

UNIVERSITÉ PAUL SABATIER – TOULOUSE III
FACULTÉ DE MÉDECINE TOULOUSE RANGUEIL
ENSEIGNEMENT DES TECHNIQUES DE RÉADAPTATION

Mémoire présenté en vue de l'obtention du Certificat de Capacité d'Orthophonie

ADAPTATION DU PROJET APHASIABANK
À LA LANGUE FRANÇAISE

—

Contribution pour une évaluation informatisée
du discours oral de patients aphasiques

Capucine COLIN
Camille LE MEUR

Sous la direction de :

Halima SAHRAOUI

*Docteur en Sciences du langage, Maître de conférences, Université de Toulouse 2 Jean-Jaurès
Laboratoire Octogone-Lordat EA4156*

Katia LABRUNEE – PROD'HOMME

*Orthophoniste et Docteur en Sciences du langage
Ecole d'Orthophonie de Toulouse, Université Paul Sabatier*

UNIVERSITÉ PAUL SABATIER – TOULOUSE III
FACULTÉ DE MÉDECINE TOULOUSE RANGUEIL
ENSEIGNEMENT DES TECHNIQUES DE RÉADAPTATION

Mémoire présenté en vue de l'obtention du Certificat de Capacité d'Orthophonie

ADAPTATION DU PROJET APHASIABANK
À LA LANGUE FRANÇAISE

—

Contribution pour une évaluation informatisée
du discours oral de patients aphasiques

Capucine COLIN
Camille LE MEUR

Sous la direction de :

Halima SAHRAOUI

*Docteur en Sciences du langage, Maître de conférences, Université de Toulouse 2 Jean-Jaurès
Laboratoire Octogone-Lordat EA4156*

Katia LABRUNEE – PROD'HOMME

*Orthophoniste et Docteur en Sciences du langage
Ecole d'Orthophonie de Toulouse, Université Paul Sabatier*

REMERCIEMENTS

Nous tenons à remercier principalement nos maîtres de mémoire, Halima Sahraoui et Katia Labrunée Prod'homme, sans qui nous n'aurions pas pu mener ce projet. Nous vous remercions pour la confiance que vous nous avez accordée et pour la qualité de nos échanges qui nous ont beaucoup apporté et ont enrichi notre réflexion. Merci également pour votre grande disponibilité, vos encouragements, et votre bonne humeur, durant toute cette année.

Nous tenons également à remercier toutes les personnes qui ont accepté de participer à notre protocole. Merci pour votre accueil et le temps que vous nous avez accordé.

Nous remercions tous les orthophonistes qui nous ont mis en relation avec leurs patients afin de nous aider dans la réalisation de ce projet.

Nous remercions chaleureusement les concepteurs du projet AphasiaBank, particulièrement Brian MacWhinney, Audrey Holland, Margie Forbes et Davida Fromm pour leur engagement, leur disponibilité et leur gentillesse. Merci pour toute l'aide que vous nous avez apportée durant toutes les étapes de notre projet et pour la richesse de nos échanges.

Nous tenons enfin à remercier nos familles et nos proches pour leur soutien durant l'ensemble de ces quatre années.

ADAPTATION DU PROJET APHASIABANK À LA LANGUE FRANÇAISE

–

Contribution pour une évaluation informatisée du discours oral de patients aphasiques

*Mémoire présenté en vue de l'obtention du Certificat de Capacité d'Orthophoniste, à la Faculté de
Médecine Toulouse-Rangueil le 30 juin 2016*

Par Capucine COLIN et Camille LE MEUR

Sous la direction de Halima SAHRAOUI et Katia LABRUNÉE PROD'HOMME

1. Introduction et problématique

Les accidents vasculaires cérébraux (AVC) représentent la première cause de handicap acquis de l'adulte. L'aphasie résultant d'un AVC peut considérablement perturber la communication entre la personne aphasique et son entourage. Sa prise en charge est donc essentielle afin d'améliorer la qualité de vie des personnes aphasiques.

Le bilan orthophonique permet d'adapter au mieux la prise en charge des patients aphasiques. Le discours oral représente un enjeu majeur, car il est central dans la communication de la personne aphasique, c'est pourquoi il est évalué dans de nombreux bilans orthophoniques. Les critères retenus sont néanmoins essentiellement qualitatifs et subjectifs. Une analyse plus objective basée sur des critères quantitatifs et qualitatifs précis pourrait contribuer à une évaluation plus complète du discours oral de la personne aphasique.

C'est dans cette perspective qu'a été conçu le programme EVAL, qui fonctionne via le logiciel CLAN, associé à la base de données AphasiaBank. Le projet AphasiaBank, mis au point par MacWhinney, Fromm, Holland et Forbes, est constitué d'un protocole de recueil de données en langue anglaise permettant de recueillir des échantillons de discours oral et de les analyser de façon automatique grâce au programme EVAL. Ce programme permet ainsi d'obtenir des données précises sur le profil discursif du sujet, en fournissant des indicateurs sur les aspects formels de son discours oral continu. Les résultats obtenus sont ensuite comparés aux données incluses dans la base de données, comprenant des corpus recueillis auprès d'un certain nombre de sujets contrôles et aphasiques, permettant de situer les performances du patient par rapport à d'autres sujets aphasiques ou non aphasiques.

Notre propos est d'explorer la possibilité d'utiliser le protocole AphasiaBank et le programme EVAL sur des sujets contrôles et aphasiques francophones, et ainsi de questionner la possibilité et la pertinence de l'utilisation d'EVAL en clinique orthophonique. Nous espérons ainsi contribuer à la mise en place d'un nouvel outil d'évaluation du discours oral aphasique.

2. Méthodologie

Afin de répondre à notre problématique de recherche, nous avons posé les hypothèses suivantes :

- Le protocole AphasiaBank est adaptable à la langue française
- Le protocole AphasiaBank est utilisable avec des sujets contrôles et des sujets aphasiques francophones
- EVAL est opérationnel avec des corpus issus de sujets contrôles et de sujets aphasiques francophones

Afin de répondre à ces hypothèses nous avons procédé à une adaptation du protocole AphasiaBank à la langue française. Pour ce faire, nous avons dû adapter le protocole de recueil de données démographiques, le protocole de recueil de données discursives et le protocole de recueil de données issues des épreuves complémentaires destinées à évaluer les autres domaines du langage oral des participants aphasiques. Nous avons essayé de rester le plus proche possible du protocole AphasiaBank original. Cependant, une traduction littérale des épreuves discursives et complémentaires n'aurait pas été pertinente. Concernant l'adaptation des épreuves complémentaires, nous avons choisi de proposer des épreuves équivalentes couramment utilisées en pratique clinique orthophonique en France, ce qui nous a conduit à proposer des épreuves différentes de celles du protocole original.

Nous avons ensuite procédé à la passation de notre protocole adapté auprès d'une population constituée de 14 sujets contrôles et de 11 sujets aphasiques, afin d'initier la création de la base de données française de référence. Les données ainsi recueillies ont ensuite été transcrites et codées sur le logiciel CLAN et ont été préparées à l'analyse informatisée du programme EVAL. Cette étape a nécessité la sélection des normes de transcription et de codage pertinentes, la transcription et le codage des données selon ces normes, la vérification informatisée des transcriptions et le lancement du programme d'étiquetage morphologique. Cette dernière étape implique la mise à jour des fichiers lexiques de la grammaire française du logiciel. Une fois les données langagières préparées à l'analyse EVAL, nous les avons intégrées à la base de données AphasiaBank afin que le programme EVAL puisse procéder à des comparaisons entre les corpus de langue française. Enfin, nous avons lancé le programme EVAL sur nos corpus.

3. Résultats

Nous avons pu effectuer une adaptation française de l'ensemble du protocole de recueil de données AphasiaBank. Les protocoles de recueil de données démographiques et discursives ont été adaptés en procédant à certaines modifications afin que le protocole proposé soit plus précis sur certains points et qu'il corresponde davantage aux particularités culturelles françaises. Nous avons également pu adapter le protocole de recueil de données issues des épreuves complémentaires en proposant un équivalent pour tous les tests présentés même si cela a nécessité de s'éloigner davantage du protocole original.

Une fois le protocole adapté, les passations ont pu se dérouler avec tous les participants contrôles et aphasiques. Nous avons ainsi pu recueillir des données démographiques et langagières pour tous les participants. La passation du protocole auprès des participants contrôles comme des participants aphasiques a permis de mettre en évidence certaines améliorations à apporter au protocole adapté.

L'ensemble de nos données a pu être transcrit et codé selon les normes CHAT. Les corpus ont ensuite pu être vérifiés par le programme de vérification et les erreurs détectées ont pu être corrigées. Le programme d'étiquetage morphologique a également pu être appliqué sur nos corpus. Les transcriptions et les fichiers audio correspondants ont ensuite pu être mis en ligne sur la base de données AphasiaBank. Le programme EVAL a pu être lancé sur nos transcriptions mais les résultats obtenus n'ont pu être exploités car EVAL ne fonctionne pas encore de façon opérationnelle sur les corpus en langue française. Les informaticiens du projet AphasiaBank vont prochainement travailler sur une reconfiguration du programme afin qu'EVAL fonctionne sur des données en langue française.

4. Discussion et conclusion

Nous nous sommes tout d'abord interrogées sur la possibilité d'adapter le protocole AphasiaBank à la langue française. L'ensemble du protocole a effectivement pu être adapté. L'hypothèse 1 est donc validée.

Nous avons ensuite testé la possibilité d'utilisation du protocole auprès de sujets contrôles et aphasiques francophones. L'ensemble du protocole a pu être utilisé auprès de tous les participants, l'hypothèse 2 est donc validée.

Nous avons enfin postulé qu'EVAL était opérationnel avec des corpus issus de sujets contrôles et de sujets aphasiques francophones. L'ensemble des étapes nécessaires à la préparation des corpus à l'analyse EVAL a pu être effectué, mais le programme EVAL n'est pas encore configuré pour fonctionner de façon optimale sur des corpus en langue française. De ce fait, l'hypothèse 3 est invalidée.

Nous avons ainsi pu constituer une adaptation française du protocole de recueil de données AphasiaBank qui est désormais accessible aux cliniciens et aux chercheurs membres de la communauté AphasiaBank. Cependant, nous n'avons pas pu mener notre travail à son terme puisque l'utilisation effective d'EVAL sur nos corpus n'a pas permis d'obtenir des analyses exploitables. L'adaptation et l'utilisation d'EVAL à la langue française doivent donc encore se poursuivre afin de permettre son utilisation optimale par les cliniciens et les chercheurs.

