou (Koelsch, 2011; Peretz and Zatorre, 2005) pour des revues sur la neuroanatomie de la perception de la musique).

Différences anatomiques entre les musiciens et les non-musiciens

Ces différences fonctionnelles pourraient être sous-tendues par des différences anatomiques. Une analyse de morphométrie basée sur les voxels (VBM) a été réalisée pour chercher des différences anatomiques entre nos 2 groupes de sujets. L'analyse statistique par test t entre musiciens et non-musiciens révèle des zones de taille réduite dont le volume est plus important pour les musiciens par rapport aux non-musiciens. La figure 5.23 page ci-contre et la table associée 5.7 révèlent une zone de quelques voxels à la jonction temporo pariétal droit (temporal supérieur), une petite partie du cortex médian frontal droit et une partie supérieure du frontal médian gauche, à un seuil non corrigé, $p < 0.001$ ¹¹. Les régions isolées par le contraste musicien - non-musiciens en musique ne sont donc pas retrouvées dans l'analyse anatomique. L'analyse statistique ne révèle pas de régions dont le volume est plus important chez les non-musiciens que chez les non-musiciens. Ces analyses ne permettent pas d'imputer les différences d'activations à des différences anatomiques entre les 2 groupes.

Analyse VBM (voxel-based morphometry)

Musiciens - non-musiciens

Analyse seuil non corrigé, $p < 0.001$

	Taille du cluster	T (df=33)	x	y	z
Temporal supérieur Droit	120	4.20	48	-39	19
Frontal mid Droit	15	3.69	44	27	49
Frontal supérieur médian Gauche	18	3.69	-3	64	24

TABLE 5.7 – Table statistique de la VBM comparant la taille de la matière grise entre les musiciens et les non-musiciens.

11. Rien ne passe les seuils corrigés, et même à un seuil non corrigé $p < 0.001$ la correction au niveau du cluster n'est pas significative.

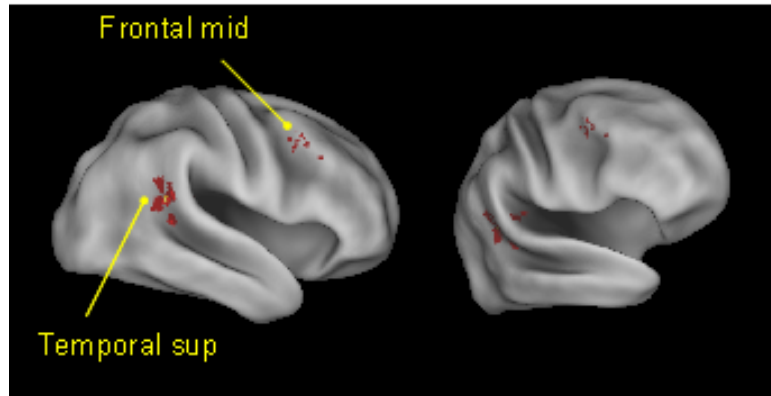


FIGURE 5.23 – Carte statistique du résultat de l'analyse par VBM. Seul l'hémisphère droit est représenté pour plus de clarté. Les 2 seules zones visibles sont la région temporal supérieure droite et frontal médiane droite. La région frontale supérieure gauche n'est pas représentée ici.

Quelques études s'intéressant particulièrement aux différences anatomiques induites par la pratique de la musique montrent des différences entre cerveaux de musiciens et non-musiciens. La plupart d'entre elles se focalisent sur une région cérébrale sans réaliser d'analyse sur cerveau entier (par exemple, (Abdul-Kareem et al., 2011; Sluming et al., 2002) trouvent des différences de volume de l'aire de Broca, pars opercularis pour (Abdul-Kareem et al., 2011) et dans la partie triangulaire pour (Sluming et al., 2002) avec une augmentation chez les musiciens) comme celle que nous avons effectuée. Une des quelques analyses réalisées en cerveau entier (Bermudez et al., 2009) montre, sur une relativement grande cohorte de sujets (71 musiciens et 64 non-musiciens), des différences d'épaisseur corticale avec une plus grosse épaisseur pour les musiciens dans une région temporale supérieure (localisée de façon plus antérieure que celle que nous mettons en évidence), une large région bilatérale dorsolatérale frontale. Les mêmes auteurs montrent avec une analyse VBM sur les mêmes données une augmentation de matière grise chez les musiciens dans la portion postéro-latérale du gyrus de Heschl. Cette zone se trouve de façon plus antérieure à celle que nous trouvons. Il convient de remarquer que les différentes techniques de mesures anatomiques (VBM, épaisseur corticale etc.) ne possèdent pas la même sensibilité, l'épaisseur corticale apparaît comme étant plus précise. Notre analyse VBM a révélé une région légèrement plus postérieure que celle

trouvée avec beaucoup plus de sujets dans l'expérience de Bermudez. Ce nombre différent de sujet pourrait être à l'origine de la différence de localisation. L'algorithme de segmentation entre matière blanche et matière grise diffère également entre notre étude et celle des chercheurs canadiens. De plus, il n'existe pas de consensus sur les différences anatomiques entre musiciens et non-musiciens. Par exemple, (Foster and Zatorre, 2010a) montrent que l'épaisseur corticale du gyrus de Heschl droit (cortex auditif) et des sulci intrapariétaux droit et gauche et leur volume de matière grise prédit la performance à une tâche portant sur la hauteur des notes. Ou encore, (Ono et al., 2011) ont effectué une analyse VBM entre 8 musiciens et 8 non-musiciens qui pointe une différence en faveur des musiciens dans les régions antérieure et postérieure du cortex cingulaire¹². Une autre étude (Gaser and Schlaug, 2003) qui étudie des musiciens professionnels (20), amateurs (20) et des non musiciens (40), uniquement des hommes appariés pour leur âge, met en évidence une corrélation entre la qualification en musique et le volume de matière grise dans plusieurs régions dont des régions motrices, prémotrices, des aires pariétales supérieures antérieures, le gyrus inféro-temporal de façon bilatéral mais aussi dans le gyrus de Heschl gauche et dans le gyrus inféro frontal de ce même hémisphère.

Pour conclure, avec des cohortes nombreuses de sujets, il est possible de trouver des différences anatomiques entre musiciens et non-musiciens notamment dans les gyri intrapariétaux et dans le cortex auditif (gyrus de Heschl) à droite (entre autres (Foster and Zatorre, 2010a; Bermudez et al., 2009)), endroit supposé du traitement de la hauteur des notes (pitch). Des études avec un plus petit nombre d'individus, comme la nôtre, n'observe que des différences sur quelques voxels dans l'analyse VBM. Ono et ses collègues (Ono et al., 2011) montrent quelques voxels du cingulaire antérieur et postérieur qui diffèrent entre les musiciens et les non-musiciens. Cependant, ces zones sont différentes de celles que nous observons. Ces auteurs ont constaté sur leurs groupes restreints d'individus des différences de localisation entre les analyses anatomiques et les analyses fonctionnelles, comme dans notre expérience, puisque les réponses cérébrales de

12. Cette étude montre une différence de latéralisation entre les musiciens et les non-musiciens. Grâce un paradigme oddball en MEG, ils mettent en évidence un mismatch field latéralisé dans le cortex auditif droit pour les non-musiciens et un mismatch field bilatéral dans le cortex auditif chez les musiciens.

type MMF à un paradigme oddball étaient latéralisées différemment chez les musiciens (bilatérales) et chez les non-musiciens (latéralisées à droite) dans le cortex auditif.

Structure en langage

L'objectif de cette expérience était l'étude de la structure syntaxique en utilisant des stimuli linguistiques et musicaux construits de façon similaire. Dans cette section, nous nous intéressons à la structure syntaxique linguistique. L'augmentation de la taille des constituants syntaxiques en langage sollicite, de la même façon pour les musiciens et les non-musiciens, les régions gauches du pôle temporal, aSTS, TPJ, IFG orb et tri, le pôle temporal droit et l'IFG orb droit comme dans les expériences en langue écrite utilisant une présentation sérielle rapide (Pallier et al., 2011; Snijders et al., 2009), représentés respectivement figures 5.24 page suivante et 5.25 page suivante. Dans l'expérience de (Pallier et al., 2011), la taille des constituants syntaxiques variait de façon paramétrique, tandis que dans l'expérience de (Snijders et al., 2009) seules des phrases et des listes de mots étaient contrastées. Malgré ces différences, le réseau dont l'activité est modulée par la taille des constituants correspond au réseau mis en évidence dans notre expérience auditive ; que ce soit de façon positive lorsque les aires cérébrales sont plus activées pour les phrases par rapport aux constituants de plus petite taille et aux listes de mots ou que ce soit de façon négative quand les zones sont moins activées par les phrases par rapport aux constituants syntaxiques plus courts. Il est intéressant de noter que comme dans l'expérience de Snijders, ces diminutions d'activité à mesure de l'augmentation de taille des constituants (représentées en vert dans la figure 5.25 page suivante) sont des déactivations. Une partie de ce réseau de déactivation correspond au réseau de mode par défaut qui est systématiquement désactivé dès lors que le sujet effectue une tâche. Cependant, l'interprétation de ces déactivations n'est pas évidente notamment parce que les mécanismes neuronaux sous-tendus par les déactivations du signal BOLD ne sont pas bien compris (Shmuel et al., 2006).

En plus des régions repérées dans les deux expériences mentionnées ci-dessus,

Langage écrit, expérience (Devauchelle, 2008)
Analyse basée sur le voxel, $p < 0.05$ correction FDR
Sans correction

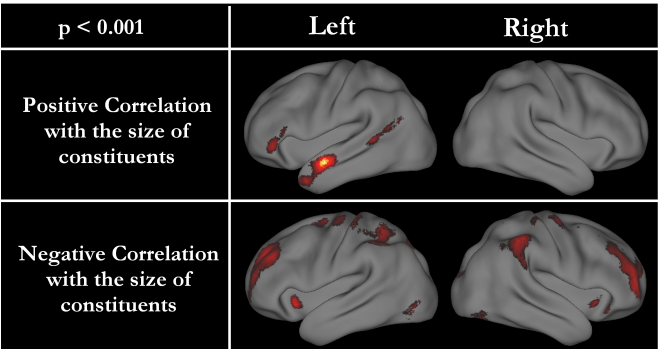


FIGURE 5.24 – Carte statistique représentant les zones sensibles à la structure en langage dans l’expérience en langue écrite (Devauchelle, 2008) à un seuil corrigé au niveau du voxel à $p < 0.05$ FDR.

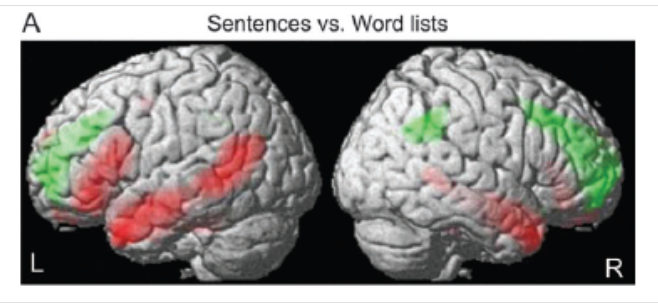


FIGURE 5.25 – Effet de grammaticalité - différence d’activation entre des phrases entières et des listes de mots présentés de façon sérielle rapide visuellement. L’augmentation d’activité pour les phrases par rapport aux listes de mots est représentée en rouge, l’augmentation pour les listes de mots par rapport aux phrases en vert. Analyse cerveau entier. $p < 0.05$ FWE corrigé, taille des clusters de 20 voxels minimum. Extrait de (Snijders et al., 2009)

d’autres régions sont également engagées spécifiquement dans notre expérience : la région frontale médiane supérieure gauche, l’hippocampe et une partie du cervelet de façon bilatérale. L’utilisation de stimuli auditifs en langage répliquent donc essentiellement les expériences avec des stimuli visuels.

Le peu de différences entre le langage écrit et oral confirme les résultats obtenus dans 2 autres études comparant spécifiquement ces 2 modalités et qui montrent un réseau semblable pour la complexité écrite et orale. Constable et ses collègues (Constable

et al., 2004) contrastent des phrases contenant des propositions relative sujet avec des propositions relatives objet. Ils montrent que le réseau qui réagit différemment aux deux types de phrase, à savoir des régions de l'hémisphère gauche comprenant l'IFG, le gyrus temporal supérieur et médian et le gyrus angulaire, est indépendant de la modalité de présentation. Dans une autre étude, Braze et ses collègues (Braze et al., 2011) ont présenté de courtes phrases bien construites ou contenant une anomalie pragmatique, par exemple « * New cars might shrink when washed\ *Les nouvelles voitures peuvent rétrécir lorsqu'elles sont lavées* », ou une anomalie syntaxique, par exemple « * New shirts have shrink when washed.\ *Les nouvelles chemises ont rétrécir lorsqu'elles sont lavées*. Les deux modalités de présentation évoquent le même type de réponses hémodynamiques dans six régions d'intérêt de l'hémisphère gauche (IFG ventral et dorsal, STG, MTG, occipito-temporal, région inférieure pariétale) qui sont plus activées lorsque les phrases contiennent des anomalies.

Les seules différences notables entre notre expérience auditive et celle de (Pallier et al., 2011) en langue écrite se situent dans la présence de l'IFG orb droit, de la zone frontale médiale gauche en langue parlée et de la différence de localisation du pSTS gauche qui est localisé de façon plus antérieure en langue parlé qu'en langue écrite. Lorsqu'on analyse spécifiquement l'activation du pSTS dans nos données, l'amplitude de la réponse hémodynamique au langage est importante mais n'est pas ou peu modulée par la taille des constituants dans l'analyse de groupe, pourtant on retrouve cette zone dans l'analyse par ROIs individuelles.

Cette même zone pSTS gauche est également retrouvée dans une autre expérience, (Humphries et al., 2005). Des sujets ont entendu des phrases ou des mots mélangés ou des listes de mots de contenu dont la prosodie pouvait varier c'est-à-dire ou enregistrés avec une prosodie de phrase ou de liste de mots. Les auteurs montrent que les parties antérieure et postérieure du STS sont plus activées pour les phrases par rapport aux listes de mots. Une autre expérience de la même équipe (Humphries et al., 2006) portant sur l'interaction entre le contenu syntaxique et lexical implique également la partie

antérieure du cortex temporal gauche (aSTS, gyrus temporal inférieur) et la partie antérieure du gyrus médian temporal droit lorsque des sujets entendent des phrases par rapport à des listes de mots, qu'ils soient en langage normal ou en jabberwocky. La jonction temporo-pariétale gauche est spécifiquement sensible à la structure (phrase - liste de mots) en présence de sémantique comme dans l'expérience de (Pallier et al., 2011). Nos résultats sont en accords avec ces deux expériences. Cependant, aucune région inféro-frontale n'est révélée par les contrastes de ces études. L'absence de telle région s'explique par le fait que l'analyse de l'expérience de 2006 ne porte que sur les régions temporo-pariétales du fait des hypothèses de travail de l'équipe. L'expérience de 2005 a, quant à elle, été effectuée sur cerveau entier. Cependant, il existe quelques différences dans les stimuli utilisés : la condition mots de contenu mélangés ne contient que des mots de contenu au contraire de notre expérience, tandis que la condition liste de mots ne contient pas que des constituants de taille 1 comme dans notre expérience. De plus, l'étude ne compte que 12 sujets alors que la nôtre en contient 35 ; cette différence numérique pourrait être à l'origine d'un manque de sensibilité dans l'expérience de 2005 et donc de la mise en évidence de régions plus nombreuses notamment inféro-frontale dans notre cas.

Cette zone inféro-frontale dite de Broca a, en effet, été impliquée dans le traitement de la syntaxe par des expériences de neuro-imagerie et par la neuropsychologie. Cette région de Broca n'est, malgré tout, pas définie de façon consensuelle. Le plus souvent, les chercheurs s'accordent sur le fait qu'elle contient les aires de brodmann numéro 44 et 45 représentée sur la figure 5.26 page suivante.

Dans notre expérience nous mettons en évidence la partie la plus frontale de Broca, la pars triangularis ainsi que la région orbitalis (BA44) toutes deux souvent impliquées dans le traitement syntaxique (Tyler et al., 2010; Moro et al., 2001; Snijders et al., 2009; Pallier et al., 2011). Cependant, l'IFG opercularis est détectée dans d'autres expériences qui testent des structures syntaxiques plus complexes complexes.

Par exemple, Caplan et ses collègues (Caplan et al., 2008), ont présenté visuellement

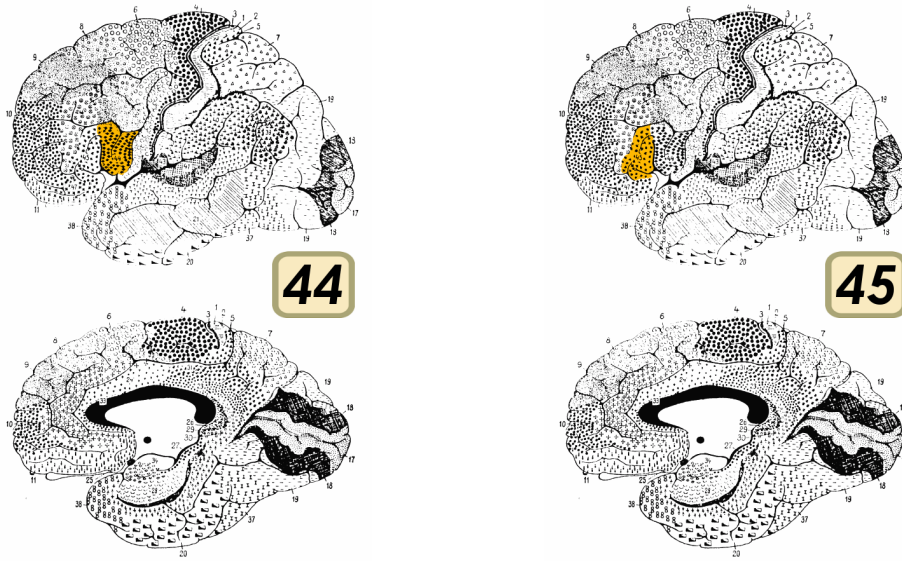


FIGURE 5.26 – Illustration des aires de brodmann 44 et 45 : respectivement IFG opercularis et IFG triangularis.

des phrases contenant des propositions relatives objet ou des propositions relatives sujet à des sujets pendant que leur activité cérébrale était enregistrée en IRMf. Les participants ont réalisé trois tâches très différentes sur ces phrases dans des blocs d'acquisition différents :

- une première de vérification de sens où ils devaient presser un bouton pour dire si la phrase cible correspondait à la phrase de test précédemment entendue ;
- une deuxième de jugement de plausibilité dans laquelle ils devaient dire si la phrase entendue possédait un sens imaginable dans le monde réel ;
- une troisième de détection de non-mots.