L'utilisation d'EVAL en clinique orthophonique pourrait à terme constituer un outil supplémentaire à disposition du clinicien afin d'améliorer l'évaluation du discours oral chez la personne aphasique, et ainsi contribuer à l'établissement d'une prise en charge la plus adaptée possible. Cependant, ce travail devra encore faire l'objet d'études ultérieures afin de faire davantage correspondre le protocole de recueil de données et le programme EVAL aux besoins spécifiques de la clinique orthophonique, mais également d'évaluer précisément l'utilisabilité et l'acceptabilité, ainsi que l'utilité d'un tel outil en clinique.

Bibliographie

Charaudeau P, Maingueneau D. *Dictionnaire d'analyse du discours*. Paris : Editions du Seuil; 2002:661.

Chomel-Guillaume S, Leloup G, Bernard I. *Les aphasies, évaluation et rééducation*. Issy-les-Moulineaux: Elsevier; 2010: 268.

EVAL Manual. (page consultée le 08/04/16). [en ligne].
<http://talkbank.org/APhasiaBank/manuals/eval.pdf>

Fromkin AV, Ratner NB. *Speech Production*. In: Gleason JB, Ratner NB. editors. *Psycholinguistics*. 2nd ed. New York: Harcourt Brace; 1998. p.309-346.

MacWhinney B. *The CHILDES Project: Tools for Analyzing Talk 3rd Edition Part 1: The CHAT Transcription Format*. Mahwah, NJ: Lawrence Erlbaum Associates (ed). 2000a. [en ligne] [mise à jour le 14/04/16 et consulté le 14/04/16]. Disponibilité sur internet :
<<http://childes.talkbank.org/manuals/chat.pdf>>

MacWhinney B. *The CHILDES Project: Tools for Analyzing Talk 3rd Edition Part 2: The CLAN Programs*. Mahwah NJ : Lawrence Erlbaum Associates (ed). 2000b. [en ligne] [mise à jour le 06/04/16 et consulté le 17/04/16]. Disponibilité sur internet :
<<http://childes.talkbank.org/manuals/clan.pdf>>

MacWhinney B, Fromm D, Forbes M, Holland A. *AphasiaBank : Methods for studying discourse*. *Aphasiology*. Nov 2011;25(11):1286-1307.

AphasiaBank. (dernière date de consultation : 22/05/16). [en ligne].
<http://talkbank.org/APhasiaBank/>

SOMMAIRE

INTRODUCTION	1
PARTIE 1 : PARTIE THEORIQUE	3
1. Le discours	3
1.1. Le discours : approche linguistique	3
1.1.1. Le discours : généralités (d'après Charaudeau et Maingueneau, 2002)	3
1.1.2. Les différents modes d'organisation du discours : prolongement pour la clinique	6
1.1.3. Notion de compétence discursive et enjeu pour la clinique	7
1.1.4. Les règles du discours : apport de la pragmatique	8
1.1.5. Les particularités du discours oral	9
1.2. Le discours oral : approche psycholinguistique (d'après Fromkin et Ratner, 1998).....	12
1.2.1. Les phénomènes de l'oral : une fenêtre sur les processus psycholinguistiques sous-jacents	13
1.2.2. Les principaux modèles de production du discours oral.....	18
2. L'aphasie	22
2.1. Définition de l'aphasie	22
2.2. Etiologie des aphasies.....	23
2.3. Perturbations langagières orales observées	23
2.4. Les différents types d'aphasie	24
2.4.1. Les aphasies à langage réduit (d'après Gil, 1999).....	25
2.4.2. Les aphasies à langage fluide (d'après Gil, 1999).....	26
2.5. Handicap et communication dans l'aphasie : mieux évaluer le discours déviant.....	27
3. L'évaluation en clinique orthophonique	29
3.1. L'évaluation orthophonique de l'aphasie.....	29
3.1.1. Domaines de l'évaluation et outils existants	29
3.1.2. L'analyse et l'interprétation des résultats	33
3.2. L'évaluation orthophonique du discours continu oral en aphasiologie.....	33
3.2.1. Les paramètres d'évaluation du discours	34
3.2.2. Les outils d'évaluation orthophoniques du discours en aphasiologie	34
3.2.3. Synthèse des outils actuels.....	38
4. L'utilisation des outils informatiques dans l'étude du discours en aphasiologie	39
4.1. Outils informatisés actuels dédiés à l'évaluation orthophonique de l'aphasie en France...39	
4.2. Les bases de données et les projets de développement collaboratif dans l'étude du discours en aphasiologie	41
4.2.1. Les principes d'une base de données.....	41
4.2.2. La base de données internationale AphasiaBank : intérêt pour la recherche fondamentale.....	42
4.3. Présentation des outils utilisés pour le traitement automatique informatisé du langage oral au sein du projet AphasiaBank	45
4.3.1. Le système CHAT	45
4.3.2. Le logiciel CLAN	46
4.4. Analyse informatisée du discours en aphasiologie : fonctionnement du programme EVAL	46
4.4.1. EVAL : généralités	46
4.4.2. Préparation d'une transcription pour l'analyse EVAL	47
4.4.3. Les mesures linguistiques calculées par EVAL	50
4.4.4. Le lancement du programme EVAL	52

PARTIE 2 : PROBLEMATIQUE ET HYPOTHESES	54
1. Problématique	54
1.1. Raisons du choix de cette problématique : Intérêts potentiels de l'analyse informatisée par EVAL du discours aphasique dans l'évaluation clinique orthophonique	54
1.1.1. L'importance de l'évaluation du discours en aphasiologie	54
1.1.2. Apport potentiel d'EVAL par rapport aux autres outils d'évaluation du discours	54
1.1.3. L'intérêt de l'utilisation d'EVAL dans l'évaluation du discours	55
12. Problématique résumée	57
2. Hypothèses	58
PARTIE 3 : PARTIE EXPERIMENTALE	60
1. Adaptation du protocole de recueil de données AphasiaBank	60
1.1. Méthodologie dédiée à l'adaptation du protocole de recueil de données AphasiaBank ...	60
1.1.1. Adaptation des documents permettant le recueil des données démographiques	61
1.1.2. Adaptation du protocole de recueil de données langagières AphasiaBank	64
1.2. Résultats concernant l'adaptation du protocole de recueil de données AphasiaBank	74
1.2.1. Résultats concernant l'adaptation des documents permettant le recueil de données démographiques	74
1.2.2. Résultats concernant l'adaptation des épreuves de discours AphasiaBank	75
2. Population	79
2.1. Critères de sélection des sujets	79
2.1.1. Critères de sélection de la population contrôle	79
2.1.2. Critères de sélection de la population aphasique	80
2.2. Description de l'échantillon de population	81
2.2.1. Description de la population contrôle	81
2.2.2. Description de la population aphasique	82
3. Le recueil de données	85
3.1. Méthodologie dédiée au recueil de données	85
3.1.1. Création d'une lettre d'information et d'un formulaire de consentement éclairé	85
3.1.2. Le recueil de données démographiques	85
3.1.3. Le recueil de données langagières	85
3.2. Résultats concernant l'utilisation du protocole AphasiaBank avec des sujets contrôles et des sujets aphasiques francophones	86
3.2.1. Résultats concernant le recueil de données démographiques	86
3.2.2. Résultats concernant le recueil de données langagières	87
4. L'opérationnalisation des données pour l'utilisation d'EVAL	94
4.1. Méthodologie de l'opérationnalisation des données pour l'utilisation d'EVAL	94
4.1.1. Sélection des conventions de transcription et des codages pertinents	94
4.1.2. Transcription et codage des données	99
4.1.3. Vérification des transcriptions	105
4.1.4. Alignement des transcriptions avec les fichiers audio	105
4.1.5. Le lancement du programme d'étiquetage morphologique	105
4.1.6. Contribution à la base de données AphasiaBank	107
4.1.7. Le lancement du programme EVAL	108
4.2. Résultats de l'opérationnalisation des données pour l'utilisation d'EVAL	108
4.2.1. Transcription et codage des données selon les normes CHAT	108
4.2.2. Vérification des transcriptions par le programme CHECK	112
4.2.3. Alignement aux fichiers audio	113
4.2.4. L'étape d'étiquetage morphologique	113
4.2.5. La mise en ligne de nos corpus sur la base de données	115
4.2.6. Le lancement du programme EVAL	116

PARTIE 4 : DISCUSSION DES RÉSULTATS	117
1. Interprétation des résultats et validation ou invalidation des hypothèses	117
1.1. Hypothèse 1 : Le protocole AphasiaBank est adaptable à la langue française	117
1.1.1. Sous-hypothèse 1 : Le protocole de recueil de données démographiques AphasiaBank est adaptable à la langue française	117
1.1.2. Sous-hypothèse 2 : Le protocole de recueil de données discursives AphasiaBank est adaptable à la langue française.....	118
1.1.3. Sous-hypothèse 3 : Le protocole de recueil de données issues des épreuves complémentaires AphasiaBank est adaptable à la langue française.....	120
1.2. Hypothèse 2 : Le protocole AphasiaBank est utilisable avec des sujets contrôles et des sujets aphasiques francophones	123
1.2.1. Sous-hypothèse 1 : Le protocole de recueil de données démographiques AphasiaBank est utilisable avec des sujets contrôles et des sujets aphasiques francophones	124
1.2.2. Sous-hypothèse 2 : Le protocole de recueil de données discursives AphasiaBank est utilisable avec des sujets contrôles et des sujets aphasiques francophones	125
1.2.3. Sous-hypothèse 3 : Le protocole de recueil de données des épreuves complémentaires AphasiaBank est utilisable avec des sujets contrôles et des sujets aphasiques francophones	129
1.3. Hypothèse 3 : EVAL est opérationnel avec des corpus issus de sujets contrôles et de sujets aphasiques francophones	130
1.3.1. Sous-hypothèse 1 Les corpus francophones peuvent être transcrits et codés selon les normes CHAT	130
1.3.2. Sous-hypothèse 2 : Les corpus francophones peuvent être vérifiés par le programme CHECK et les erreurs détectées peuvent être corrigées.....	135
1.3.3. Sous-hypothèse 3 : Le programme d'étiquetage morphologique peut être lancé sur les corpus francophones	136
1.3.4. Sous-hypothèse 4 : Les corpus francophones peuvent être mis en ligne sur la base de données française d'AphasiaBank.....	137
1.3.5. Sous-hypothèse 5 : Le programme EVAL peut être lancé de façon opérationnelle sur les corpus francophones.....	139
2. Les limites	139
2.1. Concernant la population	139
2.2. Concernant le projet AphasiaBank	140
2.2.1. Limites des critères d'inclusion.....	140
2.2.2. Limites de la transcription et du codage des données	141
3. Perspectives	143
3.1. Perspectives en recherche fondamentale	143
3.1.1. Perspectives à court terme.....	143
3.1.2. Perspectives à moyen et long terme.....	144
3.2. Perspectives concernant l'utilisation du protocole AphasiaBank et d'EVAL en clinique orthophonique.....	145
3.2.1. Perspectives à court terme.....	145
3.2.2. Perspectives à moyen et long terme.....	145
CONCLUSION	149
BIBLIOGRAPHIE	150
BIBLIOGRAPHIE D'AUTEURS CITÉS PAR D'AUTRES AUTEURS	153
LISTE DES FIGURES, TABLEAUX, DIAGRAMMES	156
TABLE DES MATIERES	157
RESUME	173

GLOSSAIRE

La base de données AphasiaBank : La base de données regroupe des échantillons de discours oral de sujets contrôles et aphasiques. Ces échantillons sont recueillis grâce à un protocole de recueil de données permettant de recueillir des données démographiques, discursives et langagières.