Si une zone cérébrale effectue des calculs syntaxiques différents pour ces 2 types de phrases alors elle doit pouvoir être repérée durant les différentes tâches. Les auteurs montrent que la partie postérieure du gyrus frontal inférieur, BA44, la pars opercularis, est plus activée par les phrases contenant des propositions relatives objet, plus complexes, que les relatives sujet, plus simples. Cette zone effectuerait donc un certain type de traitement syntaxique. A la différence de notre expérience, les auteurs ont utilisé

une propriété particulière de la syntaxe, en faisant varier la difficulté de traitement syntaxique par 2 types de phrases sous-tendues par des arbres syntaxiques différents : avec les relatives-objet qui nécessitent une mise en mémoire plus importante que les relatives-sujet. Notre paradigme n'étudie pas ce type de calcul syntaxique. En particulier, nous avons évité toute notion de mouvement et donc de mise en mémoire différentielle due à cette propriété linguistique. Il est donc possible que l'aire BA44 soit spécifiquement associée au traitement de structures syntaxiques plus complexes non testées dans notre expérience. A l'appui de cette hypothèse, plusieurs expériences de Friederici (par exemple (Makuuchi et al., 2009; Friederici et al., 2009)) présentent des phrases dans lesquelles le nombre de propositions avec enchâssement central varie ainsi que la distance entre le sujet et le verbe. Un exemple de tel stimuli est présenté figure 5.27.

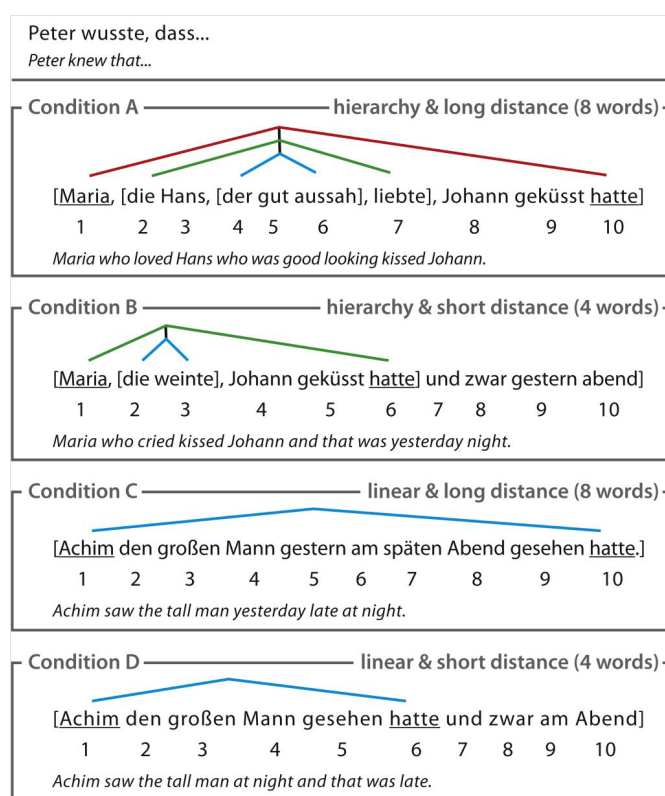


FIGURE 5.27 – Exemples de stimuli utilisés dans l'expérience (Makuuchi et al., 2009). Le nombre de proposition avec enchâssement central est varié pour moduler la structure hiérarchique. La distance entre le sujet et le verbe change pour étudier la mise en mémoire.

Avec cette manipulation, les auteurs espèrent différencier les effets dûs au calcul de structure syntaxique hiérarchique - aux propositions avec enchâssement central - de ceux dû à la mémoire de travail (distance entre le sujet et le verbe). Cependant, les différents types de phrases utilisés font varier plus que simplement l'enchâssement des phrases. En effet, elles sont sous-tendues par des arbres syntaxiques très différents. D'abord, les mouvements qu'elles contiennent sont quantitativement différents, ceci est visible notamment par le nombre de pronoms relatifs des propositions enchâssées. De plus, la mise en mémoire varie elle-aussi, selon le nombre de propositions avec enchâssement central. En outre, rien ne prouve que la distance entre le sujet et le verbe soit le seul élément dans ces phrases qui fasse varier la mise en mémoire de travail. Par exemple les phrases contenant plusieurs propositions avec enchâssement centrale varient également dans la quantité de mémoire de travail qu'elles requièrent. Malgré la covariance des facteurs manipulés, les auteurs montrent que le pSTS et l'IFG opercularis (BA44) bilatéralement pour (Friederici et al., 2009) sont plus activés par les phrases contenant des structures avec enchâssement central dites hiérarchiques par rapport aux structures syntaxique dites linéaires (dont l'arbre syntaxique n'est pas linéaire) qui ne contiennent pas d'enchâssement. Une analyse spécifique (Makuuchi et al., 2009) sur la région d'intérêt l'IFG gauche définie anatomiquement (Amunts et al., 1999) montre une spécificité de BA44, et pas BA45, pour le traitement des structures syntaxiques contenant des proposition avec enchâssement central. Etant donné ces résultats, l'aire BA44 pourrait être nécessaire au traitement de structures syntaxiques complexes comme des phrases contenant des propositions relatives objet ou des propositions avec enchâssement central. Ceci expliquerait pourquoi cette aire n'est pas mise en évidence dans notre expérience ni dans celles étudiant des constructions syntaxiques plus simples.

Il existe d'autres hypothèses quand à la fonction que remplit l'aire de Broca et particulièrement BA44. Ainsi Thompson-Schill (Thompson-Schill, 2005) propose que cette zone serve à sélectionner une information/hypothèse parmi plusieurs sources d'informations ou hypothèses concurrentes. Cette zone ne serait donc pas propre au traitement

du langage. Il faut souligner que cette zone BA44 a une anatomie très variable selon les sujets (Amunts et al., 1999). Premièrement, son volume peut varier selon un facteur 10. Deuxièmement, il existe une asymétrie en faveur de l'hémisphère gauche. Et enfin, on retrouve une différence entre les hommes et les femmes. La variabilité macroscopique de cette zone est également importante et n'est pas sous-tendue par des différences architectoniques. On comprend donc la difficulté d'étudier au niveau du groupe la spécificité des activations de cette zone cérébrale. Dans notre analyse par région d'intérêt individuel, l'activation de BA44 est corrélée faiblement mais de façon significative et positive à l'augmentation de la taille des constituants en langage.

Dans cette même analyse par région individuelle, le pôle temporal gauche correspond à la région dont l'activation croît le plus à mesure que la taille des constituants augmente. L'analyse de groupe confirme cette analyse qui va dans le sens d'une implication de cette aire dans la construction de structure syntaxique comme proposé notamment dans (Snijders et al., 2009; Pallier et al., 2011). Une autre expérience (Brennan et al., 2012) a utilisé des stimuli plus écologiques (écoute d'une histoire) pour étudier le réseau cérébral traitant la structure syntaxique. Brennan et ses collègues ont enregistré l'activité cérébrale (IRMf) pendant que des sujets écoutaient un extrait d'*Alice au pays des Merveilles*. Ils ont ensuite corrélié leurs activations cérébrales avec la quantité de structure syntaxique évoquée en comptant pour chaque mot le nombre de noeuds syntaxiques associés et donc le nombre d'opérations d'intégration successives pour inclure chaque mot dans la structure syntaxique. Dans notre étude, nous avons modélisé ce même nombre d'intégrations successives en augmentant la taille du plus grand constituant syntaxique. Les auteurs montrent l'implication du pôle temporal gauche et d'une région couvrant le gyrus transverse temporal et l'insula de l'hémisphère droit dans le calcul de la quantité de structure syntaxique. Ces résultats sont en accord avec notre expérience. Il faut ajouter que notre expérience montre une corrélation négative entre l'augmentation de la taille des constituants et l'activité cérébrale dans les aires droites mentionnées ci-dessus. Plusieurs facteurs peuvent expliquer le faible nombre de régions impliquées

dans la construction de structure syntaxique. D'une part, la méthode d'analyse utilisée dans cette expérience est différente. Les auteurs n'ont pas effectué d'analyse sur cerveau entier mais ont d'abord sélectionné des régions dont les décours temporels montraient une corrélation entre les sujets. Parmi ces régions, par exemple, seule une petite partie du gyrus frontal inférieur gauche de 9 voxels (à cheval sur BA47 et 45) a été isolée ; sa taille pourrait expliquer l'absence de corrélation de l'activité cérébrale avec le traitement de la structure syntaxique. D'autre part, comme souligné dans l'article, les stimuli sont totalement naturels et ne contiennent aucune violation syntaxique. Au contraire dans notre expérience, il existe une violation syntaxique entre chaque constituant qui marque justement sa limite. Enfin, comme le souligne les chercheurs, cette expérience, comme la nôtre, porte sur des mots existants et l'intégration dans la structure syntaxique est indissociable de la construction de la structure sémantique et de l'accès au lexique. Cependant, les études utilisant des pseudo-mots, comme par exemple (Pallier et al., 2011), évitent la sélection d'un item lexical¹³ et ne permettent pas de calculer de structure sémantique, c'est-à-dire d'assigner un sens aux phrases ou groupes de mots qu'elles étudient. Dans cette expérience en langue écrite où la taille des constituants syntaxiques variait en Jabberwocky, trois zones se distinguaient par leur implication dans le calcul des structures syntaxiques en absence de lexique : l'IFG tro, orb et le pSTS dont l'activité ne semble donc pas liée au contenu lexical. Cette expérience permet de postuler que ces trois zones puissent être le siège des calculs de la structure syntaxique indépendamment du type de stimulus : langage, musique, actions etc.

Réseau commun pour la structure en langage et en musique ?

L'analyse de l'activité cérébrale des régions impliquées dans le calcul des structures musicales, construites sur le même modèle que le langage, a révélé une corrélation positive avec la taille des fragments musicaux au niveau individuel dans les 4 régions du traitement de la structure en constituant en Jabberwocky isolées dans (Pallier et al.,

13. Même si on ne sait pas comment sont traités les pseudo-mots et qu'il n'est pas possible d'affirmer qu'il n'y a pas de recherche dans le lexique pour les pseudo-mots

2011; Rogalsky et al., 2011) : IFG orb, tri, pSTS et pôle temporal (seulement pour (Rogalsky et al., 2011)). Ce résultat va donc dans le sens d'une indépendance de ces zones pour le type de stimuli traités. Cependant, il convient de rappeler qu'au niveau de l'analyse de groupe, seul le pôle temporal gauche présentait une augmentation de son activité lorsque la taille des fragments augmentait. De plus, les activations obtenues pour la musique sont 2 fois inférieures à celles obtenues pour le langage (cf. 5.18 page 176 et 5.19 page 178). Il est donc possible que ces zones soient activées par la structure en musique mais dans une moindre mesure ce qui ne nous permet pas de visualiser le réseau au niveau du groupe à un seuil statistique habituel. Il est aussi possible que la taille des zones recrutées par la structure en musique soit plus faible et que leur localisation soit variable entre les sujets, et peut être entre les groupes de sujets.

Dans l'étude de Rogalsky, sont également testés des stimuli musicaux, de nouvelles mélodies. Les stimuli langagiers et musicaux sont entendus suivant trois vitesses de présentation (lent - 30% plus lent que le normal, normal, rapide - 30% plus rapide que la présentation normale). En augmentant la vitesse de présentation des stimuli et donc en modulant la phase de l'enveloppe, les chercheurs espèrent mettre en évidence des zones de traitement du langage de plus haut niveau. Cette méthode a été proposée pour éviter d'utiliser des violations dont on sait qu'elles augmentent la charge computationnelle artificiellement¹⁴ et pour sélectionner des régions responsables de calculs de haut niveau afin de tester leur implication pour le traitement de stimuli musicaux. Cependant, autant la modulation de la phase de l'enveloppe a été révélée comme étant une propriété importante du traitement du langage (Luo and Poeppel, 2007) notamment pour son intelligibilité, autant, il me semble difficile de partager cette assertion pour la musique. En effet, en accélérant ou décélérant une même mélodie, il me semble important de souligner qu'il est possible que la charge de traitement varie mais que beaucoup d'autres aspects différeront également. Le tempo est une caractéristique importante d'un morceau de musique et grâce à elle, sa perception va être modifiée; un morceau gai sera plus

14. artificiellement dans le sens où personne ne sait si ce même type de computation est requis pour le traitement du langage et qu'il ne serait pas spécifique à la détection de violation et à leur réparation

rapide, un morceau triste sera plus lent. L'émotion et la perception vont donc covarier. Le tempo est une caractéristique importante pour la compréhension et la perception d'un morceau de musique. Il me semble donc qu'en faisant varier ce paramètre en musique on étudie un phénomène très différent de celui mis en évidence en langage. Les contrastes présentés portent sur les zones communes au traitement du langage (jabberwocky) et de la musique bien construits en prenant en compte ou non la variation de la vitesse de présentation des stimuli. Ces zones correspondent à des régions auditives au sein desquelles il est possible de classifier correctement les effets provenant du langage et de la musique. Les auteurs concluent donc que les réseaux sous-tendant les 2 types de stimuli sont différents. Or, on sait que les propriétés temporelles et acoustiques du langage sont différentes, il n'est donc pas étonnant de pouvoir différencier des images provenant de ces 2 types de stimuli dans les aires auditives. De plus cette expérience détruit la structure syntaxique uniquement en langage et n'effectue pas la même manipulation pour les stimuli musicaux, ce qui aurait été crucial pour pouvoir comparer les processus syntaxiques perturbés par le mélange de l'ordre des mots. Les auteurs ne présentent pas d'analyse du traitement de la musique dans les zones sensibles à la structure (phrases entières - phrases mélangés) à part dans 2 zones particulières (antérieur temporal lobe gauche et IFG gauche voir image 5.3 page 151). Les résultats sur l'IFG montrent que cette structure est tout de même activée pour la musique même si dans une moindre mesure que pour le langage, comme dans notre expérience. Il me semble pertinent de conclure que cette première étude comparant musique et langage chez les mêmes sujets, sans tenir compte du fait qu'ils aient étudié la musique ou non, ne permet pas de répondre à la question que nous nous posons à savoir si le langage et la musique utilise en partie les mêmes réseaux lors du calcul de la structure syntaxique et donc de l'intégration d'éléments unitaires.

Une autre étude très récente s'attache à ce problème d'intégration temporelle en musique et en langage (Abrams et al., 2011). Les auteurs ont fait écouter du langage (extraits de discours) et de la musique (mélodies plus ou moins connues) dans des ver-

sions intactes ou mélangées. Des fragments de moins de 350ms ont été recombinaés, le choix empirique de 350ms au maximum permet de préserver les segments phonétiques et les mots courts dans le langage ainsi que les notes individuelles en musique. La taille des fragments variaient de façon à éviter de recréer une pulsation. La structure temporelle en langage et en musique a ici été étudiée de la même façon. L'analyse univariée comparant le traitement de la structure temporelle en langage et en musique¹⁵ ne révèle pas de différence dans le traitement de la structure temporelle des deux domaines, et ce même à des seuils statistiques libéraux comme $p < 0.05$. Cette étude ne porte que sur des non-musiciens. En effectuant la même analyse pour notre expérience et en restreignant au groupe des non-musiciens, seul un groupe de 41 voxels dans la région pariétale inférieure est plus sensible à la structure en musique qu'en langage. Cette même analyse effectuée sur nos 2 groupes de sujets (soit 35 sujets), révèle 4 régions de l'hémisphère gauche qui sont plus activées par la structure en langage qu'en musique (gyrus angulaire ou pariétal inférieur, TP, aSTS et gyrus médian supérieur, cf figure 5.15 page 171). Le nombre de sujets et le fait qu'ils soient musiciens ou non ainsi que l'existence de niveaux de structures paramétriques, peuvent expliquer ces différences de résultats. En plus de l'analyse univariée, (Abrams et al., 2011) ont effectué une analyse de classification de pattern multivariés au sein des zones sensibles au langage et à la musique (toutes les conditions par rapport au repos). Ils montrent qu'il est possible de classer les images résultantes de l'effet de structure (entière - totalement mélangée) selon le domaine. Cette classification est bonne dans les régions gauches inféro-frontales (BA44, 45, 47), bilatéralement dans les pôles temporaux, pSTG, MTG, le gyrus de Heschl, le planum temporale et polaire¹⁶. Parmi ces régions, le pSTG gauche, l'aSTG et aMTG droits sont les régions pour lesquelles la classification est la plus efficace. Ces résultats sont en accord avec nos données puisque ces mêmes régions de l'hémisphère droit (aSTG et mSTG) sont activées par la structure en musique et pas pour la structure linguistique (cf. analyse de groupe figure 5.13).

15. (musique - musique recombinaée) et (langage - langage recombinaé) sont ici comparés

16. Malheureusement, la classification selon le niveau de structure n'est pas rapportée, on ne sait donc pas si les activations pour la structure au sein de chaque domaine sont classifiables.

Les analyses par régions d'intérêt individuelles soulignent que les régions inféro-frontales et temporales gauches sont marginalement ou pas utilisées chez les non-musiciens pour le traitement de la structure en musique à la différence du langage. Nos résultats sont en accord avec la littérature pour les non-musiciens. Quant à l'effet de structure en musique chez les musiciens, il semble que les aires sensibles à la structure en langage soient recrutées pour le traitement de la structure en musique (IFG orb, tri, TP, pSTS gauches), une légère tendance est néanmoins visible chez les non-musiciens dans les zones inféro-frontales.

Notre étude est la première à étudier l'effet de la structure en musique et en langage en comparant musiciens et non-musiciens.¹⁷ Nous montrons que les musiciens et les non-musiciens ne recrutent pas les mêmes aires cérébrales pour calculer la structure temporelle en musique. Les mêmes zones sont engagées par le traitement de la structure linguistique et musicale pour les musiciens, à savoir l'IFG orb et tri, le TP et le pSTS gauches. De plus, l'analyse de groupe a identifié un réseau dont l'activité décroît à mesure que la taille des fragments augmente. Ce réseau sensible négativement à la structure présente des similitudes avec le réseau du mode par défaut voir figure 5.28 page suivante.

Une analyse utilisant la classification de pattern multivariés est actuellement en cours sur nos données dans le but d'approfondir les résultats existants et de chercher si une classification selon le niveau de structure et le domaine est possible (classification selon 10 condition) : 5 tailles de constituants en langage et 5 tailles de fragments en musique.

Il serait intéressant d'étudier la réponse à la structure au langage dans les zones sensibles à la structure à la musique. Malheureusement, notre manipulation de la structure en musique ne nous permet pas d'identifier des régions d'intérêt, tant au niveau individuel, que dans l'analyse de groupe, sur lesquelles baser une analyse par régions d'intérêt individuelles. Il serait possible d'utiliser les régions d'intérêt mises en évidence dans l'expérience d'Abrams pour effectuer cette analyse.

17. Le plus souvent les études qui étudient les différences entre musiciens et non-musiciens s'intéressent à la perception de la hauteur des notes. par exemple (Foster and Zatorre, 2010b)

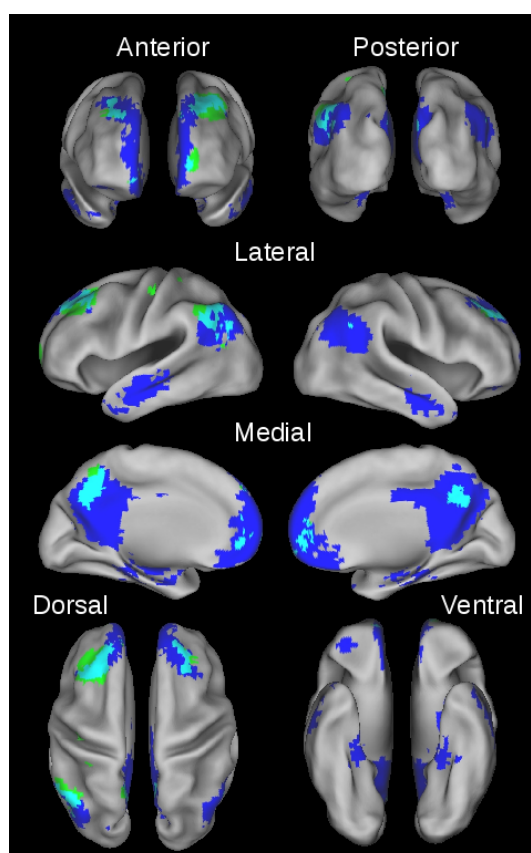


FIGURE 5.28 – Superposition du réseau du mode par défaut (en bleu) et du réseau sensible négativement à la structure chez les musiciens (en vert, $p < 0.001$ non corrigé). Les zones en cyan correspondent aux zones communes.

Il est possible cependant que notre manque de sensibilité quand à la structure en musique puisse être dû à la manipulation de la structure temporelle dans nos stimuli. En effet, nous avons modifié de façon conjointe plusieurs types de structure en musique (harmonique, rythmique particulièrement structure de groupement etc.). Dans le but de comprendre les différents calculs syntaxiques réalisés lors de l'écoute et analyse d'un morceau de musique, il serait intéressant de manipuler séparément les différents composants structuraux de la musique. Il est possible de construire des stimuli musicaux en ne manipulant qu'un seul type de structure. Par exemple, il est possible de ne faire varier que la structure mélodique : la façon dont la hauteur des notes sont reliés les unes aux autres ; en ne touchant pas à la structure rythmique par exemple en égalisant les durées des notes. Une autre étude pourrait ne s'intéresser qu'à la structure rythmique et ainsi

faire abstraction des hauteurs de notes en utilisant une seule et même note ou un son de percussion.

5.3 Conclusion

Nos résultats sont en accord avec la littérature tant pour le traitement de la structure en langage qu'en musique. Certaines zones seraient recrutées par les deux modalités mais de façon différente. Nous proposons, en effet, qu'un réseau comprenant l'IFG orb, tri, le pôle temporal et le pSTS de l'hémisphère gauche serait engagé dans le traitement de la structure en langage et en musique particulièrement par les musiciens. A l'appui de cette hypothèse, plusieurs études basées sur la connectivité structurelle ((Catani et al., 2005; Anwander et al., 2007; Saur et al., 2008) ou encore (Friederici, 2009) pour une revue à ce sujet) montrent qu'il existe des tractus de fibres connectant les régions inféro-frontal et le cortex temporal dans sa partie antérieure et postérieure comme sur la figure 5.29.

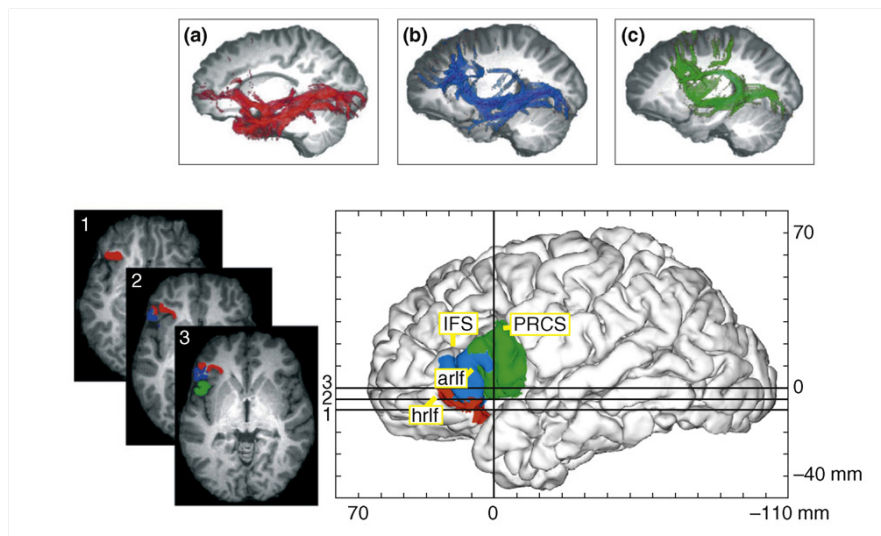


FIGURE 5.29 – Carte des territoires de la zone inféro-frontale et sa connectivité. Exemple pour un sujet. En rouge, une région à cheval entre BA45 et BA47, en bleue l'IFG triangularis (B45) et en vert l'IFG opercularis BA44 avec les fibres de matières blanches qui les connectent au cortex temporal et pariétal. Image provenant de (Friederici, 2009)

Ces zones inféro-frontales et temporelles seraient responsables de l'intégration d'éléments pour former une structure syntaxique indépendamment du sens, comme dans

l'expérience de (Pallier et al., 2011). Nos données pointent vers le fait que ce réseau serait recruté de façon différente par le langage et la musique. Même si notre effet de structure en musique n'est que peu ou pas visible chez les non-musiciens, les données de la littérature (Abrams et al., 2011) nous permettent d'espérer qu'une analyse utilisant la classification de pattern multivariés pourrait améliorer la sensibilité de nos analyses. De façon générale les musiciens tendent à recruter les mêmes zones pour le traitement du langage et de la musique.

Résumé du chapitre

- Nous répliquons avec des stimuli linguistiques à l’oral les résultats obtenus dans l’expérience en langue écrite (Pallier et al., 2011). Les mêmes réseaux sont impliqués dans le traitement de la structure en constituant en langage oral et en langage écrit.
- Le réseau sensible de façon positive à la structure en constituant en langage contient l’IFG gauche Tri et orb, aSTS, pSTS et gyrus angulaire, putamen avec en plus le pôle temporal et IFG orb droit et l’hippocampe bilatéralement. Les réseaux sensibles de façon négative à la structure comportent des aires frontales médiales bilatérales, des régions bilatérales cingulaires antérieures et médianes, le precuneus et une région pariétale inférieure à gauche, une partie du cervelet, le gyrus frontal inférieur opercularis droit, une région supramarginale gauche.
- Le réseau sensible à la quantité de structure indépendamment du sens (jabberwocky) à savoir l’IFG orb, tri et pSTS de l’hémisphère gauche (proposé par (Pallier et al., 2011)) est aussi sensible à la structure en musique ainsi que le pôle temporal gauche. Nous permettant d’aller dans le sens que ce réseau traiterait la structure indépendamment du domaine mais dans une mesure différente.
- Les musiciens et les non-musiciens ont des activations différentes quand au traitement de la structure en musique.
- Les musiciens ont tendance à recruter le même réseau pour traiter la structure en musique et en langage.

Discussion générale

Ce travail de thèse s’articule autour de deux axes :

1. l’implication de la structure syntaxique et de la prosodie dans l’accès au lexique
2. les bases neurales du traitement de la structure syntaxique linguistique et musicale.

Nous avons montré dans une première série d’expériences comportementales que les mots grammaticaux et la prosodie sont exploités en temps réel par les enfants de 18 mois et les adultes pour contraindre l’accès lexical. Les enfants détectent un mot s’il est présenté dans le contexte propre à sa catégorie syntaxique (nom ou verbe). Les adultes de la même façon repèrent un pseudo-mot (par exemple *la crépouille*) s’il est présenté dans un contexte syntaxique et prosodique congruent avec sa catégorie (par exemple *Tu ne douras pas le nabalon que la crépouille te maleille*). Nous avons proposé un modèle d’accès lexical qui utilise simultanément ces deux types d’indices ainsi que les premiers phonèmes du mot pour sélectionner un sous-ensemble d’items lexicaux probables.

La deuxième partie de ce manuscrit explore les bases neurales du traitement de la structure syntaxique en langage et en musique.

Dans une première expérience d’IRMf, nous avons fait varier la taille et la position des frontières prosodiques dans des suites de mots coordonnés, comme par exemple « Le chien ou la poule et la vache ou le lapin ». Ainsi, plusieurs structures syntaxiques ont pu être produites à partir des mêmes mots exactement. Les deux analyses menées sur les données fonctionnelles enregistrées lors de l’écoute de tels stimuli impliquent un réseau de l’hémisphère gauche composé de régions frontales, temporales et temporo-pariétale dans le traitement des différentes structures syntaxiques. La complexité prosodique est

traitée par un réseau d'aires frontales et temporales de l'hémisphère droit et gauche. Dans une seconde expérience d'IRMf, nous avons examiné si les réseaux engagés dans le traitement de la structure syntaxique en langage étaient également impliqués dans le traitement des structures musicales. Un réseau de l'hémisphère gauche contenant des aires inféro-frontale, temporale et temporo-pariétale est d'autant plus activé que la taille des constituants linguistiques augmente. Les résultats obtenus avec le groupe de musiciens montrent qu'une partie de ce réseau est recrutée pour traiter la structure musicale. De plus, les régions inféro-frontales montrent aussi une tendance à l'activation en fonction de la structure musicale chez les non musiciens. Ces zones pourraient être le siège de l'intégration d'éléments sous forme de structure hiérarchique, que ce soit en musique, en langage ou dans d'autres domaines.

Nous discuterons plus précisément les résultats obtenus dans chacun des deux axes de cette thèse dans les paragraphes suivants.

6.1 Mots grammaticaux et prosodie dans l'accès lexical

Nous avons d'abord montré par une expérience de conditionnement de l'orientation du regard, que les enfants de 18 mois utilisent les mots grammaticaux pour reconnaître un mot. En effet, les enfants repèrent un nom lorsqu'il est précédé d'un déterminant mais pas s'il est précédé d'un pronom. De plus, ils détectent un verbe s'il est précédé d'un pronom mais pas d'un déterminant. Les enfants, à l'âge où ils sont en train d'apprendre le sens des mots, exploitent donc le contexte syntaxique des mots, en temps réel, pour contraindre leur accès au lexique. Dans la deuxième expérience, nous avons montré que les adultes utilisent aussi la catégorie syntaxique des mots pour contraindre leur accès au lexique. Ainsi, si on les pousse à répondre rapidement, ils effectueront des erreurs sur des mots commençant par la même syllabe que le mot cible si et seulement s'il partage la même catégorie. Par exemple, lorsqu'ils devaient détecter le mot *chat*, ils se sont trompés en entendant le mot *chapeau* mais pas le verbe *chatouiller*. Ces données sont en accord avec la théorie de la cohorte de Marslen-Wilson ([Marslen-Wilson, 1987](#)). Une cohorte

de mots probables est formée dès les premiers phonèmes entendus. Puis s'effectue une sélection parmi les mots de la cohorte selon des informations phonologiques (les premiers phonèmes) mais aussi selon des indices syntaxiques (catégorie grammaticale et genre). Nos deux premières expériences montrent que la catégorie des mots (nom ou verbe) est effectivement utilisée. D'autres expériences montrent que le genre influence aussi l'accès lexical (Dahan et al., 2000; Lew-Williams and Fernald, 2010). Nous proposons que ces indices phonologiques et syntaxiques contraignent simultanément la sélection de l'item lexical. La figure 6.1 page suivante formalise nos résultats. La morphosyntaxe et la prosodie permettent d'effectuer une prédiction quant à la catégorie syntaxique de l'item à venir. Cette information syntaxique est combinée avec les premiers phonèmes entendus pour sélectionner le ou les items les plus probables dans le lexique. Cette double source d'informations phonologiques et syntaxiques expliquerait pourquoi les homophones ne présentent pas d'ambiguïté pour les adultes qui ne confondent pas le nom *porte* dans *la porte* et le verbe *porter* dans *je la porte*.

Si ces informations syntaxiques servent à restreindre la recherche lexicale à un sous ensemble d'items (commençant par la bonne syllabe, de la catégorie grammaticale correcte et d'un genre congruent), il faut que ces propriétés soient encodées dans le lexique. Bock et Levelt (Bock, 1994) postulent qu'un item lexical est encodé dans le lexique mental sous la forme de trois informations : le sens du mot, ses propriétés syntaxiques (genre, catégorie grammaticale) et ses propriétés phonologiques. Nos données s'inscrivent donc dans ce cadre théorique qui permet, en outre, de faire des prédictions concernant l'acquisition du lexique. Si les items sont encodés de cette façon dans le lexique mental, il faut donc que les enfants qui découvrent le sens des mots stockent non seulement leur forme phonologique mais aussi leurs informations syntaxiques¹.

1. Ou les enfants encoderaient simultanément les trois informations lors de la découverte d'un nouveau mot, ou bien ils pourraient le faire en plusieurs étapes. Mais en fin de compte, ils devraient encoder les trois informations.

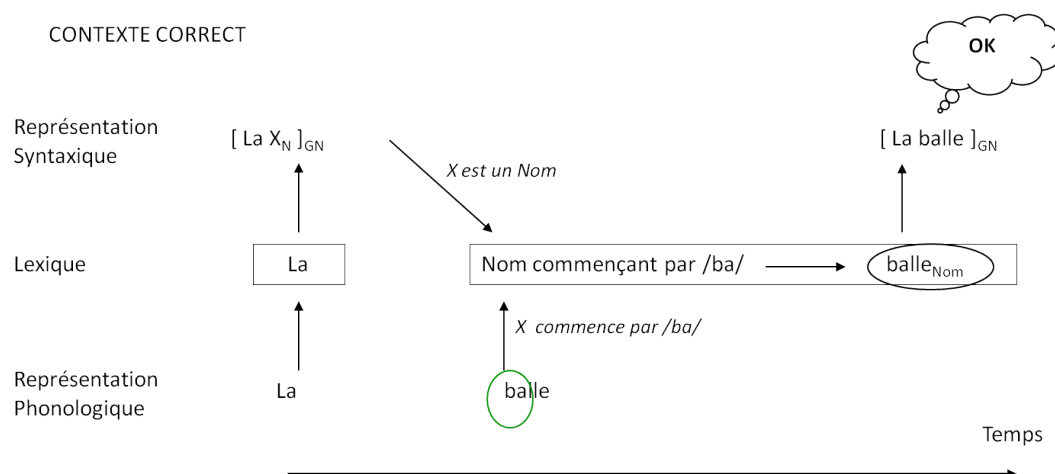


FIGURE 6.1 – Modèle d'accès au lexique dans lequel les indices phonologiques et syntaxiques contraignent simultanément la recherche de l'item.

6.2 Mots grammaticaux et prosodie dans l'acquisition lexicale

Nous avons montré dans l'expérience 1 que les enfants de 18 mois se servent déjà des mots grammaticaux pour contraindre leur recherche lexicale. Or, le calcul des catégories syntaxiques serait particulièrement avantageux pour contraindre le sens de mots inconnus. En effet, savoir qu'un mot est un nom ou un verbe permet de restreindre le nombre d'éléments possibles pour une forme phonologique inconnue. Un nom réfère le plus souvent à un objet tandis qu'un verbe représente typiquement une action (Pinker, 1984). Dans l'expérience 3 (section 1.7 page 51), nous avons placé des adultes dans une situation proche de celle rencontrée par les enfants qui cherchent le sens des mots. Les participants ont entendu des phrases en Jabberwocky, langue dans laquelle la phonotactique, la morphosyntaxe et la prosodie sont conservées et où tous les mots de contenu sont remplacés par des pseudo-mots. Ils devaient détecter un pseudo-mot dont la catégorie grammaticale était précisée. Même lorsque ce pseudo-mot était précédé d'un mot de fonction ambigu comme *le/la/les* (qui peut être soit un déterminant soit un clitique objet), les sujets ont calculé sa catégorie syntaxique en temps réel. Ils s'appuient donc sur d'autres informations syntaxiques et prosodiques que le seul mot grammatical précédant

directement le pseudo-mot cible pour calculer sa catégorie syntaxique.