Système CHAT : Les échantillons stockés sur la base de données AphasiaBank sont tous transcrits et codés selon un système de conventions de transcription et de codage uniformisé appelé le système CHAT (*Codes for the Human Analysis of Transcripts*). Ce système explicite les règles permettant de transcrire les données langagières afin que celles-ci puissent ensuite être analysées par le logiciel CLAN.

Logiciel CLAN : Le logiciel CLAN (*Computerised Language Analysis*) est un logiciel libre et gratuit qui permet d'effectuer de multiples analyses informatisées sur des données langagières transcrites en format CHAT.

Programme EVAL : Le programme EVAL est l'un des nombreux programmes proposés par le logiciel CLAN. Il permet d'analyser les productions langagières d'un sujet et de dresser un profil discursif de ses performances linguistiques par la production d'un tableau regroupant 30 mesures qualitatives et quantitatives calculées automatiquement par CLAN. Le programme permet également de comparer les résultats obtenus à ceux d'un groupe de comparaison de la base de données en fournissant des écart-types pour chacune de ces 30 mesures.

Programme CHECK : Le programme CHECK est un programme intégré à CLAN qui permet d'effectuer une vérification informatisée des transcriptions. Ce programme vérifie que les corpus ont été transcrits de façon à respecter les normes de transcription CHAT et détecte les erreurs relatives au non respect de ces normes. Les erreurs détectées par le programme CHECK doivent être corrigées afin de permettre à CLAN de fonctionner de façon optimale sur les transcriptions.

Programme MOR : Le programme MOR est un programme d'étiquetage morphologique. Il permet d'attribuer à chaque mot de la transcription la catégorie morphologique correspondante (nom, verbe, adverbe, etc.), ce qui permettra ensuite au programme EVAL d'effectuer des calculs sur ces différentes catégories. Ce programme permet d'afficher sous chaque énoncé une ligne d'étiquetage morphologique qui contient la catégorie morphologique de chaque mot de la ligne principale.

INTRODUCTION

Les accidents vasculaires cérébraux (AVC) représentent la troisième cause de mortalité dans les pays industrialisés et la première cause de handicap acquis de l'adulte (Chomel-Guillaume et al., 2010). En France, 150 000 AVC surviennent chaque année et un tiers des patients reste dépendant à 6 mois (Chomel-Guillaume et al., 2010). Les AVC sont responsables de diverses séquelles, motrices, sensorielles, langagières et/ou cognitives. Parmi l'ensemble de ces séquelles, la présence d'une aphasie peut considérablement perturber la communication entre la personne aphasique et son entourage. La prise en charge de l'aphasie constitue donc un enjeu majeur afin d'améliorer la qualité de vie de ces personnes suite à leur AVC.

En permettant d'identifier les capacités et les difficultés propres à chaque patient, l'évaluation orthophonique permet d'adapter au mieux la prise en charge des personnes aphasiques. Il existe de nombreuses batteries d'évaluation du langage et de la communication dans le cadre de l'aphasie en orthophonie, telles que *l'échelle française pour l'examen de l'aphasie (HDAE)* ou le *protocole Montréal-Toulouse d'examen linguistique de l'aphasie (MT86)*. Ces différents protocoles évaluent divers aspects du langage oral et écrit.

Parmi ces aspects, le discours oral, central dans la communication, est un domaine clé. Il est évalué dans la majeure partie des bilans orthophoniques existants en aphasie. Mais l'analyse qui en est proposée repose essentiellement sur des critères subjectifs liés à l'intuition clinique de l'orthophoniste, fruit de son expérience et de sa pratique de l'utilisation des batteries de tests « papier-crayon » dans lesquels les critères de cotation du discours continu oral sont qualitatifs. Une évaluation plus objectivée selon des critères quantitatifs et qualitatifs, automatisée et plus affinée, pourrait contribuer à une évaluation du langage oral de la personne aphasique plus complète, afin de proposer une prise en charge plus adaptée.

C'est dans cette perspective qu'a été conçu le programme EVAL, qui fonctionne par le biais du logiciel CLAN, associé à la base de données AphasiaBank. Ce programme a été mis au point par MacWhinney, Fromm, Holland et Forbes (MacWhinney et al., 2011), un groupe de chercheurs basés aux Etats-Unis impliqués dans la recherche clinique en aphasiologie et en psycholinguistique. Ils ont élaboré un protocole de recueil de données en langue anglaise permettant de recueillir des échantillons de discours oral représentatifs du comportement verbal et non verbal de locuteurs aphasiques (sous la forme vidéo et/ou audio), projet soutenu par le *National Health Institute*. Les données recueillies sont ensuite traitées par le programme EVAL qui intègre plusieurs mesures quantitatives et qualitatives dans une procédure automatisée. EVAL permet ainsi d'obtenir des données précises sur le profil discursif du sujet, en fournissant des indicateurs sur les aspects formels du discours oral continu. Les résultats obtenus sont ensuite

comparés aux données incluses dans la base de données, comprenant des corpus recueillis auprès d'un certain nombre de sujets contrôles et aphasiques, permettant de situer les performances du patient par rapport à d'autres sujets aphasiques ou non aphasiques.

Ce type de protocole d'évaluation informatisée formelle du discours oral n'existe pas à l'heure actuelle en orthophonie en France. EVAL est actuellement utilisé et développé pour différentes langues et dans différents pays, à travers la base de données collaborative AphasiaBank, à des fins de recherche clinique et fondamentale. Partant de cet intérêt pour l'évaluation du discours continu aphasique, nous nous sommes donc interrogées sur la possibilité d'adapter et d'utiliser ce type d'outil en langue française à des fins cliniques en orthophonie, dans un contexte où le domaine du numérique et du traitement automatique peut ouvrir de nouvelles perspectives pour l'évaluation et la prise en charge.

Nous aborderons dans un premier temps les aspects théoriques relatifs aux notions de discours, d'aphasie, d'évaluation orthophonique du discours pour envisager la place des outils d'évaluation informatisés en aphasiologie. Nous présenterons ensuite notre démarche méthodologique, et les résultats que nous avons obtenus à l'issue de chaque étape de cette démarche. Enfin, dans la dernière partie de ce mémoire, nous proposerons une réflexion critique sur les différents aspects de notre travail en abordant les intérêts, les limites et les perspectives de notre étude.

PARTIE 1 : PARTIE THEORIQUE

1. LE DISCOURS

Nous allons aborder la notion de discours selon deux dimensions : la dimension linguistique et la dimension psycholinguistique.

1.1. Le discours : approche linguistique

1.1.1. Le discours : généralités (d'après Charaudeau et Maingueneau, 2002)

La notion de discours recouvre de nombreuses acceptions en linguistique. Elle est classiquement définie à travers une série d'oppositions détaillées par Maingueneau (Charaudeau et Maingueneau, 2002). Nous allons reprendre quelques-unes de ces oppositions afin de mieux appréhender ce que recouvre la notion de discours.

• **Discours, phrase, langue et texte : distinctions et continuités**

Le discours peut tout d'abord être opposé à la phrase. En effet, il constitue une unité linguistique composée d'une succession de phrases.

Le discours s'oppose également à la langue. La langue, définie comme un système de valeurs virtuelles, se distingue du discours qui constitue une actualisation de la langue dans un contexte particulier. Cette conception est proche de l'opposition traditionnelle établie par De Saussure (1916) entre langue et parole (cité par Charaudeau et Maingueneau, 2002), la parole étant considérée comme l'actualisation concrète de la langue, elle-même considérée comme une entité abstraite. Le discours correspond alors à un usage de la langue. Certains auteurs orientent la notion de discours vers une dimension sociale, comme Gardiner (1989), pour qui le discours est « *l'utilisation, entre les hommes, de signes sonores articulés, pour communiquer leurs désirs et leurs opinions sur les choses* » (Gardiner, 1989, p. 24, cité par Charaudeau et Maingueneau, 2002, p. 186). D'autres auteurs orientent cette notion vers une dimension mentale. Ainsi, d'après Guillaume (1973) : « *Dans le discours (...) le physique qu'est la parole en soi se présente effectif, matérialisé, et donc, en ce qui le concerne, sorti de la condition psychique de départ. Au niveau du discours, la parole a pris corps, réalité : elle existe physiquement* » (Guillaume, 1973, p. 71, cité par Charaudeau et Maingueneau, 2002, p. 186). Benveniste (1966) rapproche la notion de discours et celle d'énonciation. Il conçoit en effet le discours comme « *la langue en tant qu'assumée par l'homme qui parle, et dans la condition d'intersubjectivité qui seule rend possible la communication linguistique* » (Benveniste, 1966, p. 266, cité par Charaudeau et Maingueneau,

2002, p. 186).

Le discours englobe également la notion de texte, en ce sens que le discours inclut un texte dans un contexte, c'est-à-dire dans des conditions de production et de réception particulières.

- **Le discours qualifié selon un type pour une communauté de locuteurs donnée**

Le discours s'oppose également à une autre acception de la langue, celle-ci pouvant être considérée comme un système partagé par l'ensemble des membres d'une communauté linguistique et le discours comme un usage restreint de ce système. Il peut s'agir d'un positionnement, on parle alors de « discours communiste » ou de « discours surréaliste ». Il peut également s'agir d'un type particulier de discours (« discours journalistique », « discours administratif »,...), des productions verbales spécifiques d'une catégorie de locuteurs (« discours des infirmières », « discours des mères de familles »...) ou d'une fonction du langage (« discours polémique », « discours prescriptif »,...) (Charaudeau et Maingueneau, 2002).

- **Caractéristiques fondamentales du discours**

Le développement de la linguistique du discours s'est effectué parallèlement à l'émergence d'une nouvelle conception du langage. Cette nouvelle façon d'appréhender le langage repose sur un certain nombre d'idées directrices qui nous permettent de mieux cerner la notion très complexe qu'est le discours :

Le discours suppose une organisation transphrastique. Même si le discours est composé d'une phrase unique, il mobilise des structures d'un autre ordre que les structures et les unités de base de la phrase (nom, verbe, complément,...). Le discours est donc une unité transphrastique soumise à des règles d'organisation en vigueur dans une communauté donnée. Ces règles sont propres aux différents genres de discours, chaque genre de discours étant soumis à des règles particulières (Charaudeau et Maingueneau, 2002).

Le discours est orienté. Le discours est une parole orientée, d'une part car il est conçu par le locuteur en fonction d'une intention particulière, d'autre part parce qu'il se développe dans le temps. Le discours se construit selon une finalité, il vise à aller dans une direction déterminée à l'avance. Mais cela n'empêche pas qu'il puisse être l'objet de digressions ou de changements de direction. Dans le discours, le locuteur exerce un véritable « guidage » sur ses propos. Ce « guidage » peut être très variable selon les conditions de l'émission du discours, selon si le locuteur est le seul énonciateur dans l'échange (monologue, écriture,...) ou si son discours peut être dévié par les réactions de son interlocuteur. Dans le cadre des interactions orales, le locuteur doit exercer un contrôle pour maintenir la cohérence de son discours en tenant compte des réactions de son interlocuteur (Charaudeau et Maingueneau, 2002).

Le discours est une forme d'action. La notion d'actes de langage développée par Austin (1962) puis Searle (1969) a posé l'idée que toute énonciation constitue un acte (comme promettre, suggérer, interroger, inviter, condamner,...) visant à modifier une situation et à agir sur une situation donnée ou sur autrui par le langage (Charaudeau et Maingueneau, 2002).

Le discours est interactif. L'interactivité du discours, évidente en conversation, l'est beaucoup moins dans le cas de discours écrit (interaction en différé) ou de certaines formes de discours oraux telle que le discours d'un conférencier (monologue) ou d'un présentateur de télévision (à distance). Cependant, on peut considérer que toute énonciation, même produite sans la présence *in situ* dans le même contexte d'un destinataire, est fondamentalement interactive. En effet, une énonciation suppose toujours un échange avec d'autres locuteurs, réels ou virtuels, à laquelle elle s'adresse et par rapport à laquelle le locuteur construit son discours (Charaudeau et Maingueneau, 2002).

Le discours est contextualisé. Le discours n'intervient pas dans un contexte duquel il pourrait être dissocié mais il est nécessairement contextualisé, il prend son sens en contexte. Il contribue également à définir le contexte dans lequel il s'exerce. Le domaine de la pragmatique s'intéresse en particulier à tout ce qui détermine la formulation ou l'interprétation d'un message, en lien avec le contexte où il est produit ou reçu (Charaudeau et Maingueneau, 2002).

Le discours est pris en charge. Le discours existe par rapport à celui qui le produit. Il est « *rapporté à une instance qui à la fois se pose comme source des repérages personnels, temporels, spatiaux et indique quelle attitude il adopte à l'égard de ce qu'il dit et de son interlocuteur* » (Charaudeau et Maingueneau, 2002, p. 189). La subjectivité du locuteur se retrouve dans son discours par le processus de modalisation. Le locuteur peut ainsi se positionner dans son propre discours, par exemple en commentant sa propre parole ou en précisant son degré d'adhésion à ses propos. Chez les patients aphasiques, cette fonction de modalisation peut être préservée de façon spécifique. Certains patients peuvent juger de la correction ou des déviations de leur production de façon explicite (« *je ne sais plus* », « *ah non, c'est pas ça* », « *c'est ça, oui je pense* »,...) (Nespoulous, 2010).

Le discours est régi par des normes. En tant que conduite sociale, le discours est soumis à des normes sociales, mais il est également soumis à des normes spécifiques, propres à chaque acte de langage et à chaque genre de discours (Charaudeau et Maingueneau, 2002).

Le discours est pris dans un interdiscours. Le discours n'a de sens qu'au sein d'un univers d'autres discours auquel il se réfère ou se positionne. On ne peut interpréter un discours sans le mettre en relation avec un ensemble d'autres discours. Le seul fait de parler de genres de discours suppose que chaque discours est pris dans des relations interdiscursives et qu'il n'existe que par rapport à d'autres (Charaudeau et Maingueneau, 2002).

Ainsi défini, et bien que certains linguistes postulent l'existence d'une « linguistique du discours » qui s'opposerait à une « linguistique de la langue », « *le discours ne délimite pas un*

domaine qui puisse être étudié par une discipline consistante. C'est davantage une manière d'appréhender le langage. » (Charaudeau et Maingueneau, 2002, p. 190).

Il est important d'avoir à l'esprit que le discours est une notion complexe qui a beaucoup été étudiée en sciences du langage, notamment en linguistique fondamentale. L'approche linguistique évoquée ici nous permet de cadrer notre objet d'étude en définissant le niveau d'organisation de la langue sur lequel porte l'outil d'évaluation proposé dans le cadre de ce mémoire.

1.1.2. Les différents modes d'organisation du discours : prolongement pour la clinique

Face à la pluralité des productions discursives empiriquement observées, il est apparu indispensable de rendre compte de l'existence de différents types de discours.

Malgré la diversité des productions observables, les sciences du langage ont cherché à extraire des principes stables et récurrents permettant d'identifier des modes d'organisation spécifiques du discours. Un nombre restreint de catégories discursives a ainsi pu être mis en évidence. Ces catégories permettent de distinguer différents types de discours se caractérisant chacun par des modalités variables présidant à la combinaison des segments qu'ils contiennent.

Il existe donc différents modes d'organisation du discours et ceux-ci mettent en jeu des processus cognitifs et langagiers spécifiques.

La notion de mode d'organisation du discours est définie par Charaudeau (1992) comme « *l'ensemble des procédés de mise en scène de l'acte de communication qui correspondent à certaines finalités (décrire, raconter, argumenter...)* » (Charaudeau, 1992, p. 635, cité par Charaudeau et Maingueneau, 2002, p. 386).

Parmi la multiplicité des propositions de classement des types de discours qui ont été faites, nous avons choisi de retenir celle de Charaudeau car cette classification nous semble pertinente par rapport aux épreuves discursives généralement proposées en évaluation clinique orthophonique. Il distingue quatre modes d'organisation du discours : le mode énonciatif, le mode descriptif, le mode narratif et le mode argumentatif. Le locuteur emploie l'un de ces modes en fonction de son intention communicative.

Charaudeau définit ainsi ces quatre modes d'organisation du discours :

« *Le mode **énonciatif** permet d'organiser la mise en scène des protagonistes de l'énonciation (Je, Tu, Il), leur identité, leurs relations, à l'aide de procédés de modalisation, également appelés « rôles énonciatifs » (..)*

*Le mode **descriptif** permet de faire exister les êtres du monde en les nommant et en les qualifiant de façon particulière.*

*Le mode **narratif** permet d'organiser la succession des actions et des événements dans lesquels ces êtres sont impliqués.*

*Le mode **argumentatif**, enfin, permet d'organiser les rapports de causalité qui s'instaurent*

entre ces actions, à l'aide de divers procédés portant sur l'enchaînement et la valeur des arguments. » (Charaudeau et Maingueneau, 2002, p. 387).

À ces quatre modes d'organisation du discours définis par Charaudeau, nous souhaiterions ajouter le discours procédural. Le discours procédural consiste en une description d'actions et peut donc être considéré comme un sous-type du discours descriptif.

Le protocole d'évaluation du discours que nous avons adapté dans le cadre de ce mémoire se compose de plusieurs tâches de discours destinées à évaluer les différents modes d'organisation du discours. Le mode énonciatif est évalué à travers une tâche de récit autobiographique, consistant à raconter l'histoire d'une maladie et un événement important. Les modes narratif et descriptif sont évalués à travers deux tâches de narration, à partir d'images et sans support imagé. Enfin, le discours procédural est évalué à travers l'évocation de la réalisation d'une tâche précise impliquant l'exécution d'une suite d'actions (la réalisation d'un sandwich jambon beurre). Ce protocole n'explore donc pas le mode argumentatif.

Dans toute évaluation des capacités langagières impliquant la mise en discours, les modes d'organisation sont plus ou moins dominants selon les épreuves proposées. Ils peuvent difficilement être totalement dissociés lors de l'analyse. Par exemple, dans une épreuve de récit d'après des images, le mode narratif (récit d'événements) et le mode descriptif (description à partir des éléments présents sur l'image) sont tous deux impliqués. Il est donc souvent difficile de dissocier ces deux modes d'organisation dans les tâches de production de discours classiques.

Ainsi l'analyse du discours est une entreprise complexe. En effet, la mise en discours intègre de nombreux éléments qui font du discours un objet complexe à constituer et à analyser, en particulier en clinique orthophonique où les perturbations langagières impactent la mise en discours.

1.1.3. Notion de compétence discursive et enjeu pour la clinique

La mise en discours nécessite la mise en jeu d'un certain nombre de compétences, telles que la compétence discursive. Celle-ci renvoie à la faculté de maîtrise des règles d'usage de la langue dans diverses situations. Elle se distingue de la compétence linguistique, de la compétence encyclopédique et de la compétence logique.

Pour communiquer, la compétence linguistique seule n'est pas suffisante. D'après Charaudeau (Charaudeau et Maingueneau, 2002), l'acte de communication requiert trois types de compétences : la compétence situationnelle, la compétence discursive et la compétence sémiolinguistique.

La compétence situationnelle renvoie à l'aptitude du sujet à « *construire son discours en fonction de l'identité des partenaires de l'échange, de la finalité de l'échange, du propos en jeu et*

des circonstances matérielles de l'échange » (Charaudeau et Maingueneau, 2002, p. 113-114).

La compétence discursive désigne l'aptitude à reconnaître et manipuler « *les procédés de mise en scène discursive qui feront écho aux contraintes situationnelles* » (Charaudeau et Maingueneau, 2002, p. 114) et les savoirs de connaissance et de croyance supposés partagés entre les interlocuteurs et témoignant d'un certain positionnement. Il faut également s'ajuster aux contextes sociaux dans lesquels se déroule l'acte de communication. Elle recouvre ainsi les compétences communicatives et pragmatiques.

La compétence pragmatique concerne les principes plus généraux de l'échange verbal et nécessite la maîtrise des principes de l'activité discursive, notamment les maximes conversationnelles, qui seront définies dans le paragraphe suivant.

Enfin, la compétence sémio-linguistique désigne l'aptitude du locuteur à reconnaître et manipuler « *les formes des signes, leurs règles de combinaison et leur sens* » (Charaudeau et Maingueneau, 2002, p. 113).

La compétence discursive nécessite donc la mise en jeu de multiples compétences (pragmatique, situationnelle...) et est étroitement liée à la mise en jeu de la compétence linguistique (ou sémio-linguistique).

L'outil d'évaluation EVAL, présenté dans la suite de ce mémoire, ne vise pas à évaluer la compétence discursive dans son ensemble. Il propose une évaluation des aspects formels du discours, et donc des performances du locuteur en production orale lors de la mise en œuvre de sa compétence sémio-linguistique.