La dernière expérience comportementale en Jabberwocky (section 1.8 page 61) révèle que la prosodie, et particulièrement les frontières de groupe intonational (prosodie de liste), sont informatives pour le calcul de ces catégories. Nos données en Jabberwocky corroborent l'hypothèse selon laquelle il est possible de construire le squelette syntaxique d'une phrase dont les mots de contenu sont inconnus en s'appuyant sur les indices morphosyntaxiques et prosodiques (Christophe et al., 1997; Christophe et al., 2008). Etant donné ces résultats et comme, jusqu'à preuve du contraire, on suppose que les enfants et les adultes partagent une même architecture cognitive, les enfants pourraient mettre en jeu les mêmes mécanismes pour catégoriser des mots inconnus. On sait que les enfants de deux ans attribuent un sens différent à un même mot selon le contexte syntaxique dans lequel il est placé. Ainsi, le mot *dase* sera interprété comme un objet lorsqu'il est présenté précédé d'un déterminant, il sera considéré comme action s'il suit un pronom (Bernal et al., 2007). De plus, les potentiels évoqués par des mots nouvellement appris² (par exemple *le touse* est un tricératops) sont différents selon qu'ils sont employés dans un contexte syntaxique correct (comme *L'indien pousse le touse*) par rapport à un contexte incorrect (*Martin le touse, le poisson.*) (Brusini, 2012). Donc, les enfants effectuent des traitements cérébraux différents lorsqu'ils repèrent un mot dans un contexte dans lequel il peut effectivement apparaître par rapport à un contexte où il ne le peut pas. De plus, ils ont été capables de généraliser l'apparition de ce nouveau mot dans des contextes grammaticaux jamais rencontrés³. Cette expérience EEG implique que les enfants de 24 mois ont calculé en temps réel la catégorie syntaxique de ces nouveaux mots, montrant ainsi que l'information de catégorie syntaxique a déjà été encodée avec la forme phonologique et le sens dans le lexique.

Les enfants testés dans ce paradigme d'apprentissage avait déjà deux ans, leur lexique contient déjà un nombre important d'items. Une expérience testant l'assignation du sens à un mot inventé par des enfants plus jeunes (entre 12 et 18 mois) dont le lexique mental

2. une semaine avant. Deux nouveaux noms et deux nouveaux verbes ont été appris.

3. puisque l'expérimentatrice n'avait jamais prononcé la suite de mots utilisée dans les phrases test : « *le touse* »

est de taille plus limitée permettrait de révéler un traitement similaire de la prosodie et de la morphosyntaxe dans l'apprentissage du sens des mots.

Nos données comportementales chez les adultes et les données de la littérature sur l'apprentissage de nouveaux mots par de jeunes enfants montrent que les enfants calculent en temps réel la catégorie syntaxique des mots. Il est envisageable de représenter l'apprentissage du lexique mental selon plusieurs étapes concomitantes. D'abord, les enfants réussissent à assigner un sens à quelques mots parmi les plus fréquents (Bergelson and Swingley, 2012) entre 6 et 9 mois. Ces premiers mots seraient utilisés pour repérer des régularités morphosyntaxiques et prosodiques qui conduiraient à l'établissement d'un squelette syntaxique. Ce squelette permettrait un calcul automatique des catégories syntaxiques des mots inconnus, restreignant ainsi le nombre de sens possibles pour un nouveau mot.

6.3 Les réseaux neuronaux recrutés pour la construction de la structure syntaxique sont-ils spécifiques au langage ?

Dans deux expériences d'IRMf (expériences 4 et 8), nous avons manipulé la structure syntaxique de façon différente. La première expérience repose sur un paradigme utilisant des mots ou groupes de mots coordonnés qui permet de varier orthogonalement la structure syntaxique et prosodique. La seconde modifie la taille des constituants linguistiques et musicaux de façon paramétrique. Toutes deux impliquent des régions de l'hémisphère gauche dans le traitement de la structure syntaxique : le gyrus frontal inférieur orbitalis (BA47) et triangularis (BA45), la jonction temporo-pariétale (ou gyrus angulaire), le précunéus et le sillon temporal supérieur antérieur. Ce réseau est mis en évidence par de nombreuses expériences contrastant des phrases bien construites avec des listes de mots ou des mots mélangés (sans structure syntaxique) (Pallier et al., 2011; Santi and Grodzinsky, 2010; Snijders et al., 2009; Humphries et al., 2006; Humphries et al., 2005;

Rogalsky et al., 2011). En outre, une partie de ce réseau comprenant le gyrus frontal inférieur gauche orbitalis et triangularis et la région postérieure du sillon temporal supérieur gauche, a été impliquée dans le traitement de la structure en constituant sans sémantique (en Jabberwocky) (Pallier et al., 2011). Cette réponse indépendamment du sens désigne ces zones comme de potentielles aires construisant la structure hiérarchique en général. Elles combindraient des éléments qu'ils soient linguistiques ou non. La partie musicale de l'expérience 8 avait donc pour but, de tester la spécificité au langage de ces régions. Les stimuli ont été construits empiriquement selon le même schéma en langage et en musique. Nous avons créé des fragments de taille croissante qui ont ensuite été combinés pour former des stimuli avec une quantité de structure variable. La perception des différents niveaux de structure a été vérifiée grâce à des expériences comportementales décrites dans le chapitre 4 page 119.

Nos données d'imagerie fonctionnelle nous permettent de montrer que les musiciens recrutent en partie les mêmes régions pour traiter la structure en langage et en musique. En effet, les régions inféro-frontales et le pSTS ainsi que le pôle temporal gauche sont impliquées indépendamment du domaine (figure 6.2 page suivante), même si le niveau d'activation cérébrale est différent pour les deux modalités. Le signal BOLD est plus important dans les conditions linguistiques par rapport à leurs homologues musicales. Une tendance est également observée chez les non musiciens dans les aires inféro-frontales dans l'analyse par régions d'intérêt individuelles (1 et 2 sur la figure 6.2 page suivante). L'analyse de groupe chez les non musiciens révèle que l'activité de l'IFG orbitalis ou BA47 (1 sur la figure 6.2 page suivante) est modulée par la taille des fragments musicaux. Cette même région a été mise en évidence dans le traitement de la structure musicale chez des non musiciens (Levitin and Menon, 2003; Abrams et al., 2011). Elle est retrouvée également dans le traitement des structures prosodique et syntaxique testées dans l'expérience 4. Cette aire cérébrale dont la réponse varie avec la quantité de structure indépendamment du domaine testé participerait à la construction de stimuli structurés qu'ils soient linguistiques ou musicaux sans intégrer de sémantique. Cette

région semble être une zone de création de structure multi-domaine.

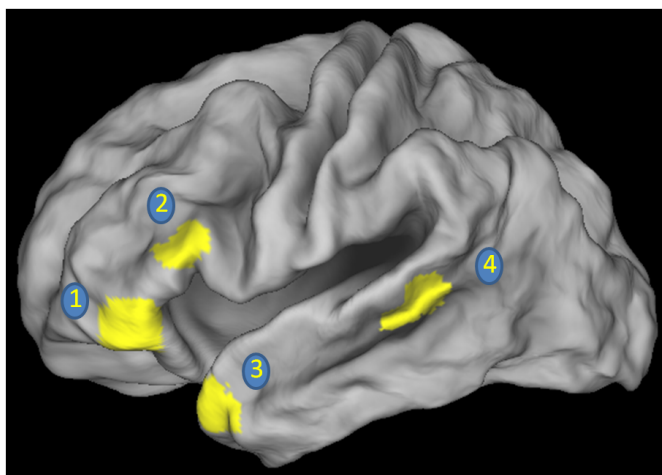


FIGURE 6.2 – Régions cérébrales impliquées dans le traitement de la structure syntaxique en langage et en musique. 1. IFG orbitalis (BA47) ; 2. IFG triangularis (BA45) ; 3. Pôle temporal ; 4. Sillon temporal supérieur postérieur.

Notre expérience révèle des différences dans le traitement de nos stimuli musicaux entre les musiciens et les non musiciens notamment dans le niveau de réponse à la structure⁴. D'une part, elles pourraient s'expliquer par des différences de compétences dans le traitement de la musique. Les musiciens étant habitués à trouver de la structure dans la musique y seraient attentifs. A contrario, les non musiciens ne la remarqueraient pas systématiquement. Cependant, l'étude comportementale de notation de nos stimuli par des non musiciens [4.1 page 120](#) montre qu'ils sont capables de percevoir la déstructuration progressive de nos stimuli. Ils traitent donc la structure de la musique lorsqu'on leur demande d'y porter attention. Ce qui nous conduit à proposer une explication alternative aux différences fonctionnelles observées. Elles pourraient, en effet, être dues à des différences d'automatisme de traitement de la musique entre les deux populations. Ainsi, pour les musiciens, experts dans ce domaine, la structure syntaxique serait construite de façon automatique ; tandis que chez les non musiciens sa construction serait optionnelle. En effet, dans notre expérience, rien ne force les sujets à utiliser la structure en musique. Il est donc possible que les non musiciens aient écouté les stimuli sans pour

4. Données présentées dans l'analyse par région d'intérêt individuelles [5.22 page 183](#)

autant construire leur structure. Pour tester cette hypothèse, une expérience forçant la construction d'une structure syntaxique musicale serait nécessaire. Si les sujets y étaient contraints, nous postulons que les mêmes zones inféro-frontales (IFG orbitalis et triangularis), le pôle temporal et la partie postérieure du sillon temporal supérieur (voir figure 6.2 page ci-contre) seraient de plus en plus recrutées à mesure que la quantité de structure augmenterait.

Notre travail et celui d'Abrams (Abrams et al., 2011) sont les deux seules expériences étudiant chez les mêmes sujets les mêmes manipulations structurales en langage et en musique. Elles appuient la théorie de partage des ressources syntaxiques (SSIRH - shared syntactic integration resource hypothesis (Patel, 2003)) puisqu'un réseau d'aires inféro-frontales et temporales sont recrutées pour construire la structure en langage et en musique. Ces régions pourraient donc intégrer sous forme de structure hiérarchique les différentes représentations, qu'elles soient musicales ou linguistiques, et pourquoi pas d'autres domaines comme les mathématiques ou l'informatique. Les représentations propres à chaque domaine seraient stockées dans différentes zones cérébrales, ce qui expliquerait qu'il soit possible d'être amusique sans être aphasique. Les ressources cérébrales recrutées par le langage et la musique pour la construction de leur structure ne semblent cependant pas utilisées de la même façon. En effet, d'une part notre expérience montre des niveaux d'activation différents dans ces régions pour les deux domaines. D'autre part, l'étude d'Abrams révèle qu'il est possible de discriminer les images cérébrales issues du contraste structural en langage, de celles issues du même contraste pour la musique. Même si nous montrons que les mêmes régions anatomiques sont recrutées pour la structure linguistique et musicale, l'imagerie par résonance magnétique ne nous permet pas d'exclure qu'il s'agisse de réseaux indépendants mais localisés au même endroit. Des sous-populations neuronales d'une même région seraient recrutées indépendamment l'une de l'autre. Cependant, les expériences comportementales (Fedorenko et al., 2009; Slevc et al., 2009) qui montrent que les processus cognitifs impliqués dans le traitement de la structure musicale intragissent avec ceux du langage, ainsi que

notre expérience d'imagerie et celle d'Abrams nous permettent de penser qu'une partie de ces réseaux pourrait être commune.

Pour étudier plus avant la spécificité du réseau activé lors de la construction de la structure syntaxique, il serait possible d'utiliser une approche un peu différente de celle proposée dans ce travail en partant des réseaux isolés pour le traitement de la structure dans d'autres domaines et en testant leur implication en langage. Ainsi, en appliquant cette idée à nos données, il serait pertinent de partir de régions répondant à la structure en musique et d'effectuer une analyse en région d'intérêt individuelle sur l'activation pour la structure linguistique. Malheureusement, notre paradigme musical n'est pas assez sensible pour isoler des régions d'intérêt pour la structure musicale. Il serait possible d'utiliser des régions d'intérêt isolées par d'autres expériences comme par exemple celle de Levitin et Menon (Levitin and Menon, 2003) et d'appliquer nos analyses sur nos données linguistiques.

Une autre piste envisagée serait de cibler un type de structure musicale comme par exemple l'harmonie ou le rythme qui dissocierait les différents aspects structuraux en musique. Plusieurs études ont ainsi impliqué l'aire de Brodmann 47 (IFG orbitalis) dans le traitement des structures rythmiques chez des sujets musiciens (Vuust et al., 2006; Vuust et al., 2011). Il serait donc pertinent de tester la structure hiérarchique rythmique de façon paramétrique pour mieux spécifier les réseaux neuronaux sous-tendant la création de ce type de structure. L'utilisation de la structure hiérarchique dans d'autres domaines comme les mathématiques, la planification de l'action, l'informatique, permettrait d'aller plus loin quant à la spécificité du réseau neuronal mis en évidence dans le traitement de la structure linguistique et musicale et de répondre à la question de l'existence d'un tel réseau amodal qui construirait des structures hiérarchiques à partir de n'importe quelle représentation.

En conclusion, dans cette thèse j'ai étudié les structures hiérarchiques, en langage et en musique. D'une part, j'ai démontré que les précurseurs de cette structure hiérarchique existaient déjà à l'âge de 18 mois, à savoir la construction de catégories syntaxiques -

Nom/Verbe -, et l'utilisation du contexte syntaxique pour l'accès au lexique. D'autre part, j'ai montré qu'il existait des régions cérébrales communes, chez l'adulte, qui sont impliquées dans le traitement des structures hiérarchiques à la fois en langage et en musique. Il est possible que ces mêmes zones cérébrales soient également impliquées dans le traitement des structures hiérarchiques dans d'autres domaines, comme les mathématiques, la planification de l'action, ou l'informatique. La capacité à représenter et à traiter des structures hiérarchiques récursives, du type de celles que nous avons étudiées ici, a été proposée comme étant potentiellement la caractéristique spécifique à l'homme qui permettrait, entre autres, l'émergence du langage humain (Hauser et al., 2002). La musique est elle aussi une capacité spécifiquement humaine. Il serait passionnant, dans des études futures, de tester les bases cérébrales de capacités susceptibles d'être partagées avec d'autres animaux, comme la planification de l'action, afin de déterminer d'une part si elles impliquent les mêmes zones cérébrales, et d'autre part si les animaux sont ou non capables d'exploiter des structures hiérarchiques dans ces domaines, et si oui à l'aide de quel substrat neural.

Appendices

Stimuli utilisés dans le chapitre 1

A.1 Phrases de test utilisées dans l'expérience comportementale 1 sur des enfants de 18 mois

Cible-Nom : *balle*

Contexte Correct

1. La balle est rouge et verte
2. C'est la balle de mon cousin
3. La balle de Paul fait du bruit
4. Des balles sont tombées par terre
5. J'ai acheté des balles de tennis
6. Il veut des balles rebondissantes
7. Une balle a roulé dans l'herbe
8. J'ai vu une balle sous le lit
9. Je voudrai une balle pour ma fête
10. Les balles de ping pong sont légères
11. J'adore les balles en mousse
12. Les balles flottent dans la piscine

Phrases Distractrices

1. La fraise est belle et fraîche
2. C'est la vache de mon père
3. La chaise de Jeanne est en bois
4. Des vaches sont tombées dans l'eau
5. J'ai mangé des fraises à la crème
6. Elle veut des chaises très confortables
7. Une fraise a poussé à l'ombre
8. J'ai pris une chaise pour m'asseoir
9. Elle voudrait une vache pour jouer
10. Les chaises de Mamie sont belles
11. J'adore les fraises au sucre
12. Les vaches marchent très lentement

Contexte Incorrect

1. Il balle une petite pomme
2. Pierre, il balle du chocolat
3. Il balle une glace au citron
4. Je balle de la tarte aux fraises
5. A midi, je balle des saucisses
6. Ce soir, je balle au restaurant
7. On balle une pizza au thon
8. Ce soir, on balle chez Mamie
9. Aujourd'hui, on balle du poisson
10. Tu balles des gâteaux à la crème
11. Demain, tu balles chez moi
12. Tu balles bien tes légumes verts

Cible-Verbe : *mange***Contexte Correct**

1. Il mange une petite pomme
2. Pierre, il mange du chocolat
3. Il mange une glace au citron
4. Je mange de la tarte aux fraises
5. A midi, je mange des saucisses
6. Ce soir, je mange au restaurant
7. On mange une pizza au thon
8. Ce soir, on mange chez Mamie
9. Aujourd'hui, on mange du poisson
10. Tu manges des gâteaux à la crème
11. Demain, tu manges chez moi
12. Tu manges bien tes légumes verts

Phrases distractrices

1. Il donne un livre à Pierre
2. Max, il marche à reculons
3. Il chante un air très joli
4. Je donne des bonbons au chien
5. Quand ça glisse, je marche lentement
6. Ce soir, je chante à l'opéra

7. On chante à la fête foraine
8. Ce soir, on donne des gâteaux
9. A la plage, on marche sur le sable
10. Tu donnes des cadeaux à ton frère
11. Demain, tu chantes chez Paul
12. Tu marches seul dans la forêt

Contexte Incorrect

1. La mange est rouge et verte
2. C'est la mange de mon cousin
3. La mange de Paul fait du bruit
4. Des manges sont tombées par terre
5. J'ai acheté des manges de tennis
6. Il veut des manges rebondissantes
7. Une mange a roulé dans l'herbe
8. J'ai vu une mange sous le lit
9. Je voudrai une mange pour ma fête
10. Les manges de ping pong sont légères
11. J'adore les manges en mousse
12. Les manges flottent dans la piscine

A.2 Phrases pour l'expérience comportementale 2

Items Nom

1. Père
Aujourd'hui, mon père est parti en voyage.
Il s'inguiète du bonheur de son père.
Le personnel est toujours aussi radieux.
Je permettrai qu'il sorte enfin de l'hôpital demain.
2. Mot
Juste un mot avec vous avant de m'en aller.
Il avait ce mot sur le bout de la langue.
Non, ce n'était pas le moment approprié.
De temps en temps, nous nous moquons de ces étranges manies.
3. Gars
Les gars du village ne sont pas très tolérants.
Je t'assure que ce gars m'exaspère.
Je t'assure que ce gamin m'exaspère.

A cet exercice, nous gagnons toujours beaucoup d'argent.

4. Trou

Le trou n'avait jamais été aussi grand.
Il n'avait jamais vu un aussi grand trou.
Le troupeau n'avait jamais été aussi grand.
Vous trouvez ce garçon bizarre.

5. Thé

Il aime le thé plus que tout.
Fais vite, le thé va refroidir.
Il adore le tennis féminin.
Demain, il téléphonera à sa grand mère.

6. Thé

Il fait la grimace, ce thé est vraiment affreux.
Tu sais pourtant bien qu'il faut éviter le thé le soir.
Le témoin va s'en aller s'il continue ainsi.
Je témoigne dans quelques instants.

7. Rat

Le soir, les rats sortent de leur repère.
Il avait si peur des rats.
Il faut en finir avec ces rapports.
Dès que possible, il rappelle la banque.

8. Sang

Ce sang n'est pas le sien assurément.
Il voit le sang couler sur sa main.
Il regarde le sentier à sa droite.
Sans savoir pourquoi, je sentais qu'il n'allait pas être d'accord.

9. Bout

Le bout du tunnel apparaît toujours quand on s'y attend le moins.
Il en a assez de ces bouts de ficelle.
Il est perdu sans ses bouquins toute la journée.
Il faut que nous bougions si nous voulons rester éveillés.

10. Chat

Le chat de mes voisins est très câlin.
Je regarde le chat qui saute sur le muret d'en face.
Le chapeau de paille de ma soeur la protège bien des insulations.
Il chatouille sa nièce à chaque fois qu'il la voit.

11. Dos

Le dos des adolescents est vraiment fragile.
En effet, son dos est magnifiquement sculpté.
Avec son dossier il n'a rien à craindre.
Vous donnez à cette association une somme modique.

12. Rang

Le rang dans l'armée est quelque chose d'important.
Il a fait tomber le rang des pioches.
Un rendement élevé est exigé.
Nous rentrons à la maison en autobus faute de voiture.

13. Jus

Ce jus de pomme est délicieux.
Elle préfère de loin ce jus-là.
Ce jugement est inéquitable.
Vous jugez tout le monde tout le temps.

14. Pouls

Le pouls est plus facile à trouver au niveau du cou.
Il est inquiet, son pouls n'est pas très bon.
Il affirme que le poulet est périmé.
Si vous vous décidez, nous pouvons tout régler aujourd'hui.

15. Part

La part de travailleur pauvre augmente d'année en année.
Il est content, la part qui lui revient est correcte.
Il renonce à la partie de pêche pour travailler.
S'il pouvait, il partirait loin, très loin.

16. Vie

La vie des animaux est passionnante à étudier.
J'ai un livre sur la vie des saints.
La visite médicale est un moment important.
Nous vivons depuis toujours à la campagne dans une grande maison.

Items Verbe

1. Dis

Je dis n'importe quoi quand je le veux.
En général, je dis ce que je pense.
Je dirigerai une équipe de football plus tard.
Le dimanche est réservé à la relaxation.

2. Dis

Il dit des choses sans importance.
Pour être hommêtes, nous disons tout sauf le contraire.
A cause de ce travail, je dînerai vers quatre heure.
Je n'aime pas ses distractions puérides.

3. Ment

Il ment comme un arracheur de dents.
Quand il fait cette grimace, il ment c'est infaillible.
Et bien, vous mangez de bon appétit.
Sous le menton, il avait une horrible cicatrice.

4. Ment

Je l'ai vu, il ment à sa femme et ses enfants.
Dis le tout de suite que je mens.
Il mange comme quatre dès qu'il est chez sa mère.
Dans cette mansarde, je jouais quand j'étais enfant.

5. Tais

Je tais mon innocence pour la protéger, je le reconnais.
Si tu veux, je me tais dès son arrivée.
C'est d'accord, on se téléphone.
Une fois la terrasse achevée, nous emménagerons.

6. Rend

Je rends les clés de l'appartement cet après-midi.
Ce matin, il rend la voiture au garagiste.
Demain, si tout se passe bien, je rencontre mon patron.
Lors de leur première rencontre, il la trouva exquise.

7. Va

Il va à la piscine tous les matins.
De temps à autre, je me demande s'il va bien.
Je ne pense pas qu'un valet ferait l'affaire.
Il valait mieux pour lui qu'il ait un alibi.

8. Sais

Je sais que j'ai raison et vous aussi d'ailleurs.
Arrête, je sais que tu ne m'en veux pas.
Attention, tu serraais beaucoup trop fort !
Le seigneur veut un nouveau cheval.

9. Vais

Je vais voir ma tante ce week-end.

Dans une minute, je vais au marché faire les courses.
Fais moi confiance, je verrai s'il est honnête.
Dans la cuisine, la vaisselle sale encombre l'évier.

10. Suis

Je suis un génie incompris, mais je m'en fiche.
Ayez confiance, je suis un guide plus que qualifié.
J'ai parfois des remords, je suivais le mouvement.
Il faut le dire, le suivant avait une voix atroce.

11. plais

Je plais à mes patrons sans savoir pourquoi.
Il est vrai que je plais beaucoup aux animaux.
Je plaisante de temps à autre sur mes défauts.
Cela suffit, la plaisanterie a assez duré.

12. Vois

Je vois que tu n'es pas dans ton assiette.
Je ne peux pas dire que je vois où tu veux en venir.
Honteuse, elle voyage sous un faux nom.
La voisine est superbe dans sa robe de mariée.

13. Vois

Je vois une procession dans la plaine au loin.
Personnellement, je vois juste un rond noir.
Une fois par an, je voyage hors de France.
Je m'imagine tout à fait dans cette voiture.

14. Sens

Il sent très vite ce genre de choses.
Vu les circonstances tout me semble fade.
Il semblait perdu dans ses pensées.
J'ai des doutes, le sentier ne mène nul part.

15. Vend

Le matin, je vends des petits pains au marché.
Il n'a plus le choix, il vend sa maison.
Je vengerai ma femme et mes enfants.
Le vendredi, je ne mange jamais de poisson.

16. Perd

Il perd sans cesse ses clefs de voiture.
Plus le temps passe, plus il perd de choses.
Si je le peux, je permettrai qu'il s'en aille.

J'aime cette personne sans me l'avouer.

A.3 Phrases utilisées dans l'expérience comportementale 3 : influence des mots grammaticaux dans le calcul des catégories syntaxique

Condition 1

Construction du type « Quand *le/la/les* X » pour les phrases contenant la cible-Nom et du type « Je *le/la/les* X » pour les phrases incluant la cible-Verbe. Le mot cible est en gras ci-dessous.

1. Phrase Nom

[Quand la tempête] [souffle fort] [la température] [semble glaciale]
[Quand la **baruche**] [daloube guire] [la marnicature] [timbe goutal]

Phrase Verbe

[Je la transporte] [dans son lit] [car elle s'est endormie]
[Je la **baruche**] [dans son tit] [car elle s'est amutie]

2. Phrase Nom

[Si le volet] [reste ouvert] [la température] [diminuera]
[Si le **manatre**] [eneste oulerte] [la capotaline] [churanera]

Phrase Verbe

[Ils le surprennent] [en train de manger] [dans la cuisine]
[Ils le **manatrent**] [en glon de mucher] [dans la badole]

3. Phrase Nom

[Comme la douceur] [est annoncée], [les feuilles] [vont repousser] [sur les arbres]
[Comme la **crépouille**] [est dacrinée], [les meuches] [gont doriper] [sur les atrules]

Phrase Verbe

[Tu la détends] [de son stress] [en la massant]
[Tu la **crépouilles**] [de son prasse] [en la mèrant]

4. Phrase Nom

[Chez le voisin] [est organisée une fête] [la musique] [est donc forte]
[Chez le **cramone**] [est apuriquée une goufe], [la piluche] [est donc nirte]

Phrase Verbe

[Je le sermonne] [à plusieurs reprises] [pour qu'il ne recommence pas]
[Je le **cramone**] [à pileurs titouses] [pour qu'il ne dabilise pas]

5. Phrase Nom

[Si la fillette] [joue aux échecs], [elle apprendra] [à se concentrer]

[Si la **jigoule**] [droue aux arecs], [elle mopalera] [à se guirouter]

Phrase Verbe

[Il la surveille] [du coin de l'oeil] [en travaillant]

[Il la **jigoule**] [du prand de l'oille] [en samirant]

6. Phrase Nom

[Sous la commode] [tu trouveras] [la serpillère]

[Sous la **ramode**] [tu louilleras] [la sadillère]

Phrase Verbe

[Je la secoue] [par les épaules] [jusqu'à ce qu'elle me dise] [la vérité]

[Je la **ramode**] [par les élaudes] [pousqu'à ce qu'elle me rise] [la doubité]

7. Phrase Nom

[Même le facteur] [vient nous dire bonjour] [tous les matins]

[Même le **falore**] [drine nous ripe jambour] [tous les galins]

Phrase Verbe

[Elle le retient] [dans sa chute vertigineuse]

[Elle le **falore**] [dans sa frute bertalineuse]

8. Phrase Nom

[Dans les racines] [apparaissent les jeunes pousses]

[Dans les **dabines**] [conirassent les jounes deusses]

Phrase Verbe

[Elles les recouvrent] [car le froid est mordant]

[Elles les **dabinent**] [car le prouille est dorlant]

9. Phrase Nom

[Sur les tables] [on trouve les saladiers] [remplis de fruits]

[Sur les **pliches**] [on drate les satafiers] [noublis de crouits]

Phrase Verbe

[On les trouve] [dans le patio]

[On les **pliche**] [dans le tarío]

10. Phrase Nom

[Par la vitre] [vois-tu mon père arriver] ?

[Par la **chitre**] [pouilles-tu mon gore analer] ?

Phrase Verbe

[Tu la jettes] [par la fenêtre] [immédiatement]

[Tu la **chitres**] [par la nabouche] [tobutalement]

11. Phrase Nom

[Sans les mains] [on n'arrivera jamais] [à écrire]

[Sans les **gounes**] [on n'abulera jamais] [à odire]

Phrase Verbe

[On les laisse] [à la prochaine station]
 [On les **goune**] [à la bolaine pichon]

12. Phrase Nom

[Même le fer] [ne résistera pas] [à notre nouvelle trouvaille]
 [Même le **lipe**] [ne vasoupera pas] [à notre mastelle bruchaille]

Phrase Verbe

[Elle le lance] [assez fort] [pour qu'il vole haut]
 [Elle le **lipe**] [napi dour] [pour qu'il daille pi]

13. Phrase Nom

[Quand le train] [arrive en gare] [le sifflet retentit]
 [Quand le **troune**] [cagouille en rule] [le nifflot retonguit]

Phrase Verbe

[Ils le traînent] [derrière eux] [tous les jours]
 [Ils le **troune**] [canière eux] [tous les gueilles]

14. Phrase Nom

[Sous les feuilles] [cueilles les framboises] [les plus mûres] [pour la confiture]
 [Sous les **frailles**] [vripes les maroises] [les plus dares] [pour la fabiture]

Phrase Verbe

[Il les frotte] [vigoureusement] [pour enlever la terre]
 [Il les **fraille**] [vitupérisement] [pour arouter la grèche]

15. Phrase Nom

[Comme les fraises] [sont bien mûres] [j'en mange beaucoup]
 [Comme les **crupes**] [sont bine deilles] [j'en gouche badou]

Phrase Verbe

[Elles les coupent] [sur la table] [en petits morceaux]
 [Elles les **crupent**] [sur la bloute] [en tipit corneux]

16. Phrase Nom

[Sans les mains] [réussis-tu] [à faire du vélo] ?
 [Sans les **dranes**] [sabouris-tu] [à grare du lémo] ?

Phrase Verbe

[Tu les prends] [sans rien dire]
 [Tu les **dranes**] [sans nire dare]

Condition 2

Construction du type « que *le/la/les* X » pour les phrases contenant la cible-Nom et du type « qui *le/la/les* X » pour les phrases incluant la cible-Verbe. Le mot cible est en gras ci-dessous.

1. Phrase Nom

[Elle voudrait] [la poupée] [que la fillette] [a acheté] [au marché]
 [Elle coutait] [la pépouille] [que la **baruche**] [a poudé] [au biché]

Phrase Verbe

[Je joue] [avec la jolie jument] [qui la regarde] [avec tendresse]
 [Je gouille] [avec la liji mulant] [qui la **baruche**] [arec mudesse]

2. Phrase Nom

[Il a peur] [du gros monstre] [que le joueur] [doit affronter]
 [Il a nir] [du pros noustre] [que le **manatre**] [da agronter]

Phrase Verbe

[Il a relu] [le dossier] [qui le concerne] [directement]
 [Il a guilé] [le ronier] [qui le **manatre**] [daléquiment]

3. Phrase Nom

[Tu ne prendras pas] [le pantalon] [que la vendeuse] [te conseille]
 [Tu ne douras pas] [le nabalon] [que la **crépouille**] [te maleille]

Phrase Verbe

[Sens-tu] [cette chaleur] [qui la réchauffe] [tout doucement] ?
 [Doules-tu] [cette bouneur] [qui la **crépouille**] [tiau douliment] ?

4. Phrase Nom

[Ils achèteront] [le grand cahier] [que le garçon] [veut pour la rentrée]
 [Ils alouilleront] [le dran dailler] [que le **cramone**] [jouille pour la fautrée]

Phrase Verbe

[J'ai lu] [le courrier] [qui le réjouit] [énormément]
 [J'ai groul] [le roubier] [qui le **cramone**] [jagoulament]

5. Phrase Nom

[Je possède] [le parapluie] [que la voisine] [prend pour sortir]
 [Je salède] [le galachoui] [que la **jigoule**] [truche pour guélire]

Phrase Verbe

[J'entends] [l'oiseau] [qui la réveille] [tous les matins]
 [J'onken] [l'éprou] [qui la **jigoule**] [tous les tachins]

6. Phrase Nom

[Il regarde] [la caméra] [que la cadreuse] [a en main]
 [Il crébarde] [la fémara] [que la **ramode**] [a en chouin]

Phrase Verbe

[J'ai enroulé] [la corde] [qui la retient] [à la taille]
 [J'ai enfaré] [la forne] [qui la **ramode**] [à la gruche]

7. Phrase Nom

[Ils regardent] [le trait lumineux] [que le concorde] [dessine dans le ciel]

[Ils agrament] [le duche goumaneux] [que le **falore**] [dérine dans le doune]

Phrase Verbe

[Je travaille] [ce morceau] [qui le séduit] [depuis toujours]

[Je grassine] [ce curo] [qui le **falore**] [dabi talou]

8. Phrase Nom

[Elle cueille] [les bambous] [que les pandas] [adorent manger]

[Elle deugue] [les fadous] [que les **dabines**] [damolent vanprer]

Phrase Verbe

[Ils suivent] [la grosse lampe] [qui les éclairent] [dans le noir]

[Ils doubent] [la sode lanche] [qui les **dabinent**] [dans le boine]

9. Phrase Nom

[On s'assoit] [sur l'herbe] [que les vaches] [ont broutée]

[On s'agroune] [sur l'ourbe] [que les **pliches**] [ont ariée]

Phrase Verbe

[Nous avons joué] [avec le petit chat] [qui les suit] [où qu'ils aillent]

[Nous avons gruché] [avec le tipi fra] [qui les **pliche**] [où qu'ils dabent]

10. Phrase Nom

[Nous prendrons] [les chatons] [que la vieille] [aimait temps]

[Nous mouprons] [les farons] [que la **chitre**] [agait trume]

Phrase Verbe

[Sent-elle] [la pluie] [qui la mouille] [abondamment] ?

[Leume-t-elle] [la gruié] [qui la **chitre**] [agoulament] ?

11. Phrase Nom

[Elle représente] [la princesse] [que les princes] [doivent sauver]

[Elle refudente] [la minesse] [que les **gounes**] [dachent mider]

Phrase Verbe

[Elle aime] [le boulanger] [qui les cuit] [dans son four]

[Elle bine] [le souvinier] [qui les **goune**] [dans son mour]

12. Phrase Nom

[Ils observent] [la gazelle] [que le lion] [essaie d'attraper]

[Ils sidervent] [la mafelle] [que le **lipe**] [ébai d'agrouner]

Phrase Verbe

[Ils cherchent] [L'arbre magique] [que le roi] [a découvert]

[Ils darchent] [l'afe ganide] [que le **lipe**] [a vaboulert]

13. Phrase Nom

[Il fait signe] [aux enfants] [que le train] [ramène chez eux]

[Il fait douque] [aux gounants] [que le **troune**] [salène chez zui]

Phrase Verbe

[Vous observez] [la gourmande] [qui le mange] [goulûment]
[Vous saguepez] [la fourdanche] [qui le **troune**] [saroument]

14. Phrase Nom

[Elles retiennent] [les images] [que les clips] [utilisent souvent]
[Elles repouillent] [les sidages] [que les **frailles**] [équilipent sament]
Phrase Verbe
[Ils remercient] [la jeune fille] [qui les aide] [gentiment]
[Ils sabermient] [la feune diche] [qui les **fraille**] [éboument]

15. Phrase Nom

[Il ne croit pas] [aux histoires] [que les films] [racontent]
[Il ne lour pas] [aux famoires] [que les **crupes**] [damoulement]
Phrase Verbe
[Ils chantent] [avec le monsieur] [qui les guide] [dans la forêt]
[Ils daillent] [avec le foimeu] [qui les **crupe**] [dans la pomine]

16. Phrase Nom

[On plaint] [les animaux] [que les zoo] [enferment] [dans des cages]
[On lome] [les olinaux] [que les **dranes**] [oncirment] [dans des goutes]
Phrase Verbe
[On suit] [les archéologues] [qui les fouillent] [depuis des mois]
[On chouvre] [les gasidélogues] [qui les **dranent**] [onpi des muches]

A.4 Stimuli utilisés dans l'expérience 3bis : influence de la prosodie dans le calcul de la catégorie syntaxique en Jabberwocky

1. Phrase Nom

[La fraise], [la framboise] [et les groseilles] [sont des fruits rouges]
[La draise], [la **baruche**] [et les nicrons] [sont des chouï drones]
Phrase Verbe
[Sa mère] [la retient] [en l'attrapant] [par la main]
[Sa lure] [la **baruche**] [en l'apramant] [par la blain]

2. Phrase Nom

[Le requin], [le dauphin] [et la baleine] [nagent dans l'océan]
[Le balu], [le **manatre**] [et la baride] [jalent dans l'odouillant]
Phrase Verbe
[Ma petite soeur] [le sirote] [avec une paille]
[Ma banite jeure] [le **manatre**] [avec une chouille]

3. Phrase Nom

[Le cheval], [la jument], [et leur poulain] [se baladent dans la prairie]
 [Le bardou], [la **crépouille**] [et leur grichin] [se mouladent dans la grailie]

Phrase Verbe

[L'inquiète maman] [la surveillance] [ouvertement]
 [L'inbiète malan] [la **crépouille**] [irachement]

4. Phrase Nom

[Le dresseur] [le jongleur] [et la magicienne] [se succèdent] [sans interruption]
 [Le cédreur] [le **cramone**] [et la mogrivienne] [se jumèdent] [sans onbéripou]

Phrase Verbe

[La méchante sorcière] [le transforme] [en gros crapaud]
 [La chémante bormière] [le **cramone**] [en blon cripou]