Cette analyse formelle s'intégrera dans l'analyse plus globale des autres compétences impliquées dans la compétence discursive, effectuée qualitativement par le clinicien.

1.1.4. Les règles du discours : apport de la pragmatique

Afin d'assurer l'efficacité de l'échange entre les interlocuteurs, certains linguistes ont mis en évidence la nécessité de respecter certaines règles. Les partenaires de l'échange se soumettent à un principe de coopération, que Grice (1979) a défini ainsi : « *Que votre contribution conversationnelle corresponde à ce qui est exigé de vous par le but ou la direction acceptés de l'échange parlé dans lequel vous êtes engagé* ». (Grice, 1979, cité par Charaudeau et Maingueneau, 2002, p. 368). D'après Grice (1979), ce principe nécessite le respect de règles spécifiques : de qualité, de quantité, de relation et de modalité.

Les règles de qualité consistent à donner à l'interlocuteur des informations que l'on considère comme vraies ou avérées.

Les règles de quantité nécessitent que le locuteur fournisse la quantité d'informations

suffisantes à son interlocuteur. Il ne doit en fournir ni trop ni trop peu.

La règle de relation désigne le fait que le locuteur doit s'exprimer en cohérence avec le thème du discours.

Les règles de modalité consistent à s'exprimer de façon claire et non ambiguë.

Les maximes conversationnelles décrites par Grice (1979) peuvent être transgressées et elles peuvent entrer en conflit les unes avec les autres. Les maximes de qualité et de quantité peuvent faire peser une double contrainte, nécessitant d'arbitrer entre le respect de l'une ou de l'autre règle : par exemple lorsqu'un locuteur hésite entre donner une information trop vague, qui violerait la maxime de quantité, et une information plus précise mais incertaine, qui pourrait violer la maxime de qualité.

D'autres linguistes ont postulé l'existence de règles similaires comme les lois de discours de Ducrot (1972) ou les postulats de conversation de Gordon et Lakoff (1973) (cités par Charaudeau et Maingueneau, 2002, p. 369).

1.1.5. Les particularités du discours oral

Afin de mieux cerner l'objet de notre étude, nous allons à présent aborder les spécificités du discours oral.

1.1.5.1. Discours oral et discours écrit

• **Caractéristiques de l'oral par rapport à l'écrit**

La distinction entre oral et écrit repose sur une série d'oppositions classiques, dont les caractéristiques sont relevées par Charaudeau et Maingueneau (Charaudeau et Maingueneau, 2002, p. 202) et Blanche-Benveniste et Bilger, 1999 (Blanche-Benveniste et Bilger, 1999).

La distinction première entre oral et écrit repose sur le canal utilisé pour transmettre le message verbal, chacun impliquant une relation particulière au contexte d'énonciation. Le canal oral est considéré comme étroitement dépendant du contexte dans lequel il s'inscrit, alors que le canal graphique serait indépendant du contexte d'énonciation. De cette opposition découle un certain nombre de conséquences, qui nous permettent de mettre en évidence les particularités de l'oral. La première est que le canal écrit permettrait de stocker les informations et de les manipuler dans le temps et dans l'espace, ce que ne permettrait pas le canal oral. L'oral est éphémère, sans cesse changeant, ce qui rend moins aisée son analyse. Cependant le développement des médias audiovisuels et des enregistrements sonores permet désormais de stocker, conserver et analyser les productions orales en dehors de leur contexte de production, de la même façon que les productions écrites. La deuxième conséquence est que le message verbal ne s'élabore pas de la même façon. En effet, dans la production orale, le locuteur ne peut appréhender la globalité de son message, celui-ci se construit au fur et à mesure. De plus, en situation orale, le locuteur peut être

interrompu par ses interlocuteurs, ce qui peut perturber le déroulement du message.

À travers cette opposition classique, le terme d'« oral », tel qu'il est aujourd'hui couramment utilisé, renvoie implicitement à un usage particulier et restreint de la langue, considéré comme familier et n'obéissant pas aux règles conventionnelles de l'usage de la langue.

Il renvoie à une conception de la langue parlée qui serait « naturelle » et qu'on pourrait opposer à une langue écrite soumise à des règles spécifiques, résultant d'un apprentissage académique (Blanche-Benveniste et Bilger, 1999). Ceci met bien en évidence la difficulté d'appréhender l'oral par rapport à l'écrit à la fois pour son analyse et pour son évaluation. On comprend ainsi l'intérêt de développer des outils permettant de stocker les productions orales.

- **Les particularités du discours oral dans la normalité** (d'après Blanche-Benveniste et Bilger, 1999)

De cette opposition classique établie entre oral et écrit ont émergé différentes caractéristiques attribuées à la langue orale.

L'oral est considéré comme un lieu de manque car il se soustrairait aux règles conventionnelles du langage, auxquelles serait soumis l'écrit. L'oral dépendrait étroitement du contexte, de ce fait le recours à certaines formes grammaticales et structures syntaxiques complexes serait moins nécessaire. Ainsi Blanche-Benveniste et Bilger (1999) montrent que la relation entre les pronoms et leurs référents n'a pas besoin d'être clairement marquée à l'oral comme elle devrait l'être à l'écrit, l'ancrage dans la situation permettant de résoudre ces ambiguïtés. Certaines tournures syntaxiques seraient également rendues inutiles du fait de la gestuelle et des comportements paraverbaux accompagnant l'expression orale. En outre, les groupes rythmiques délimités par la prosodie permettent certaines fois de remplacer la phrase syntaxiquement construite de l'écrit. Les auteurs citent pour exemple la phrase orale « *il faisait un pas, il se faisait tuer* », montrant que le rapport d'hypothèse marqué par l'intonation à l'oral nécessiterait à l'écrit une phrase syntaxiquement plus complexe telle que « *s'il avait fait un seul pas, il se serait fait tuer* » (Blanche-Benveniste et Bilger, 1999, p. 3). Cependant, on retrouve également à l'oral des subordinations avec des marqueurs traditionnels et les productions orales présentent sous certains aspects une complexité particulière par rapport à l'écrit. En effet, la présence de nombreuses parenthèses et commentaires dans le discours oral, interrompant le déroulement de la construction verbale, a tendance à complexifier la construction du message. Enfin, l'oral contiendrait de nombreux énoncés laissés inachevés, notamment dans les situations de conversation. Cependant, l'analyse des corpus de langue parlée révèle que ces énoncés, qui semblent a priori être laissés inachevés, s'intègrent souvent dans une continuité syntaxique se déroulant sur une séquence plus longue du discours.

L'oral est donc souvent désigné comme un lieu de manque. Mais, parallèlement, il est

également désigné comme un lieu de surplus. En effet, l'oral apparaît comme étant un lieu de répétition qu'on pourrait opposer à la qualité de concision de la langue écrite. À l'oral, les thèmes du discours ont ainsi tendance à être répétés de nombreuses fois sous des formes différentes. De plus, l'oral est considéré comme un lieu de répétitions de formes lexicales et de schémas syntaxiques identiques. Ces répétitions relèvent parfois de procédés d'insistance (ex : « *il fallait tout tout tout briquer à fond* ») (Blanche-Benveniste et Bilger, 1999, p. 5), d'autres fois d'hésitations involontaires liées à la construction du discours. Les répétitions de structures syntaxiques identiques dans plusieurs énoncés successifs peuvent aussi servir à produire un effet de contraste. (Blanche-Benveniste et Bilger, 1999). Elles citent pour exemple les énoncés suivants : « *moi je jouais dans un orchestre* » et « *elle elle chantait dans les cours* » (Blanche-Benveniste et Bilger, 1999, p. 5). Enfin, elles relèvent que les répétitions peuvent constituer un élément de rythme du discours. Certains locuteurs semblent ainsi s'appuyer sur un schéma rythmique en utilisant la répétition, la symétrie et la rupture. Les répétitions, fréquemment associées à l'oral, pourraient donc être interprétées comme des outils visant à favoriser une meilleure perception du discours.

Ces particularités, propres au discours oral, sont importantes à considérer dans l'approche clinique de l'évaluation du discours. En effet, ces particularités, décrites dans le discours oral non pathologique, peuvent également être retrouvées dans le langage pathologique. De ce fait, dans l'analyse du discours pathologique, il est parfois difficile de distinguer ce qui relève de la normalité et ce qui relève de la pathologie. En effet, un patient aphasique peut produire certaines déviations qu'on pourrait également retrouver dans le discours d'un sujet normal. Ainsi des erreurs telles qu'une confusion entre deux mots sémantiquement proches, l'omission d'un mot grammatical ou la présence de répétitions ou de reformulations, peuvent traduire, chez le sujet normal comme chez le sujet aphasique, des stratégies d'organisation et de planification du discours relevant de processus sous-jacents normaux et non pathologiques. Des disfluences orales, relevant pourtant de processus normaux, risquent donc à tort d'être interprétées comme des processus pathologiques, si elles sont observées chez un locuteur aphasique.

Cette difficulté à distinguer le caractère normal ou pathologique d'une même manifestation de surface met en avant la nécessité de disposer d'un référentiel de discours « typique », avec ses particularités et ses disfluences « normales », lorsqu'on aborde l'évaluation du discours pathologique, ce qui est rendu possible par le biais d'ÉVAL et la constitution d'une base de données représentative du comportement langagier « normal » ou « ordinaire ». Ceci peut permettre de mieux caractériser ce qui relève de la pathologie dans les manifestations de surface du discours oral aphasique.

Ce faisant, ÉVAL s'inscrit dans une base de données regroupant des échantillons de discours issus de sujets contrôles, non aphasiques, et de sujets aphasiques. Ceci permet d'établir des comparaisons et des distinctions entre les discours pathologiques et non pathologiques, et de

mieux caractériser les déviations linguistiques pathologiques, sur les plans qualitatifs et quantitatifs.

1.1.5.2. Les différents registres de langue

L'opposition classiquement établie entre langue écrite et langue orale est actuellement remise en cause par l'analyse des données fournies par les corpus de français parlé. Ces données ont permis de mettre en évidence que la langue se manifestait sous différents registres, et ce, chez un même locuteur, indépendamment du canal utilisé, oral ou écrit.

Blanche-Benveniste et Bilger (1999) ont ainsi montré l'existence de deux registres de langue, qu'elles qualifient de « spontané » et « soutenu ». Elles ont également présenté un certain nombre de phénomènes grammaticaux sensibles à la différence de registre, et ceci quel que soit le canal utilisé (Blanche-Benveniste et Bilger, 1999). Le langage « soutenu » se caractérise ainsi par la plus grande occurrence de certaines formes lexicales et grammaticales, telles que les conjonctions *car* et *lorsque*, le pronom *nous*, et le *ne* de négation, ainsi que l'utilisation d'un même sujet pour plusieurs formes verbales. À l'inverse, dans un registre de langage « spontané », le locuteur utilisera préférentiellement leurs équivalents moins soutenus : *parce que*, *quand*, le pronom *on*, et l'absence du *ne* de négation. Elles ajoutent que l'utilisation du passé simple, caractéristique du registre « soutenu » et classiquement associé à l'écrit, est fréquemment retrouvé à l'oral, notamment lorsque les locuteurs sont invités à raconter un événement de leur vie.

Ces différents registres peuvent être utilisés à tour de rôle par un même locuteur au sein d'une même prise de parole. Cette variabilité des registres au sein de la langue orale (et de la langue écrite) révèle qu'ils ne sont pas spécifiques au canal oral ou écrit et qu'on ne peut pas réduire l'oral au langage « spontané » et l'écrit au langage « soutenu ».

Après avoir présenté les caractéristiques du discours continu oral d'après des critères et des modèles linguistiques, nous allons à présent développer les aspects psycholinguistiques en jeu dans la mise en discours.

1.2. Le discours oral : approche psycholinguistique (d'après Fromkin et Ratner, 1998)

Si les aspects physiologiques, articulatoires et acoustiques impliqués dans la production du langage oral sont bien connus, les processus permettant au locuteur de traduire une pensée en forme linguistique, et la façon dont les mots et les phrases sont sélectionnés, construits et agencés, sont encore peu compris.

Fromkin et Ratner (1998) rassemblent les fondamentaux concernant les processus de production du discours oral (Fromkin et Ratner, 1998). Nous allons présenter les principaux apports fournis par ces données dans la compréhension de ces processus.

Après avoir brièvement rappelé les différentes méthodes permettant de construire les modèles de production du discours oral, nous aborderons les unités de production et les fondamentaux concernant les processus de production de ce discours. Enfin, nous présenterons les principaux modèles permettant de rendre compte de ces données.

1.2.1. Les phénomènes de l'oral : une fenêtre sur les processus psycholinguistiques sous-jacents

1.2.1.1. Généralités

Afin de comprendre les processus impliqués dans la production du discours oral et de construire des modèles de production, les chercheurs se sont historiquement basés sur deux types de données : les phénomènes d'erreurs et de disfluences.

Le discours est généralement produit à un niveau de 150 mots par minute (Maclay et Osgood, 1959, cités par Fromkin et Ratner, 1998). Cette rapidité de production explique que des erreurs surviennent fréquemment dans le discours. Tout le monde, en tant que locuteur ou interlocuteur, a fait l'expérience de moments où les énoncés produits ne correspondaient pas à ceux qui étaient voulus. Ces erreurs révèlent que le discours est composé d'unités distinctes, de différents types et de différentes tailles. Elles montrent également que le discours n'est pas simplement produit par l'émission d'un son, d'une syllabe, d'un mot à la fois, mais que ces éléments sont ordonnés de manière sérielle et assemblés au moyen d'étapes précises (Fromkin et Ratner, 1998).

Le discours oral est également caractérisé par la présence de nombreuses disfluences, telles que des hésitations, des répétitions, des faux départs et des mots de remplissage. Goldman-Eisler (1968) suggère que les hésitations se produisent tous les 7-8 mots lors des conversations naturelles (Goldman-Eisler, 1968, cité par Fromkin et Ratner, 1998).

1.2.1.2. Apports de l'étude des erreurs de performance : de la normalité à la pathologie

L'analyse des erreurs et des disfluences produites dans le discours normal fournit des indications indirectes sur les unités, les étapes et les processus cognitifs impliqués dans les productions langagières. Les chercheurs analysent donc ce type de données afin de mettre en évidence les différents processus à l'œuvre dans la production du discours oral, et de nombreux modèles de production ont été construits à partir de l'analyse de ces données (Fromkin et Ratner, 1998).

1.2.1.2.1. Les unités de production du discours oral

La production orale repose sur l'utilisation d'un nombre limité d'items stockés au niveau mental. En effet, les capacités de stockage du cerveau ne sont pas illimitées et ne permettent pas de stocker un nombre infini d'énoncés. Aussi, lors de la production orale, nous puisons dans un

stock limité d'items que nous devons ensuite réarranger et agencer afin de produire l'énoncé voulu.

Le flux discursif s'écoule de façon continue et linéaire dans une temporalité particulièrement rapide et rythmée, il est donc difficile d'isoler les unités distinctes impliquées dans sa production. Cependant, les erreurs qui se produisent dans le discours permettent de mettre en évidence l'existence de certaines unités distinctes, qui ne sont pas identifiables lorsque le discours est produit sans erreur. En effet, si des segments sont affectés par des erreurs, qu'ils sont produits de façon incorrecte ou à une mauvaise position dans l'énoncé, on peut alors les identifier comme des unités distinctes impliquées dans la production du langage.

L'analyse des erreurs révèle que les erreurs langagières peuvent affecter des unités de taille et de catégories différentes. Ainsi, différents types d'unités distinctes seraient impliquées dans la production du discours oral : les traits phonétiques, les segments phonémiques, les syllabes, les accents, les morphèmes et les mots, les syntagmes et propositions syntaxiques (Fromkin et Ratner, 1998).

Les erreurs portant sur les traits distinctifs de la réalisation des phonèmes ont permis de mettre en évidence la place des traits phonétiques comme unités distinctes de production orale.

Les segments phonémiques peuvent également être affectés d'un certain nombre d'erreurs telles que les ajouts, les suppressions, les substitutions, les inversions et les assimilations. Ils constituent donc également des unités distinctes de production.

Les syllabes qui n'ont pas le statut de morphèmes peuvent également être l'objet d'erreurs dans le discours.

Le fait que les accents toniques puissent être également affectés d'erreurs et ainsi être considérés comme des unités indépendantes de production du discours a été montré par certains chercheurs mais fait encore l'objet de controverses (Fromkin et Ratner, 1998).

L'analyse des erreurs et des disfluences a également montré le statut d'unités de base des mots, des morphèmes, et des syntagmes composant la phrase dans la production du discours.

Les mots constituent des unités distinctes impliquées dans la production orale comme en témoignent les erreurs de sélection et de placement de mot, ainsi que les phénomènes de recherche lexicale. Ces perturbations donnent des indications sur l'organisation du lexique mental et les mécanismes de récupération lexicale. Les chercheurs ont notamment mis en évidence que les pauses étaient plus fréquentes et plus longues avant les mots signifiants (noms, verbes,...) qu'avant les mots fonctions (articles,...). Les hésitations se produisent également plus souvent avant les mots de basse fréquence. La longueur et la fréquence des pauses semblent révéler que le processus d'accès lexical est plus difficile pour les mots signifiants que pour les mots fonction, et plus difficile pour les mots de basse fréquence que pour les mots de haute fréquence (Fromkin et

Ratner, 1998).

Les morphèmes lexicaux (lexèmes) et les morphèmes grammaticaux (dérivationnels et flexionnels) peuvent être affectés d'erreurs et donc aussi être considérés comme des unités de traitement indépendantes. De plus, la production de formes dérivées possibles mais non admises dans la langue montre que, dans certains cas, les mots complexes sont formés pendant la production du discours en plus d'être sélectionnés à partir du lexique mental (Fromkin et Ratner, 1998).

Des erreurs ont également été observées au sein d'unités linguistiques plus larges comme les syntagmes et les propositions. Les phrases ou propositions sont composées de syntagmes syntaxiques ayant une structure hiérarchique, des syntagmes plus larges incluant des syntagmes plus petits. Ainsi, les syntagmes nominaux, verbaux ou prépositionnels, ont été identifiés comme des unités de traitement impliquées dans les processus de production. Le statut d'unités de ces structures est également suggéré par le fait que, dans la majeure partie des cas, lorsqu'un locuteur repère une erreur et procède à sa correction, la correction se produit généralement au début du composant syntaxique dans lequel l'erreur s'est produite plutôt qu'à l'endroit précis où s'est produite l'erreur. De même, quand un locuteur procède à une répétition, il revient le plus souvent au point de rupture entre les composants de la phrase pour commencer la répétition. Par exemple, si un locuteur procède à une répétition, il aura tendance à la formuler ainsi : « *Speech errors provide evidence (pause) provide evidence for ...* » plutôt que « *Speech errors provide evidence (pause) evidence for...* ». (Maclay et Osgood, 1959, cités par Fromkin et Ratner, 1998, p.321).

L'organisation du discours en unités propositionnelles a également été mise en évidence par Garrett (1975, 1976). Il a relevé que 85 % des interventions de mots se produisaient au sein d'une même proposition. Le fait que 15 % des erreurs impliquent des mots de différentes propositions révèle également que le locuteur planifie et construit les différentes structures syntaxiques à l'avance (Garrett, 1975, 1976, cité par Fromkin et Ratner, 1998).

L'analyse des pauses accrédite également l'idée selon laquelle le locuteur procède à un encodage au moment des propositions. Les pauses surviennent ainsi plus fréquemment aux limites des propositions ou autres points de rupture importants. Ce temps de pause est utilisé pour encoder la proposition suivante. Plus le discours est syntaxiquement complexe, plus les hésitations et les pauses pleines sont nombreuses. Le locuteur a alors besoin de plus de temps pour planifier son discours (Fromkin et Ratner, 1998).

Les unités utilisées dans la production du discours doivent être récupérées, sélectionnées, manipulées et organisées en structures hiérarchiques. Un certain nombre de principes à l'œuvre dans les processus de production du discours ont été mis en évidence. Nous allons exposer ces principes et montrerons ensuite comment les différents modèles de production ont justifié ces différents principes (Fromkin et Ratner, 1998).

1.2.1.2.2. Les principes fondamentaux à l'œuvre dans les processus de production du discours oral

1.2.1.2.2.1. Le discours est organisé à l'avance

Les erreurs d'anticipation ou d'échange (phonologiques, lexicales ou syntaxiques) montrent que la production du discours ne s'accomplit pas par la production d'une unité à la fois. Pour produire un discours, le locuteur doit avoir accès à une représentation qui inclut plus qu'un mot, et même plus qu'une proposition. Le discours est donc planifié à l'avance et les unités qui le composent sont donc impliquées dans toute une série de processus permettant d'aboutir à la production effective du message. Lashley (1951) a ainsi proposé : « *une série de hiérarchies d'organisation : l'ordre des mouvements vocaux dans la prononciation des mots, l'ordre des mots dans la phrase, l'ordre des phrases dans le paragraphe.* » (Fournié, 1887, cité par Lashley, 1951, cité par Fromkin et Ratner, 1998, p. 323). Cette hiérarchie suppose donc l'existence de différents niveaux de représentation et de différentes étapes dans le processus de production du discours. Pour rendre compte de la production orale, les modèles de production doivent donc proposer toutes les étapes nécessaires à l'élaboration du message, montrer quelles erreurs peuvent se produire à quelles étapes, et prédire la forme de représentation des énoncés aux différents niveaux (Fromkin et Ratner, 1998).