5. Phrase Nom

[Le téléphone] [la télé] [et l'ordinateur] [font partie] [de la maison]
 [Le boguidone] [la **jigoule**] [et l'agibaleur] [fonent larmi] [de la dason]

Phrase Verbe

[Cette jupe longue] [la grossit] [énormément]
 [Cette poume gonre] [la **jigoule**] [olounément]

6. Phrase Nom

[La digue] [la chemise] [et le nourrisson] [sont des mots]
 [La plougue] [la **ramode**] [et le peminon] [sont des mouc]

Phrase Verbe

[La nounou] [la nourrit] [et la change] [tous les soirs]
 [La bouchou] [la **ramode**] [et la choune] [tous les joibes]

7. Phrase Nom

[La maison] [le jardin] [et les dépendances] [appartiennent] [à mon voisin]
 [La saimu] [le **falore**] [et les pélamances] [olarvinrent] [à mon soimon]

Phrase Verbe

[Les saisons] [le surprennent] [beaucoup]
 [Les méchas] [le **falorent**] [alou]

8. Phrase Nom

[Les volets] [les rideaux] [et les voilages] [les protègent] [du froid]
 [Les lurets] [les **dabines**] [et les voidouges] [les goprègent] [du frou]

Phrase Verbe

[Le professeur] [les accepte] [aujourd'hui] [toutes les deux]
 [Le mabiteur] [les **dabine**] [abouti] [toutes les reux]

9. Phrase Nom

[Les renards] [les loups] [et les belettes] [connaissent la chanson]

[Les medars] [les **pliches**] [et les telèbes] [nocancent la panchon]

Phrase Verbe

[Une gourmande] [les mangent] [avidement]

[Une mourtan] [les **plichent**] [adivement]

10. Phrase Nom

[Le rouge] [le bleu] [et le jaune] [sont des couleurs primaires]

[Le froume] [la **chitre**] [et le naude] [sont des badeurs milaires]

Phrase Verbe

[Le joli chaton] [la suit] [sans hésiter]

[Le lobi gatru] [la **chitre**] [sans loubiter]

11. Phrase Nom

[Les garçons] [les filles] [et les instituteurs] [regardent] [par la fenêtre]

[Les rabous] [les **gounes**] [et les jubicateurs] [guédardent] [par la fachètre]

Phrase Verbe

[Les moustiques] [les piquent] [sauvagement] [toute la nuit]

[Les kismotes] [les **gounent**] [oubagement] [toute la ruie]

12. Phrase Nom

[La petite soeur] [le frère] [et le cousin] [s'amusent ensemble]

[La fichoute leille] [le **lipe**] [et le foulin] [s'arulent ennable]

Phrase Verbe

[La jolie fille] [le prend] [avec tendresse]

[La fili jouille] [le **lipe**] [avec craillesse]

13. Phrase Nom

[Le cabas] [le sac] [et le panier] [permettent] [le transport d'objets]

[Le jabas] [le **troune**] [et le naguier] [moulettent] [le tandor d'oulets]

Phrase Verbe

[Ses animaux] [le suivent] [où qu'il aille]

[Ses enipas] [le **trounent**] [où qu'il rine]

14. Phrase Nom

[L'encre] [les plumes] [et le papier] [servaient à écrire] [autrefois]

[L'ambe] [les **frailles**] [et le gapier] [sendaient à goulir] [boutreloi]

Phrase Verbe

[Un gaspilleur] [les met] [à la poubelle]

[Un groubilleur] [les **fraille**] [à la joubette]

15. Phrase Nom

[Les pianos] [les harpes] [et les violons] [sont beaucoup joués]

[Les niados] [les **crupes**] [et les joilons] [sont oulou gunés]

Phrase Verbe

[Mon petit fils] [les prend] [dans ses bras] [doucelement]
[Mon pagui disse] [les **crupe**] [dans ses fras] [daloument]

16. Phrase Nom

[Les pages] [les feuilles][et les livres] [valsent] [dans tous les sens]
[Les tumes] [les **dranes**] [et les mouvres] [dirnent] [dans tous les mense]

Phrase Verbe

[L'examen][les stresse] [à l'avance]
[L'étadin] [les **drane**] [à l'abonce]

Stimuli langagier pour l'expérience Muslang

1. Ils sortent le chien au bord du fleuve qui passe à côté du parc du manoir.
2. Paul mange la tarte sur laquelle se trouvait une bougie en cire pour fêter son succès.
3. Elle couvre la marmite qui réchauffe tout doucement la soupe de légumes du jardin de mamie.
4. Il passe la balle de foot à Luc qui tire très fort à côté du but.
5. Ils manquent de force pour suivre le guide qui grimpe sans peine des pentes très raides.
6. Je porte des croissants aux femmes très gourmandes qui finissent leurs pains aux noix du Brésil.
7. Je frappe la cloche du portail en fer du château qui date du siècle des lumières.
8. Elle coupe avec soin les fleurs qui poussent au bord des champs de blé bien jaunes.
9. La neige est tombée en couche très fine au nord du village où habite ma cousine.
10. La fillette a lu les lettres très longues du russe qui écrit un français presque parfait.
11. Il savoure avec délice le thé de Chine qui sent les fleurs de Jasmin toutes fraîches.
12. Elle cueille mille roses pour faire un don aux prêtres du temple du dieu du soleil.
13. Le vampire est fou du sang des monstres qui rodent en groupe les nuits sans lune.
14. Mon père parle souvent de choses très savantes qui fascinent les copains qui viennent chez moi.
15. La biche sort vite du bois très sombre pour fuir les chasseurs qui tirent sur elle.
16. Le cirque plaît beaucoup aux garçons qui préfèrent les grimaces des clowns aux dresseurs de tigres.

Un tout petit peu de théorie de détection du signal.

Pour comprendre ce que signifie cet indice de sensibilité, appliquons la théorie de détection du signal à notre cas.

Nous considérons ici que la tâche était de détecter les différences pour plus de clarté dans l'explication. 4 types de réponse sont possibles (tableau C.1).

		Cible	
		Identiques	Différentes
Réponse Donnée	Identiques	Détection Correcte (Hit)	Erreur - Détection manquée (Miss)
	Différentes	Erreur- Fausse Alarme	Rejet correct (correct rejection)

TABLE C.1 – Différents types de réponses possible

Dans 2 cas, la réponse est juste : détection correcte de la différence et rejet correct lorsque les 2 stimuli sont identiques, c'est la partie qui nous intéresse. Les 2 autres cas illustrent le bruit dans les données c'est-à-dire lorsque les sujets ont détecté une différence inexistante (fausse alarme) ou lorsqu'ils en ont manqué une. Le calcul de sensibilité d'une mesure reflète la séparation entre la moyenne des distributions du signal d'une part et le bruit d'autre part et ce en unités de déviation standard de la distribution du bruit. Plus le d' est grand, plus le signal est facile à dissocier du bruit, la réponse est plus nette et donc moins due au hasard dans notre cas. Plus le d' est proche de 1, plus il est difficile de séparer le signal du bruit, puisque les 2 mesures ont le même ordre de grandeur.

Bibliographie

- Abdul-Kareem, I. A., Stancak, A., Parkes, L. M., and Sluming, V. (2011). Increased gray matter volume of left pars opercularis in male orchestral musicians correlate positively with years of musical performance. *J Magn Reson Imaging*, 33(1) :24–32. (Cité en page 187.)
- Abney, S. P. and Johnson, M. (1991). Memory requirements and local ambiguities of parsing strategies. *Journal of Psycholinguistic Research*, 20(3) :233–250. (Cité en pages 72 et 76.)
- Abrams, D. A., Bhatara, A., Ryali, S., Balaban, E., Levitin, D. J., and Menon, V. (2011). Decoding temporal structure in music and speech relies on shared brain resources but elicits different fine-scale spatial patterns. *Cerebral Cortex*, 21(7) :1507–1518. (Cité en pages 108, 152, 199, 200, 204, 213 et 215.)
- Ackermann, H. and Hertrich, I. (2000). The contribution of the cerebellum to speech processing. *Journal of Neurolinguistics*, 13(2-3) :95 – 116. (Cité en page 96.)
- Amunts, K., Schleicher, A., Bürgel, U., Mohlberg, H., Uylings, H. B., and Zilles, K. (1999). Broca’s region revisited : cytoarchitecture and intersubject variability. *Journal of Comparative Neurology*, 412(2) :319–341. (Cité en pages 172, 195 et 196.)
- Amunts, K., Weiss, P. H., Mohlberg, H., Pieperhoff, P., Eickhoff, S., Gurd, J. M., Marshall, J. C., Shah, N. J., Fink, G. R., and Zilles, K. (2004). Analysis of neural mechanisms underlying verbal fluency in cytoarchitectonically defined stereotaxic space—the roles of brodmann areas 44 and 45. *Neuroimage*, 22(1) :42–56. (Cité en pages 158, 171 et 172.)
- Anwander, A., Tittgemeyer, M., von Cramon, D. Y., Friederici, A. D., and Knösche, T. R. (2007). Connectivity-based parcellation of broca’s area. *Cereb Cortex*, 17(4) :816–825. (Cité en page 203.)
- Ashby, J. and Clifton, C. (2005). The prosodic property of lexical stress affects eye movements during silent reading. *Cognition*, 96(3) :B89–100. (Cité en page 70.)
- Baayen, R. H. (2008). *Analyzing linguistic data : a practical introduction to statistics using R*. Cambridge university press. (Cité en page 87.)
- Bader, M. and Meng, M. (1999). Subject-object ambiguities in german embedded clauses : An across-the-board comparison. *Journal of Psycholinguistic Research*, 28(2) :121–143. (Cité en page 72.)
- Bahlmann, J., Schubotz, R. I., and Friederici, A. D. (2008). Hierarchical artificial grammar processing engages broca’s area. *Neuroimage*, 42(2) :525–534. (Cité en page 78.)
- Baldwin, D. A., Markman, E. M., Bill, B., Desjardins, R. N., Irwin, J. M., and Tidball, G. (1996). Infants’ reliance on a social criterion for establishing word-object relations. *Child Dev*, 67(6) :3135–3153. (Cité en page 11.)
- Beach, C. M. (1991). The interpretation of prosodic patterns at points of syntactic structure ambiguity : Evidence for cue trading relations. *Journal of Memory and Language*, 30(6) :644 – 663. (Cité en page 61.)

- Bergelson, E. and Swingle, D. (2012). At 6-9 months, human infants know the meanings of many common nouns. *Proc Natl Acad Sci U S A*, 109(9) :3253–3258. (Cité en pages 8, 20 et 212.)
- Bermudez, P., Lerch, J. P., Evans, A. C., and Zatorre, R. J. (2009). Neuroanatomical correlates of musicianship as revealed by cortical thickness and voxel-based morphometry. *Cereb Cortex*, 19(7) :1583–1596. (Cité en pages 187 et 188.)
- Bernal, S., Dehaene-Lambertz, G., Millotte, S., and Christophe, A. (2010). Two-year-olds compute syntactic structure on-line. *Dev Sci*, 13(1) :69–76. (Cité en page 20.)
- Bernal, S., Lidz, J., Millotte, S., and Christophe, A. (2007). Syntax constrains the acquisition of verb meaning. *Language Learning and Development*, 3(4) :325–341. (Cité en pages 18, 21, 22, 36 et 211.)
- Besson, M. and Faïta, F. (1995). An event-related potential (erp) study of musical expectancy : Comparison of musicians with nonmusicians. *Journal of Experimental Psychology : Human Perception and Performance*, 21(6) :1278–1296. (Cité en page 107.)
- Blasko, D. G. and Hall, M. D. (1998). Influence of prosodic boundaries on comprehension of spoken english sentences. *Percept Mot Skills*, 87(1) :3–18. (Cité en page 61.)
- Bock, Kathryn & Levelt, W. (1994). *Handbook of Psycholinguistics*, chapter Language Production : Grammatical Encoding, pages 945–983. Academic Press. (Cité en page 209.)
- Boersma, P. and Weenink, D. (2010). Praat : doing phonetics by computer (version 5.1.25). Computer program from <http://www.praat.org/>. (Cité en pages 55 et 124.)
- Brauer, J. and Friederici, A. D. (2007). Functional neural networks of semantic and syntactic processes in the developing brain. *Journal of Cognitive Neuroscience*, 19(10) :1609–1623. PMID : 18271736. (Cité en page 172.)
- Braze, D., Mencl, W. E., Tabor, W., Pugh, K. R., Constable, R. T., Fulbright, R. K., Magnuson, J. S., Dyke, J. A. V., and Shankweiler, D. P. (2011). Unification of sentence processing via ear and eye : an fmri study. *Cortex*, 47(4) :416–431. (Cité en page 191.)
- Brennan, J., Nir, Y., Hasson, U., Malach, R., Heeger, D. J., and Pykkänen, L. (2012). Syntactic structure building in the anterior temporal lobe during natural story listening. *Brain Lang*, 120 :163–173. (Cité en page 196.)
- Brett, M., Anton, J. L., Valabregue, R., and Poline, J. B. (2002). Region of interest analysis using an spm toolbox. *NeuroImage*, 16(2) :497. (Cité en page 158.)
- Brown, S., Martinez, M. J., and Parsons, L. M. (2006). Music and language side by side in the brain : a pet study of the generation of melodies and sentences. *Eur J Neurosci*, 23(10) :2791–2803. (Cité en page 108.)
- Brusini, P. (2012). *Découvrir les noms et les verbes : Quand les classes sémantiques initialisent les catégories syntaxiques*. PhD thesis, École Doctorale n°158 - Cerveau-Cognition-Comportement. (Cité en page 211.)
- Cannam, C., McIntyre, D. M., Bown, R., and Laurent, G. Rosegarden (version 10.02). from <http://www.rosegardenmusic.com/>. (Cité en page 123.)
- Caplan, D. (2009). Experimental design and interpretation of functional neuroimaging studies of cognitive processes. *Hum Brain Mapp*, 30(1) :59–77. (Cité en page 75.)

- Caplan, D., Chen, E., and Waters, G. (2008). Task-dependent and task-independent neurovascular responses to syntactic processing. *Cortex*, 44(3) :257–275. (Cité en page 192.)
- Caplan, D., Vijayan, S., Kuperberg, G., West, C., Waters, G., Greve, D., and Dale, A. M. (2002). Vascular responses to syntactic processing : event-related fmri study of relative clauses. *Hum Brain Mapp*, 15(1) :26–38. (Cité en pages 84 et 97.)
- Caplan, D., Waters, G., and Alpert, N. (2003). Effects of age and speed of processing on rcbf correlates of syntactic processing in sentence comprehension. *Hum Brain Mapp*, 19(2) :112–131. (Cité en page 95.)
- Caplan, D., Waters, G., Kennedy, D., Alpert, N., Makris, N., Dede, G., Michaud, J., and Reddy, A. (2007). A study of syntactic processing in aphasia ii : neurological aspects. *Brain Lang*, 101(2) :151–177. (Cité en page 95.)
- Carandini, M., Movshon, J. A., and Ferster, D. (1998). Pattern adaptation and cross-orientation interactions in the primary visual cortex. *Neuropharmacology*, 37(4-5) :501–511. (Cité en page 76.)
- Carlson, K., Jr., C. C., and Frazier, L. (2001). Prosodic boundaries in adjunct attachment. *Journal of Memory and Language*, 45(1) :58 – 81. (Cité en page 61.)
- Carroll, M. K., Cecchi, G. A., Rish, I., Garg, R., and Rao, A. R. (2009). Prediction and interpretation of distributed neural activity with sparse models. *Neuroimage*, 44(1) :112–122. (Cité en page 92.)
- Catani, M., Jones, D. K., and ffytche, D. H. (2005). Perisylvian language networks of the human brain. *Ann Neurol*, 57(1) :8–16. (Cité en page 203.)
- Caulo, M., Briganti, C., Mattei, P. A., Perfetti, B., Ferretti, A., Romani, G. L., Tartaro, A., and Colosimo, C. (2007). New morphologic variants of the hand motor cortex as seen with mr imaging in a large study population. *AJNR Am J Neuroradiol*, 28(8) :1480–1485. (Cité en page 163.)
- Chemla, E., Mintz, T. H., Bernal, S., and Christophe, A. (2009). Categorizing words using 'frequent frames' : what cross-linguistic analyses reveal about distributional acquisition strategies. *Dev Sci*, 12(3) :396–406. (Cité en page 43.)
- Chomsky, N. (1957). *syntactic structures*. Mouton, The Hague. (Cité en page 101.)
- Christophe, A., Guasti, M. T. and Nespors, M., Dupoux, E., and van Ooyen, B. (1997). Reflections on phonological bootstrapping : its role for lexical and syntactic acquisition. *Language and Cognitive Processes*, 12 :585–612. (Cité en pages 19 et 211.)
- Christophe, A., Millotte, S., Bernal, S., and Lidz, J. (2008). Bootstrapping lexical and syntactic acquisition. *Language and Speech*, 51(Pt 1-2) :61–75. (Cité en pages v, 19, 20, 21 et 211.)
- Clifton, C., Carlson, K., and Frazier, L. (2002). Informative prosodic boundaries. *Lang Speech*, 45(Pt 2) :87–114. (Cité en page 61.)
- Constable, R. T., Pugh, K. R., Berroya, E., Mencl, W. E., Westerveld, M., Ni, W., and Shankweiler, D. (2004). Sentence complexity and input modality effects in sentence comprehension : an fmri study. *Neuroimage*, 22(1) :11–21. (Cité en page 191.)