1.2.1.2.2.2. Le lexique est organisé sémantiquement et phonologiquement

Les erreurs de sélection lexicale observées permettent d'appréhender la nature de l'organisation lexicale. Dans la plupart des substitutions de mots, on retrouve certaines similarités sémantiques et/ou phonologiques entre le mot cible et le mot produit par erreur : le même début, la même structure morphologique, la même forme phonologique ou des traits sémantiques communs. Ces erreurs de sélection lexicale interviennent à une étape ultérieure à l'étape d'attribution de la catégorie syntaxique du mot cible. En effet, les mots sont substitués par des mots appartenant à la même catégorie grammaticale. Les noms sont substitués par des noms, les verbes par des verbes,... Ainsi, dans l'exemple suivant, la substitution implique des critères à la fois sémantiques et phonologiques et le mot substitué appartient à la même catégorie grammaticale que le mot cible : « *Arrested and prosecuted* » → « *Arrested and persecuted* » (Fromkin et Ratner, 1998, p. 323).

Toutes les erreurs lexicales ne s'expliquent cependant pas par des facteurs linguistiques, une partie d'entre elles relève de l'interférence de facteurs non linguistiques ou de lapsus inconscients (Fromkin et Ratner, 1998).

1.2.1.2.2.3. Les mots morphologiquement complexes sont assemblés

Lors de certaines erreurs, on observe des agencements morphologiques possibles mais n'existant pas dans la langue (exemple : « *a New Yorkan* », au lieu de « *a New Yorker* », sur le modèle de America/American) (Fromkin et Ratner, 1998, p. 324). On peut penser que ces erreurs dérivationnelles se produisent dans le lexique avant l'étape d'insertion des items lexicaux dans l'énoncé. Les règles morphologiques permettant la formation des mots complexes sont donc activées dans les processus de production du discours. Les erreurs affectant le positionnement des affixes suggèrent que les morphèmes flexionnels et dérivationnels sont stockés et traités différemment des mots et des radicaux dans les processus de production. Il arrive par exemple que les mots et les radicaux soient intervertis, sans que leurs morphèmes grammaticaux le soient. Par exemple dans la substitution suivante : « *I don't know that I'd know one if I heard it* » → « *I don't know that I'd hear one if I knew it.* » (Fromkin et Ratner, 1998, p. 325), les radicaux des verbes ont été intervertis mais leurs flexions verbales n'ont pas été affectées par le déplacement des radicaux. Lorsque cela se produit, les règles phonologiques et morphologiques sont appliquées aux morphèmes grammaticaux. Ainsi, dans l'exemple précédent, bien que les radicaux aient été intervertis, la flexion verbale a été agencée en tenant compte des règles d'accord régulier ou irrégulier du radical du verbe à laquelle elle a été finalement rattachée. Ceci révèle que les morphèmes grammaticaux sont traités de façon indépendante dans les processus de production et que les mots morphologiquement complexes ne sont pas « pré-agencés » dans le lexique mental mais qu'ils sont assemblés au cours du processus de production du discours (Fromkin et Ratner, 1998).

1.2.1.2.2.4. Les erreurs affectent différemment les affixes, les mots fonction, les éléments mineurs et les éléments majeurs de la phrase

Garrett (1976, 1984, 1988, cité par Fromkin et Ratner, 1998) a montré que les affixes, les mots fonction et les éléments mineurs de la phrase (adverbes, déterminants,...) d'une part, et les éléments majeurs signifiants (noms, verbes, adjectifs) d'autre part n'étaient pas affectés par le même type d'erreurs. Les affixes, les mots fonction et les éléments mineurs sont fréquemment impliqués dans des erreurs de déplacement, ce qui n'est pas le cas des éléments majeurs, qui sont plutôt affectés par des erreurs de substitution.

Exemple 1 : « *I frankly admit to being subjective* » → « *I admit to frankly being subjective* » (Fromkin et Ratner, 1998, p. 326).

Exemple 2 : « *Too many irons in the fire* » → « *Too many irons in the smoke* » (Fromkin et Ratner, 1998, p. 323).

Ainsi dans l'exemple 1, l'adverbe « *frankly* » est affecté d'une erreur de déplacement, et dans l'exemple 2, le nom « *fire* » est affecté d'une erreur de substitution (sémantique).

1.2.1.2.2.5. Les erreurs de discours reflètent la connaissance de la règle

Lors de la production du discours, le locuteur fait appel à une grammaire mentale qui inclut les règles grammaticales et syntaxiques déterminant si les phrases sont grammaticalement et syntaxiquement correctes ou non. Cette grammaire mentale est sollicitée à l'étape de planification et de construction de la structure syntaxique du discours. Son existence est mise en évidence par un certain nombre d'erreurs. On retrouve ainsi des énoncés dans lesquels une règle régulière est appliquée à une forme irrégulière ou à une exception, des règles qui sont appliquées alors qu'elles ne devraient pas l'être ou inversement, comme dans l'exemple suivant : « *He swam in the pool* » → « *He swimmed in the pool* » (Fromkin et Ratner, 1998, p. 326).

Les erreurs ainsi décrites au niveau linguistique permettent donc de mieux comprendre les processus psycholinguistiques à l'œuvre dans la production du discours non pathologique. Cette démarche est également valable dans l'étude du discours pathologique. En effet, l'étude des erreurs que l'on rencontre dans le comportement aphasique (paraphasies, agrammatisme, dyssyntaxie, anomie,...) peut permettre de fournir de précieuses indications quant aux mécanismes sous-jacents perturbés pouvant expliquer les erreurs observables en surface. La connaissance de ces mécanismes permet au clinicien de mieux appréhender la problématique singulière de son patient et ainsi de proposer une prise en charge adaptée. Ainsi, l'analyse précise des erreurs observables dans le discours de la personne aphasique revêt un aspect fondamental dans la clinique orthophonique.

1.2.2. Les principaux modèles de production du discours oral

Pendant longtemps, les travaux des psycholinguistes se sont centrés sur l'étude du versant de la compréhension de la chaîne parlée, peu de travaux ont concerné la production du langage oral. Les premiers modèles s'intéressant à la production du langage oral se sont focalisés sur la production des sons du langage, et non du langage lui-même. Cependant, depuis les années 1970, l'émergence de l'analyse des données d'erreurs de discours a permis la constitution d'un certain nombre de modèles de production du langage oral (Fromkin et Ratner, 1998).

Il existe trois grands modèles de production du langage oral. Ces principaux modèles postulent l'existence de trois étapes principales de traitement : la préparation conceptuelle du message, représentant l'intention que veut exprimer le locuteur sous une forme non linguistique, la formulation du message préverbal en forme linguistique et la mise en œuvre des processus articulatoires. Les différents modèles s'opposent cependant sur l'organisation de ces différents niveaux et sur la nature sérielle ou interactive des relations entre les niveaux sémantiques et phonologiques.

Le modèle de Fromkin, publié en 1972, constitue l'un des premiers modèles visant à rendre compte des étapes et des représentations impliquées dans la production orale. Il s'agit d'un modèle sériel qui postule l'existence de six étapes au cours desquelles différentes représentations de l'énoncé sont constituées. Ce modèle repose sur un circuit descendant sans boucle de rétro-contrôle. Il constitue la première proposition de modélisation à partir des erreurs dans la production orale et a permis le développement de la recherche dans ce domaine.

Le modèle de Garrett, publié en 1975 et révisé à plusieurs reprises, repose sur un principe de sérialité stricte. Les processus de traitements ultérieurs ne peuvent entrer en jeu qu'à partir du moment où les processus de traitements précédents sont terminés. Ce principe semble difficilement compatible avec les contraintes temporelles qui s'exercent sur la production orale, dans laquelle le discours s'exprime de façon continue et fluide.

Dans le cadre de ce mémoire, nous nous appuyerons sur le modèle de Levelt. Il apporte en effet un prolongement intéressant au modèle de Garrett, en postulant que les traitements peuvent s'opérer de manière parallèle à toutes les étapes de la production du discours.

1.2.2.1. Le modèle de Levelt

Le modèle de Levelt, publié en 1989, est également un modèle sériel, mais il autorise le traitement simultané de différentes représentations du message (Levelt, 1999).

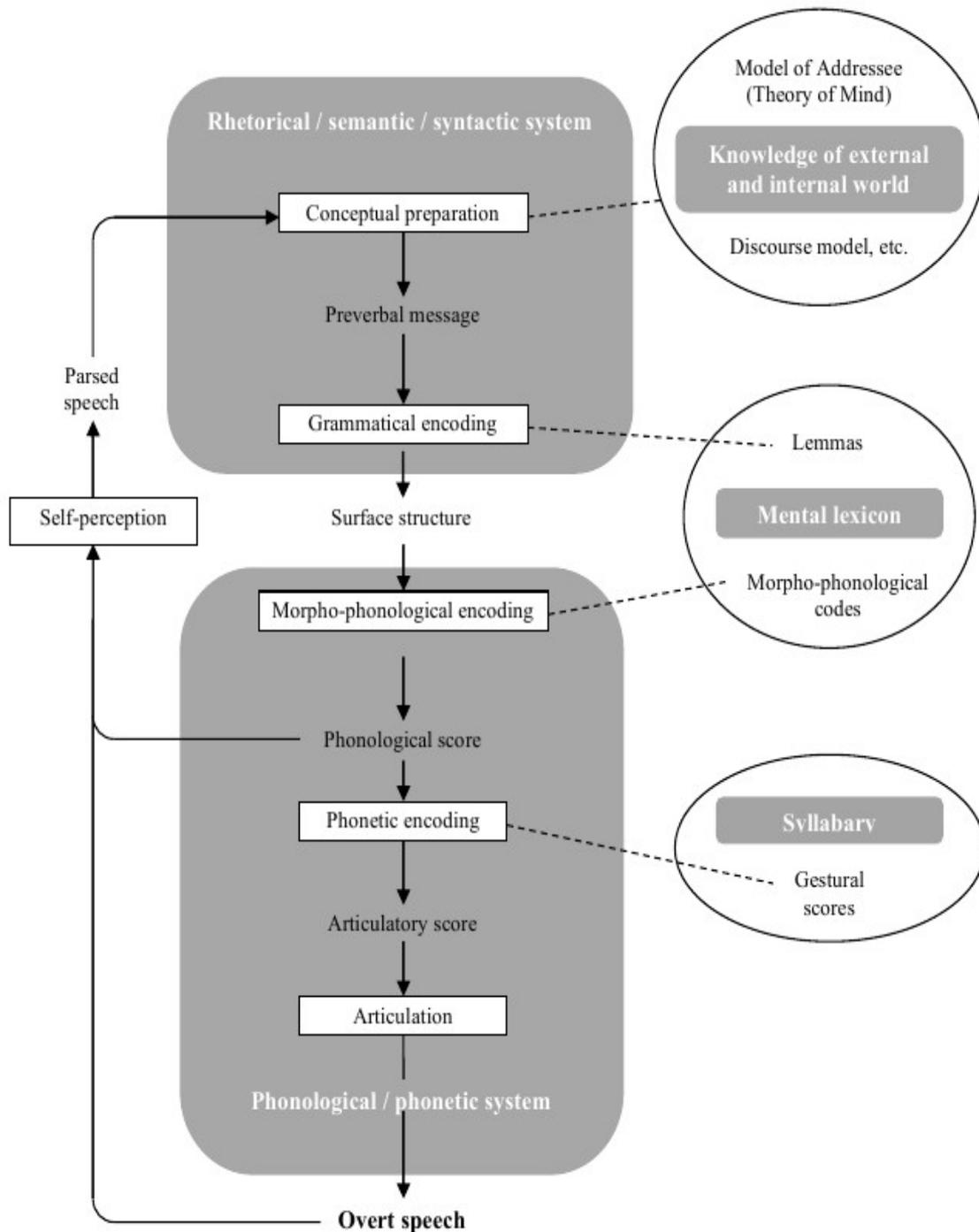


Figure n°1 : Modèle de Levelt (tiré de Levelt, 1999, p. 87)