- Cooper, R. P., Abraham, J., Berman, S., and Staska, M. (1997). The development of infants' preference for motherese. *Infant Behavior and Development*, 20(4) :477 – 488. (Cité en page 14.)
- Cox, D. D. and Savoy, R. L. (2003). Functional magnetic resonance imaging (fmri) "brain reading" : detecting and classifying distributed patterns of fmri activity in human visual cortex. *Neuroimage*, 19(2 Pt 1) :261–270. (Cité en page 76.)
- Dahan, D., Swingle, D., Tanenhaus, M. K., and Magnuson, J. S. (2000). Linguistic gender and spoken-word recognition in french. *Journal of Memory and Language*, 42 :465–480(16). (Cité en pages v, 44, 45, 50 et 209.)
- Dehaene-Lambertz, G. and Dehaene, S. (1994). Speed and cerebral correlates of syllable discrimination in infants. *Nature*, 370(6487) :292–295. (Cité en page 76.)
- Dehaene-Lambertz, G., Dehaene, S., Anton, J.-L., Campagne, A., Ciuciu, P., Dehaene, G. P., Denghien, I., Jobert, A., Lebihan, D., Sigman, M., Pallier, C., and Poline, J.-B. (2006). Functional segregation of cortical language areas by sentence repetition. *Hum Brain Mapp*, 27(5) :360–371. (Cité en page 76.)
- Demuth, K. and Tremblay, A. (2008). Prosodically-conditioned variability in children's production of french determiners. *Journal Of Child Language*, 41 :133–144. (Cité en pages xi, 38, 39 et 52.)
- Devauchelle, A.-D. (2008). *Exploration du réseau cérébral impliqué dans les traitements syntaxiques et lexico-sémantiques des phrases*. Neurosciences cognitives, Université Pierre et Marie Curie - Ecole Doctorale 158 Cerveau Cognition Comportement. (Cité en pages vii, viii, 114 et 190.)
- Devauchelle, A.-D., Oppenheim, C., Rizzi, L., Dehaene, S., and Pallier, C. (2009). Sentence syntax and content in the human temporal lobe : an fmri adaptation study in auditory and visual modalities. *J Cogn Neurosci*, 21(5) :1000–1012. (Cité en page 76.)
- Dirk, W., Thomas, E., Didier, G., and Benjamin, K. (2009). A cerebral network model of speech prosody comprehension. *International Journal of Speech-Language Pathology*, 11(4) :277–281. (Cité en page 74.)
- Doherty, C. P., West, W. C., Dilley, L. C., Shattuck-Hufnagel, S., and Caplan, D. (2004). Question/statement judgments : an fmri study of intonation processing. *Hum Brain Mapp*, 23(2) :85–98. (Cité en page 74.)
- Fan, R.-E., Chang, K.-W., Hsieh, C.-J., Wang, X.-R., and Lin, C.-J. (2008). Liblinear : A library for large linear classification. *Journal of Machine Learning Research*, 9 :1871–1874. (Cité en page 86.)
- Fedorenko, E., Hsieh, P.-J., Nieto-Castañón, A., Whitfield-Gabrieli, S., and Kanwisher, N. (2010). New method for fmri investigations of language : defining rois functionally in individual subjects. *J Neurophysiol*, 104(2) :1177–1194. (Cité en pages 174, 176, 178 et 181.)
- Fedorenko, E., Patel, A., Casasanto, D., Winawer, J., and Gibson, E. (2009). Structural integration in language and music : evidence for a shared system. *Mem Cognit*, 37(1) :1–9. (Cité en pages vii, 110, 111 et 215.)
- Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., and Pethick, S. J. (1994). Variability in early communicative development. *Monogr Soc Res Child Dev*, 59(5) :1–173; discussion 174–85. (Cité en page 22.)

- Fernald, A. (1985). Four-month-old infants prefer to listen to motherese. *Infant Behavior and Development*, 8(2) :181 – 195. (Cité en page 14.)
- Fisher, C. (1996). Structural limits on verb mapping : the role of analogy in children's interpretations of sentences. *Cognitive Psychology*, 31(1) :41–81. (Cité en page 20.)
- Fisher, C., Gertner, Y., Scott, R. M., and Yuan, S. (2010). Syntactic bootstrapping. *Wiley Interdisciplinary Reviews : Cognitive Science*, 1(2) :143–149. (Cité en page 20.)
- Fisher, C., Klingler, S. L., and Song, H.-J. (2006). What does syntax say about space ? 2-year-olds use sentence structure to learn new prepositions. *Cognition*, 101(1) :B19–B29. (Cité en page 21.)
- Flandin, G., Kherif, F., Pennec, X., Malandain, G., Ayache, N., and Poline, J.-B. (2002). Improved detection sensitivity in functional mri data using a brain parcelling technique. In *MICCAI (1)*, pages 467–474. (Cité en page 86.)
- Fodor, J. (1998). Learning to parse. *Journal of Psycholinguistic Research*, 27(2) :285–319. (Cité en page 69.)
- Foster, N. E. V. and Zatorre, R. J. (2010a). Cortical structure predicts success in performing musical transformation judgments. *Neuroimage*, 53(1) :26–36. (Cité en page 188.)
- Foster, N. E. V. and Zatorre, R. J. (2010b). A role for the intraparietal sulcus in transforming musical pitch information. *Cereb. Cortex*, 20(6) :1350–1359. (Cité en page 201.)
- Frazier, L., Carlson, K., and Clifton, C. (2006). Prosodic phrasing is central to language comprehension. *Trends Cogn Sci*, 10(6) :244–249. (Cité en page 61.)
- Friederici, A. D. (2009). Pathways to language : fiber tracts in the human brain. *Trends Cogn Sci*, 13(4) :175–181. (Cité en pages ix et 203.)
- Friederici, A. D., Kotz, S. A., Scott, S. K., and Obleser, J. (2010). Disentangling syntax and intelligibility in auditory language comprehension. *Hum Brain Mapp*, 31(3) :448–457. (Cité en page 172.)
- Friederici, A. D., Makuuchi, M., and Bahlmann, J. (2009). The role of the posterior superior temporal cortex in sentence comprehension. *Neuroreport*, 20(6) :563–568. (Cité en pages 194 et 195.)
- Friston, K., Ashburner, J., Kiebel, S., Nichols, T., and Penny, W. (2007). *Statistical Parametric Mapping : The Analysis of Functional Brain Images*. Elsevier Ltd. (Cité en page 83.)
- Gaser, C. and Schlaug, G. (2003). Brain structures differ between musicians and non-musicians. *The Journal of Neuroscience*, 23(27) :9240–9245. (Cité en page 188.)
- Gelman, A. and Hill, J. (2007). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press. (Cité en pages 87 et 88.)
- Gerken, L., Jusczyk, P. W., and Mandel, D. R. (1994). When prosody fails to cue syntactic structure : 9-month-olds' sensitivity to phonological versus syntactic phrases. *Cognition*, 51(3) :237 – 265. (Cité en page 16.)
- Gerken, L. and McIntosh, B. J. (1993). Interplay of function morphemes and prosody in early language. *Developmental Psychology*, 29(3) :448 – 457. (Cité en page 18.)

- Gillette, J., Gleitman, H., Gleitman, L., and Lederer, A. (1999). Human simulations of vocabulary learning. *Cognition*, 73(2) :135–176. (Cité en pages [v](#), [12](#) et [13](#).)
- Göksun, T., Küntay, A. C., and Naigles, L. R. (2008). Turkish children use morphosyntactic bootstrapping in interpreting verb meaning. *Journal of Child Language*, 35(2) :291–323. (Cité en page [20](#).)
- Gleitman, L. (1990). The structural sources of verb meanings. *Language Acquisition : A Journal of Developmental Linguistics*, 1 :3–55. (Cité en pages [10](#) et [11](#).)
- Golinkoff, R. M., Hirsh-Pasek, K., and & Schweisguth, M. A. (2001). *Approaches to bootstrapping : Phonological, syntactic and neurophysiological aspects of early language acquisition.*, chapter A reappraisal of young children’s knowledge of grammatical morphemes, pages 167–188. Amsterdam : Benjamins. (Cité en page [42](#).)
- Gout, A., Christophe, A., and Morgan, J. (2004). Phonological phrase boundaries constrain lexical access : Ii. infant data. *Journal of Memory and Language*, 51 :548–567. (Cité en pages [16](#), [25](#) et [27](#).)
- Gregory, E., Varley, R., and Herbert, R. (2012). Determiner primes as facilitators of lexical retrieval in english. *J Psycholinguist Res*, in press :-. (Cité en page [50](#).)
- Grice, H. P. (1975). *Syntax and semantics 3 : Speech acts*, chapter Logic and Conversation, pages 41–58. Academic Pr. (Cité en page [71](#).)
- Grill-Spector, K., Henson, R., and Martin, A. (2006). Repetition and the brain : neural models of stimulus-specific effects. *Trends Cogn Sci*, 10(1) :14–23. (Cité en page [76](#).)
- Hale, J. (2001). A probabilistic earley parser as a psycholinguistic model. In *NAACL ’01 Proceedings of the second meeting of the North American Chapter of the Association for Computational Linguistics on Language technologies*. (Cité en page [95](#).)
- Hall, D. G., Waxman, S. R., Brédart, S., and Nicolay, A.-C. (2003). Preschoolers’ use of form class cues to learn descriptive proper names. *Child Dev*, 74(5) :1547–1560. (Cité en page [21](#).)
- Hastie, T., Tibshirani, R., and Friedman, J. (2003). *The elements of Statistical Learning*. Springer. (Cité en page [86](#).)
- Hauser, M. D., Chomsky, N., and Fitch, W. T. (2002). The faculty of language : what is it, who has it, and how did it evolve? *Science*, 298(5598) :1569–1579. (Cité en page [217](#).)
- Haynes, J.-D. and Rees, G. (2005). Predicting the orientation of invisible stimuli from activity in human primary visual cortex. *Nat Neurosci*, 8(5) :686–691. (Cité en page [76](#).)
- Haynes, J.-D. and Rees, G. (2006). Decoding mental states from brain activity in humans. *Nat Rev Neurosci*, 7(7) :523–534. (Cité en page [85](#).)
- Hesling, I., Clément, S., Bordessoules, M., and Allard, M. (2005). Cerebral mechanisms of prosodic integration : evidence from connected speech. *NeuroImage*, 24(4) :937 – 947. (Cité en page [96](#).)
- Höhle, B., Schmitz, M., Santelmann, L. M., and Weissenborn, J. (2006). The recognition of discontinuous verbal dependencies by german 19-month-olds : Evidence for lexical and structural influences on children’s early processing capacities. *Language Learning and Development*, 2(4) :277–300. (Cité en page [43](#).)

- Höhle, B. and Weissenborn, J. (2003). German-learning infants' ability to detect unstressed closed-class elements in continuous speech. *Developmental Science*, 6(2) :122 – 127. (Cité en page 17.)
- Höhle, B., Weissenborn, J., Kiefer, D., Schulz, A., and Schmitz, M. (2004). Functional elements in infants' speech processing : The role of determiners in the syntactic categorization of lexical elements. *Infancy*, 5(3) :341–353. (Cité en pages 21, 22, 23 et 37.)
- Hochmann, J.-R., Endress, A. D., and Mehler, J. (2010). Word frequency as a cue for identifying function words in infancy. *Cognition*, 115(3) :444–457. (Cité en page 18.)
- Humphries, C., Binder, J. R., Medler, D. A., and Liebenthal, E. (2006). Syntactic and semantic modulation of neural activity during auditory sentence comprehension. *Journal of Cognitive Neuroscience*, 18(4) :665–679. A finir de lire. (Cité en pages 70, 191 et 213.)
- Humphries, C., Love, T., Swinney, D., and Hickok, G. (2005). Response of anterior temporal cortex to syntactic and prosodic manipulations during sentence processing. *Hum Brain Mapp*, 26(2) :128–138. (Cité en pages 70, 74, 191 et 213.)
- Ischebeck, A. K., Friederici, A. D., and Alter, K. (2008). Processing prosodic boundaries in natural and hummed speech : an fmri study. *Cereb Cortex*, 18(3) :541–552. (Cité en pages 74, 75, 96 et 97.)
- Jusczyk, P. W., Hirsh-Pasek, K., Nelson, D. G., Kennedy, L. J., Woodward, A., and Piwoz, J. (1992). Perception of acoustic correlates of major phrasal units by young infants. *Cogn Psychol*, 24(2) :252–293. (Cité en page 16.)
- Kamitani, Y. and Tong, F. (2005). Decoding the visual and subjective contents of the human brain. *Nat Neurosci*, 8(5) :679–685. (Cité en pages 76 et 85.)
- Katz, J. and Pesetsky, D. (submitted). The identity thesis for language and music. *Language*, . :. (Cité en page 106.)
- Kayne, R. S. (1994). *the Antisymmetry of Syntax*. MIT press. (Cité en page 72.)
- Kedar, Y., Casasola, M., and Lust, B. (2006). Getting there faster : 18- and 24-month-old infants' use of function words to determine reference. *Child Development*, 77(2) :325–338. (Cité en pages 18, 22, 23 et 36.)
- Koelsch, S. (2011). Toward a neural basis of music perception - a review and updated model. *Front Psychol*, 2 :110. (Cité en page 186.)
- Koelsch, S., Gunter, T. C., v Cramon, D. Y., Zysset, S., Lohmann, G., and Friederici, A. D. (2002). Bach speaks : a cortical "language-network" serves the processing of music. *Neuroimage*, 17(2) :956–966. (Cité en page 108.)
- Koelsch, S., Gunter, T. C., Wittfoth, M., and Sammler, D. (2005). Interaction between syntax processing in language and in music : an erp study. *J Cogn Neurosci*, 17(10) :1565–1577. (Cité en pages vii, 109 et 111.)
- Krekelberg, B., Boynton, G. M., and van Wezel, R. J. A. (2006). Adaptation : from single cells to bold signals. *Trends Neurosci*, 29(5) :250–256. (Cité en page 76.)
- Krumhansl, C. L. and Kessler, E. J. (1982). Tracing the dynamic changes in perceived tonal organization in a spatial representation of musical keys. *Psychological Review*, 89(4) :334–368. (Cité en page 121.)

- Landau, B. and Gleitman, L. (1985). *Language and Experience - Evidence from a Blind Child* :. Harvard University Press, Cambridge Mass. (Cité en pages 10 et 12.)
- Lau, E. F., Phillips, C., and Poeppel, D. (2008). A cortical network for semantics : (de)constructing the n400. *Nat Rev Neurosci*, 9(12) :920–933. (Cité en pages 76 et 97.)
- Lerdahl, E. and Jackendoff, R. (1983). *A Generative Theory of Tonal Music*. Cambridge, MA : MIT Press. (Cité en pages 5, 102, 103, 104, 105 et 106.)
- Levitin, D. J. and Menon, V. (2003). Musical structure is processed in "language" areas of the brain : a possible role for brodmann area 47 in temporal coherence. *Neuroimage*, 20(4) :2142–2152. (Cité en pages viii, 108, 152, 153, 213 et 216.)
- Lew-Williams, C. and Fernald, A. (2010). Real-time processing of gender-marked articles by native and non-native spanish speakers. *J Mem Lang*, 63(4) :447–464. (Cité en pages 44, 50 et 209.)
- Luo, H. and Poeppel, D. (2007). Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. *Neuron*, 54(6) :1001–1010. (Cité en page 198.)
- Maess, B., Koelsch, S., Gunter, T. C., and Friederici, A. D. (2001). Musical syntax is processed in broca's area : an meg study. *Nat Neurosci*, 4(5) :540–545. (Cité en page 108.)
- Magnuson, J. S., Tanenhaus, M. K., and Aslin, R. N. (2008). Immediate effects of form-class constraints on spoken word recognition. *Cognition*, 108(3) :866–873. (Cité en page 50.)
- Makuuchi, M., Bahlmann, J., Anwender, A., and Friederici, A. D. (2009). Segregating the core computational faculty of human language from working memory. *PNAS*, 106(20) :8362–8367. hierarchy language. (Cité en pages viii, 158, 194 et 195.)
- Marcus, M. and Hindle, D. (1990). Description theory and intonation boundaries. In Altmann, G. T. M., editor, *Cognitive models of speech processing*, chapter Description theory and intonation boundaries, pages 483–512. MIT Press, Cambridge, MA, USA. (Cité en page 61.)
- Marslen-Wilson, W. D. (1975). Sentence perception as an interactive parallel process. *Science*, 189(4198) :226–228. (Cité en page 44.)
- Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word-recognition. *Cognition*, 25(1-2) :71–102. (Cité en page 208.)
- Mazoyer, B. M., Tzourio, N., Frak, V., Syrota, A., Murayama, N., Levrier, O., Salamon, G., Dehaene, S., Cohen, L., and Mehler, J. (1993). The cortical representation of speech. *Journal of Cognitive Neuroscience*, 5(4) :467–479. (Cité en page 158.)
- Michel, V., Eger, E., Keribin, C., Poline, J., and Thirion, B. (2010). A supervised clustering approach for extracting predictive information from brain activation images. In *MMBIA'10*. (Cité en page 86.)
- Millotte, S. (2005). *Le rôle de la prosodie dans le traitement syntaxique adulte et l'acquisition de la syntaxe*. PhD thesis, Ecole des Hautes Etudes en Sciences Sociales. (Cité en page 51.)
- Millotte, S., Morgan, J., Margules, S., Bernal, S., Dutat, M., and Christophe, A. (2010). Phrasal prosody constrains word segmentation in french 16-month-olds. *Journal of Portuguese Linguistics*, 9 :67–86. (Cité en pages 16, 25 et 27.)

- Millotte, S., René, A., Wales, R., and Christophe, A. (2008). Phonological phrase boundaries constrain the online syntactic analysis of spoken sentences. *J Exp Psychol Learn Mem Cogn*, 34(4) :874–885. (Cité en pages 16 et 61.)
- Mintz, T. H. (2003). Frequent frames as a cue for grammatical categories in child directed speech. *Cognition*, 90(1) :91–117. (Cité en page 43.)
- Mitchell, T. M., Hutchinson, R., Niculescu, R. S., Pereira, F., and Wang, X. (2004). Learning to decode cognitive states from brain images. *Machine Learning*, 57(1-2) :145–175. (Cité en pages 76 et 84.)
- Mitchell, T. M., Shinkareva, S. V., Carlson, A., Chang, K.-M., Malave, V. L., Mason, R. A., and Just, M. A. (2008). Predicting human brain activity associated with the meanings of nouns. *Science*, 320(5880) :1191–1195. (Cité en page 93.)
- Morgan, Shi, and Allopenna (1996). *Signal to syntax : bootstrapping from speech to grammar in early acquisition*, chapter Chap 16 : Perceptual Bases of Rudimentary grammatical categories : Toward a Broader Conceptualization of Bootstrapping. Hillsdale, NJ, England : Lawrence Erlbaum Associates, Inc. (Cité en page 17.)
- Moro, A., Tettamanti, M., Perani, D., Donati, C., Cappa, S. F., and Fazio, F. (2001). Syntax and the brain : disentangling grammar by selective anomalies. *Neuroimage*, 13(1) :110–118. (Cité en page 192.)
- Naigles, L. G. and Kako, E. T. (1993). First contact in verb acquisition : Defining a role for syntax. *Child Development*, 64(6) :1665. (Cité en page 20.)
- Naigles, L. R. (1996). The use of multiple frames in verb learning via syntactic bootstrapping. *Cognition*, 58(2) :221–251. (Cité en page 20.)
- Nelson, D. G. K., Hirsh-Pasek, K., Jusczyk, P. W., and Cassidy, K. W. (1989). How the prosodic cues in motherese might assist language learning. *J Child Lang*, 16(1) :55–68. (Cité en page 15.)
- Nespor, M. and Vogel, I. (1986). *Prosodic Phonology*. Floris, Dordrecht, Holland and Riverton, N.J., USA. (Cité en pages v, 15 et 16.)
- Noppeney, U. and Price, C. J. (2004). An fmri study of syntactic adaptation. *J Cogn Neurosci*, 16(4) :702–713. (Cité en page 69.)
- Norman, K. A., Polyn, S. M., Detre, G. J., and Haxby, J. V. (2006). Beyond mind-reading : multi-voxel pattern analysis of fmri data. *Trends Cogn Sci*, 10(9) :424–430. (Cité en page 76.)
- Ono, K., Nakamura, A., Yoshiyama, K., Kinkori, T., Bundo, M., Kato, T., and Ito, K. (2011). The effect of musical experience on hemispheric lateralization in musical feature processing. *Neurosci Lett*, 496(2) :141–145. (Cité en page 188.)
- Oshima-Takane, Y., Ariyama, J., Kobayashi, T., Katerelos, M., and Poulin-Dubois, D. (2010). Early verb learning in 20-month-old japanese-speaking children. *J Child Lang*, FirstView :1–30. (Cité en pages 21, 22 et 36.)
- Pallier, C., Devauchelle, A.-D., and Dehaene, S. (2011). Cortical representation of the constituent structure of sentences. *Proc Natl Acad Sci U S A*, 108(6) :2522–2527. (Cité en pages vii, 69, 95, 114, 116, 117, 149, 150, 151, 153, 154, 158, 164, 171, 172, 173, 179, 189, 191, 192, 196, 197, 198, 204, 205 et 213.)

- Patel, A. (2003). Language, music, syntax and the brain. *nature neuroscience*, 6(7) :674–681. (Cité en pages 108 et 215.)
- Patel, A. D. (2008). *Music, Language and the Brain*. Oxford University Press, Inc. (Cité en page 109.)
- Patel, A. D., Gibson, E., Ratner, J., Besson, M., and Holcomb, P. J. (1998). Processing syntactic relations in language and music : an event-related potential study. *J Cogn Neurosci*, 10(6) :717–733. (Cité en page 107.)
- Peretz, I. (1993). Auditory atonalia for melodies. *Cognitive Neuropsychology*, 10(1) :21–56. (Cité en page 107.)
- Peretz, I. and Coltheart, M. (2003). Modularity of music processing. *Nat Neurosci*, 6(7) :688–691. (Cité en page 107.)
- Peretz, I., Kolinsky, R., Tramo, M., Labrecque, R., Hublet, C., Demeurisse, G., and Belleville, S. (1994). Functional dissociations following bilateral lesions of auditory cortex. *Brain*, 117 (Pt 6) :1283–1301. (Cité en page 107.)
- Peretz, I. and Zatorre, R. J. (2005). Brain organization for music processing. *Annu Rev Psychol*, 56 :89–114. (Cité en page 186.)
- Pinker, S. (1984). *Language Learnability and language development*. . (Cité en pages 3, 18 et 210.)
- Pinker, S. and Jackendoff, R. (2005). The faculty of language : what’s special about it? *Cognition*, 95(2) :201–236. (Cité en page 4.)
- Potter, M. C. and Lombardi, L. (1998). Syntactic priming in immediate recall of sentences. *Journal of Memory and Language*, 38 :265–282. (Cité en page 76.)
- Price, P. J., Ostendorf, M., Shattuck-Hufnagel, S., and Fong, C. (1991). The use of prosody in syntactic disambiguation. *The Journal of the Acoustical Society of America*, 90(6) :2956–2970. (Cité en page 61.)
- Quine, W. V. (1960). *Word and object*. The MIT Press, Cambridge MA. (Cité en page 10.)
- Rogalsky, C., Rong, F., Saberi, K., and Hickok, G. (2011). Functional anatomy of language and music perception : temporal and structural factors investigated using functional magnetic resonance imaging. *J Neurosci*, 31(10) :3843–3852. (Cité en pages viii, 150, 151, 153, 198 et 213.)
- Rohrmeier, M. (2011). Towards a generative syntax of tonal harmony. *Journal of Mathematics and Music*, 5(1) :35–53. (Cité en page 106.)
- Ryali, S., Supekar, K., Abrams, D. A., and Menon, V. (2010). Sparse logistic regression for whole-brain classification of fmri data. *Neuroimage*, 51(2) :752–764. (Cité en page 86.)
- Santelmann, L. M. and Jusczyk, P. W. (1998). Sensitivity to discontinuous dependencies in language learners : evidence for limitations in processing space. *Cognition*, 69(2) :105–134. (Cité en page 43.)
- Santi, A. and Grodzinsky, Y. (2007). Working memory and syntax interact in broca’s area. *Neuroimage*, 37(1) :8–17. (Cité en page 95.)

- Santi, A. and Grodzinsky, Y. (2010). fmri adaptation dissociates syntactic complexity dimensions. *Neuroimage*, 51(4) :1285–1293. (Cité en pages [vi](#), [75](#), [76](#), [77](#), [95](#) et [213](#).)
- Sastre-Janer, F. A., Regis, J., Belin, P., Mangin, J. F., Dormont, D., Masure, M. C., Remy, P., Frouin, V., and Samson, Y. (1998). Three-dimensional reconstruction of the human central sulcus reveals a morphological correlate of the hand area. *Cereb Cortex*, 8(7) :641–647. (Cité en page [163](#).)
- Saur, D., Kreher, B. W., Schnell, S., Kümmerer, D., Kellmeyer, P., Vry, M.-S., Umarova, R., Musso, M., Glauche, V., Abel, S., Huber, W., Rijntjes, M., Hennig, J., and Weiller, C. (2008). Ventral and dorsal pathways for language. *Proc Natl Acad Sci U S A*, 105(46) :18035–18040. (Cité en page [203](#).)
- Schmithorst, V. J. (2005). Separate cortical networks involved in music perception : preliminary functional mri evidence for modularity of music processing. *Neuroimage*, 25(2) :444–451. (Cité en page [108](#).)
- Shetreet, E., Friedmann, N., and Hadar, U. (2009). An fmri study of syntactic layers : sentential and lexical aspects of embedding. *Neuroimage*, 48(4) :707–716. (Cité en page [95](#).)
- Shi, R. and Lepage, M. (2008). The effect of functional morphemes on word segmentation in preverbal infants. *Developmental Science*, 11(3) :407–413. (Cité en pages [17](#) et [20](#).)
- Shi, R. and Melancon, A. (2010). Syntactic categorization in french-learning infants. *Infancy*, 15 :517–533. (Cité en pages [21](#), [22](#), [23](#) et [36](#).)
- Shi, R., Morgan, J. L., and Allopenna, P. (1998). Phonological and acoustic bases for earliest grammatical category assignment : a cross-linguistic perspective. *Journal of Child Language*, 25(1) :169–201. (Cité en pages [17](#) et [18](#).)
- Shmuel, A., Augath, M., Oeltermann, A., and Logothetis, N. K. (2006). Negative functional mri response correlates with decreases in neuronal activity in monkey visual area v1. *Nat Neurosci*, 9(4) :569–577. (Cité en page [189](#).)
- Slevc, L. R., Rosenberg, J. C., and Patel, A. D. (2009). Making psycholinguistics musical : Self-paced reading time evidence for shared processing of linguistic and musical syntax. *Psychonomic Bulletin & Review*, 16(2) :374–381. (Cité en page [215](#).)
- Sluming, V., Barrick, T., Howard, M., Cezayirli, E., Mayes, A., and Roberts, N. (2002). Voxel-based morphometry reveals increased gray matter density in broca’s area in male symphony orchestra musicians. *NeuroImage*, 17(3) :1613 – 1622. (Cité en page [187](#).)
- Smith, M. and Wheeldon, L. (2001). Syntactic priming in spoken sentence production - an online study. *Cognition*, 78 :123–164. (Cité en page [76](#).)
- Snijders, T. M., Vosse, T., Kempen, G., Berkum, J. J. A. V., Petersson, K. M., and Hagoort, P. (2009). Retrieval and unification of syntactic structure in sentence comprehension : an fmri study using word-category ambiguity. *Cereb Cortex*, 19(7) :1493–1503. (Cité en pages [viii](#), [153](#), [189](#), [190](#), [192](#), [196](#) et [213](#).)
- Stolterfoht, B., Friederici, A. D., Alter, K., and Steube, A. (2007). Processing focus structure and implicit prosody during reading : differential erp effects. *Cognition*, 104(3) :565–590. (Cité en page [70](#).)

- Strelnikov, K., Vorobyev, V., Chernigovskaya, T., and Medvedev, S. (2006). Prosodic clues to syntactic processing—a pet and erp study. *NeuroImage*, 29(4) :1127 – 1134. (Cité en page 96.)
- Stromswold, K., Caplan, D., Alpert, N., and Rauch, S. (1996). Localization of syntactic comprehension by positron emission tomography. *Brain Lang*, 52(3) :452–473. (Cité en page 75.)
- Thirion, B., Flandin, G., Pinel, P., Roche, A., Ciuciu, P., and Poline, J.-B. (2006). Dealing with the shortcomings of spatial normalization : multi-subject parcellation of fmri datasets. *Hum Brain Mapp*, 27(8) :678–693. (Cité en page 86.)
- Thompson-Schill, S. L. (2005). *twenty-first century psycholinguistics : Four corners-tones*, chapter Dissecting the language organ : a new look at the role of Broca’s area in language processing, pages 173–189. Lawrence Erlbaum Associates, Inc. (Cité en page 195.)
- Thompson-Schill, S. L., D’Esposito, M., Aguirre, G. K., and Farah, M. J. (1997). Role of left inferior prefrontal cortex in retrieval of semantic knowledge : a reevaluation. *Proc Natl Acad Sci U S A*, 94(26) :14792–14797. (Cité en page 95.)
- Tillmann, B., Bharucha, J. J., and Bigand, E. (2000). Implicit learning of tonality : a self-organizing approach. *Psychol Rev*, 107(4) :885–913. (Cité en page 109.)
- Tillmann, B., Janata, P., and Bharucha, J. J. (2003). Activation of the inferior frontal cortex in musical priming. *Brain Res Cogn Brain Res*, 16(2) :145–161. (Cité en page 108.)
- Toivonen and Izumo (2004). Timidity++. from <http://timidity.sourceforge.net/>. (Cité en page 124.)
- Tomasello, M. (1992). The social bases of language acquisition. *Social Development*, 1(1) :67 – 87. (Cité en page 10.)
- Tomasello, M. and Akhtar, N. (1995). Two-year-olds use pragmatic cues to differentiate reference to objects and actions. *Cognitive development*, 10 :201–224. (Cité en page 10.)
- Tyler, L. K. and Marslen-Wilson, W. (2008). Fronto-temporal brain systems supporting spoken language comprehension. *Philosophical Transactions of the Royal Society B : Biological Sciences*, 363(1493) :1037–1054. (Cité en page 97.)
- Tyler, L. K., Shafto, M. A., Randall, B., Wright, P., Marslen-Wilson, W. D., and Stamatakis, E. A. (2010). Preserving syntactic processing across the adult life span : the modulation of the frontotemporal language system in the context of age-related atrophy. *Cereb Cortex*, 20(2) :352–364. (Cité en page 192.)
- Van Heugten, M. and Shi, R. (2009). French-learning toddlers use gender information on determiners during word recognition. *Dev Sci*, 12(3) :419–425. (Cité en pages 44 et 50.)
- Van Heugten, M. and Shi, R. (2010). Infants’ sensitivity to non-adjacent dependencies across phonological phrase boundaries. *J Acoust Soc Am*, 128(5) :EL223–EL228. (Cité en page 43.)
- Veneziano, E. and Parisse, C. (2010). The early comprehension of noun-verb distinction in french : An experimental method. In *XVIIth Biennial International Conference on Infant Studies*, Baltimore, Maryland, USA. (Cité en page 21.)

- Vuust, P., Roepstorff, A., Wallentin, M., Mouridsen, K., and Østergaard, L. (2006). It don't mean a thing... keeping the rhythm during polyrhythmic tension, activates language areas (ba47). *Neuroimage*, 31(2) :832–841. (Cité en page 216.)
- Vuust, P., Wallentin, M., Mouridsen, K., Ostergaard, L., and Roepstorff, A. (2011). Tapping polyrhythms in music activates language areas. *Neurosci Lett*, 494(3) :211–216. (Cité en page 216.)
- Wagner, M. (2005). *Prosody and Recursion*. PhD thesis, Massachusetts Institut of Technology. (Cité en page 72.)
- Wagner, M. (2010). Prosody and recursion in coordinate structures and beyond. *Natural Language and Linguistic Theory*, 28 :183–237. 10.1007/s11049-009-9086-0. (Cité en page 70.)
- Ward, J. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, 58(301) :236–244. (Cité en page 86.)
- Warren, J. D., Uppenkamp, S., Patterson, R. D., and Griffiths, T. D. (2003). Separating pitch chroma and pitch height in the human brain. *Proc Natl Acad Sci U S A*, 100(17) :10038–10042. (Cité en page 186.)
- Waxman, S. R. and Booth, A. E. (2001). Seeing pink elephants : fourteen-month-olds' interpretations of novel nouns and adjectives. *Cogn Psychol*, 43(3) :217–242. (Cité en page 21.)
- Waxman, S. R., Lidz, J. L., Braun, I. E., and Lavin, T. (2009). Twenty four-month-old infants' interpretations of novel verbs and nouns in dynamic scenes. *Cognitive Psychology*, 59(1) :67–95. (Cité en pages 21, 22 et 36.)
- Weber, A., Grice, M., and Crocker, M. W. (2006). The role of prosody in the interpretation of structural ambiguities : a study of anticipatory eye movements. *Cognition*, 99(2) :B63–B72. (Cité en page 61.)
- Weber, K. and Indefrey, P. (2009). Syntactic priming in german-english bilinguals during sentence comprehension. *Neuroimage*, 46(4) :1164–1172. (Cité en page 69.)
- Weintraub, S., Mesulam, M. M., and Kramer, L. (1981). Disturbances in prosody. a right-hemisphere contribution to language. *Arch Neurol*, 38(12) :742–744. (Cité en page 96.)
- Werker, J. F. and Tees, R. C. (1984). Cross-language speech perception : Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*, 7 :49–63. (Cité en page 2.)
- Xu, J., Kemeny, S., Park, G., Frattali, C., and Braun, A. (2005). Language in context : emergent features of word, sentence, and narrative comprehension. *Neuroimage*, 25(3) :1002–1015. (Cité en page 95.)
- Yamashita, O., aki Sato, M., Yoshioka, T., Tong, F., and Kamitani, Y. (2008). Sparse estimation automatically selects voxels relevant for the decoding of fmri activity patterns. *Neuroimage*, 42(4) :1414–1429. (Cité en page 92.)
- Yuan, S. and Fisher, C. (2009). "really? she blicked the baby?" : two-year-olds learn combinatorial facts about verbs by listening. *Psychological Science*, 20(5) :619–626. (Cité en page 20.)

Zangl, R. and Fernald, A. (2007). Increasing flexibility in children's online processing of grammatical and nonce determiners in fluent speech. *Language Learning and Development*, 3 :199–231. (Cité en pages 18, 22, 23 et 36.)

Résumé :

Ce travail de thèse porte sur la structure syntaxique des phrases. La première partie de ce travail examine dans quelle mesure les mots grammaticaux et la prosodie sont utilisés pour identifier les catégories syntaxiques (Nom et Verbe). Nous montrons, grâce à une expérience comportementale avec des enfants de 18 mois, qu'ils savent déjà que les déterminants précèdent les noms et que les pronoms précèdent les verbes. Les enfants, à l'âge où ils apprennent le sens des mots, ne disposent que des informations syntaxiques et prosodiques. Trois expériences comportementales prouvent que les adultes utilisent ces indices pour déterminer automatiquement la catégorie grammaticale de pseudo-mots. Ces données nous conduisent à proposer un modèle d'accès lexical s'appuyant sur la prosodie et les mots grammaticaux pour sélectionner les mots les plus probables. Ce modèle rend également compte de l'apprentissage de nouveaux mots. La deuxième partie de ce travail explore les réseaux cérébraux impliqués dans le traitement de la structure syntaxique. Nous examinons si les traitements des structures dans le langage et dans la musique partagent les mêmes réseaux. Deux expériences d'IRMf utilisant la langue parlée montrent que les aires inféro-frontales (BA45 et BA47), temporelles (pôle temporal, aSTS et pSTS) et temporo-pariétale sont recrutées pour le traitement de la structure syntaxique linguistique. Dans une seconde étude, nous montrons qu'un sous-ensemble de ces régions est impliqué dans le traitement de la structure musicale.

Mots clés : Langage, Musique, Syntaxe, Acquisition, Traitement Cérébral, IRMf, Prosodie, Accès lexical

Syntactic structures processing in language and music

Abstract : This dissertation deals with syntactic structures. In the first part of this work, we examine how function words and prosody are exploited to recognize syntactic categories (nouns and verbs). With a behavioral experiment testing 18-months-olds, we showed that toddlers already know that determiners precede nouns and that pronouns occur before verbs in the case of known words. Furthermore, our behavioral experiments on adults proved that adult speakers use prosodic cues and function words to compute online syntactic categories for pseudowords. Those data lead us to propose a model which rely on prosody and function words to constrain lexical access. This model would also explain the acquisition of new words. In the second part of our work, we explored cerebral networks processing syntactic structure. We questioned if those networks involved in syntactic processing in language were shared with other domains such as music. Two fMRI experiments using spoken language showed that infero-frontal regions (BA45 and BA47), temporal regions (temporal pole, aSTS, pSTS) and the temporo-parietal junction are recruited to process syntactic structure in language. In a second study, we showed that a part of these networks is involved in syntactic structure processing in music.

Keywords : Language, Music, Syntax, Acquisition, Cerebral Processing, fMRI, Prosody, lexical access