La première étape de ce modèle consiste en une préparation conceptuelle du message (*conceptual preparation*). Cette étape permet de générer une première représentation, celle du « message préverbal » (*preverbal message*). Cette représentation du message contient l'intention

communicative du locuteur, sous une forme encore non linguistique. A cette étape interviennent différents processus (représentés dans la bulle en haut à droite dans le schéma), notamment la théorie de l'esprit (*theory of mind*), avec la prise en compte des connaissances de l'interlocuteur, les connaissances du monde interne et externe (*knowledge of external and internal world*) et un modèle de discours (narratif, descriptif,...) (*discourse model*). La représentation conceptuelle du message prend une forme propositionnelle qui s'établit en fonction de la perspective adoptée par le locuteur.

L'étape suivante est celle de la formulation du message, elle se divise en deux étapes : l'encodage grammatical (*grammatical encoding*) et l'encodage phonologique (*morpho-phonological encoding*).

L'encodage grammatical va donner naissance à une structure syntaxique de surface (*surface structure*) dans laquelle les unités sélectionnées seront positionnées. L'encodeur grammatical récupère les items lexicaux au sein du lexique mental (*mental lexicon*, représenté par une bulle à droite dans le schéma). D'après Levelt, les propriétés sémantiques et syntaxiques (*lemmas*) et l'information phonologique (*morpho-phonological codes*, correspondant aux lexèmes) des items lexicaux sont stockées et récupérées séparément. À cette étape, seuls les lemmes, propriétés sémantiques et syntaxiques des items lexicaux, sont récupérés au sein du lexique mental, leur forme phonologique n'est pas encore spécifiée. Les informations sémantiques et syntaxiques sont donc récupérées avant les informations phonologiques. Les propriétés syntaxiques récupérées permettent de générer les structures de phrases appropriées. Si un nom est sélectionné, les contours d'un groupe nominal seront générés, si c'est un verbe, un syntagme verbal sera généré. Les différents lemmes activés, correspondant aux différents concepts de la phrase à produire, sont ensuite assemblés entre eux en fonction de leurs propriétés syntaxiques. Ainsi, l'encodeur grammatical produit la structure de surface dans laquelle les lemmes sont ordonnés de façon linéaire, de gauche à droite.

Concernant la sélection des lemmes, Levelt a développé un modèle intitulé « Weaver » dans lequel il postule que le lexique mental est organisé sous forme de réseau. Lorsqu'un concept est activé, d'autres concepts sémantiquement reliés sont également activés, ce qui aboutit à l'activation de plusieurs lemmes. Le degré d'activation du lemme pertinent par rapport aux autres permet la sélection finale du lemme (Levelt, 1999, p. 95-99).

Après l'encodage grammatical intervient l'encodage morpho-phonologique. Les codes morpho-phonologiques sont récupérés au sein du lexique mental. Le plan phonologique de l'énoncé, qui inclut les structures intonatives, est alors généré.

Ensuite, l'encodage phonétique (*phonetic encoding*) permet la programmation articulaire et les commandes motrices sont envoyées au système neuro-musculaire afin d'obtenir la production effective de l'énoncé (*overt speech*) (Fromkin et Ratner, 1998).

Levelt postule l'existence d'un système de contrôle des erreurs au sein de son modèle de

production du discours. D'après lui, le locuteur exerce un contrôle à la fois sur les formes intermédiaires de ses énoncés au cours du processus de production et sur leurs résultats. Les nombreuses tentatives d'auto-correction constatées au cours de la production du discours oral s'expliqueraient par l'existence de ce mécanisme de rétro-contrôle (*self-perception et parsed speech*).

Levelt définit trois niveaux de contrôle exercés par le locuteur sur ses énoncés. Le premier niveau de contrôle s'exerce sur la planification du message. Lorsque le locuteur perçoit une erreur il peut réviser son message en cours de planification avant qu'il ne soit formulé. Le deuxième niveau de contrôle concerne le « discours intérieur ». Le locuteur peut corriger son message lorsque celui-ci est encore au niveau du « buffer articulatoire ». Tant que le discours n'a pas été articulé, ce mécanisme permet de prévenir l'apparition d'erreurs. Enfin, à un troisième niveau, le locuteur exerce un contrôle sur le discours effectivement produit. Généralement, dès que le locuteur détecte une erreur, il interrompt son discours et procède à une auto-correction. Il introduit parfois des mots indiquant son rejet de l'erreur produite (« *non* », « *désolé* »,...). Mais il arrive que la détection de l'erreur se fasse avec retard car le locuteur prête principalement attention à la planification du discours et non à son résultat. Dans ce cas, un ou plusieurs mots peuvent être produits après l'erreur avant que le locuteur ne perçoive son erreur. Le locuteur peut également devenir conscient de son erreur grâce à la réaction de son interlocuteur et les divers signaux d'incompréhension que celui-ci peut lui renvoyer (Levelt, 1993, p. 6-7). Il est intéressant de pouvoir aussi appréhender la possibilité de ce rétro-contrôle chez les locuteurs aphasiques. En effet, chez les sujets aphasiques, la correction des erreurs produites peut être altérée du fait d'une perturbation des processus dédiés au rétro-contrôle audio-phonatoire ou d'une perturbation des processus de base de décodage.

2. L'APHASIE

2.1. Définition de l'aphasie

Il existe de nombreuses définitions de l'aphasie. Parmi toutes ces définitions, nous pouvons retenir celle de Damasio (1991). D'après lui, l'aphasie représente « *la perturbation de la compréhension et de la formulation des messages verbaux qui résulte d'une affection nouvellement acquise du système nerveux central* » (Viader et al., 2010, p. 2). Cette définition a l'avantage d'énoncer les caractéristiques essentielles de l'aphasie permettant de la distinguer des autres pathologies langagières.

Le terme « nouvellement acquise » permet de distinguer l'aphasie des troubles du langage

d'origine congénitale ou développementale. La référence au « système nerveux central » la distingue d'une utilisation déviante du langage qui serait liée à un usage social particulier ou à une affection psychogène. En précisant que l'atteinte touche les « messages verbaux », Damasio (1991) établit également une distinction avec les troubles de la communication gestuelle ou émotionnelle. Enfin, en utilisant les termes de « compréhension des messages verbaux » et de « formulation des messages verbaux », il la distingue d'une part des troubles perceptifs visuels ou auditifs et d'autre part des troubles de la phonation et de l'articulation.

Damasio (1991) précise également dans sa définition que l'aphasie doit être distinguée des troubles langagiers survenant dans des états de confusion mentale ou d'altération de la conscience. (Damasio, 1991, cité par Viader et al., 2010).

2.2. Etiologie des aphasies

Les accidents vasculaires cérébraux (AVC) sont une cause fréquente d'aphasie. Il existe deux types principaux d'AVC, les AVC ischémiques, dus à l'occlusion d'une artère cérébrale qui provoque un arrêt du flux sanguin dans un territoire cérébral, et les AVC hémorragiques, provoqués par une rupture des vaisseaux sanguins entraînant une diffusion de sang dans le cerveau.

Il existe également d'autres causes pouvant entraîner l'apparition d'une aphasie : les tumeurs cérébrales, des causes inflammatoires et infectieuses (méningo-encéphalites herpétiques) ou les traumatismes crâniens. Une aphasie transitoire peut également être provoquée par la migraine et l'épilepsie.

2.3. Perturbations langagières orales observées

L'aphasie peut entraîner une atteinte du langage sur les versants expressifs et/ou réceptifs et concerner ses aspects oraux et/ou écrits. Nous allons présenter les principales perturbations langagières observées au niveau de l'expression orale. Nous n'aborderons pas les troubles de la compréhension et les troubles du langage écrit, qui ne concernent pas directement notre sujet de mémoire.

La description des perturbations langagières orales repose notamment sur le modèle linguistique des trois articulations du langage défini par Martinet en 1967 (Chomel-Guillaume et al., 2010). Ce modèle définit quatre types d'unités linguistiques et trois niveaux d'articulation.

Les unités de **troisième articulation** sont les traits. Ce sont les mouvements élémentaires de l'appareil bucco-phonatoire.

Le niveau de troisième articulation correspond au niveau phonétique, c'est-à-dire à la sélection d'un certain nombre de traits, leur combinaison selon les règles de convention phonétique afin de réaliser les phonèmes.

Les unités de **deuxième articulation** sont les phonèmes. Ce sont les plus petites unités de son. Elles existent en nombre stable dans la langue et peuvent être transcrites en alphabet phonétique international.

Le niveau de deuxième articulation correspond au niveau phonémique ou phonologique, c'est-à-dire à la sélection, la sériation et la combinaison de phonèmes suivant le système phonologique pour permettre la réalisation des monèmes.

Les unités de **première articulation** sont les monèmes. Ce sont les plus petites unités de sens. Elles sont composées d'un contenu sémantique, le signifié, et d'une expression phonique, le signifiant. Les monèmes regroupent les monèmes lexicaux, les affixes et les morphèmes grammaticaux.

Le niveau de première articulation correspond au niveau morpho-syntaxique. Il désigne la sélection, la sériation et la combinaison des monèmes suivant les règles syntaxiques afin de produire les syntagmes.

Les syntagmes correspondent à l'organisation des monèmes selon les règles morpho-syntaxiques (Chomel-Guillaume et al., 2010).

On peut retrouver différentes perturbations langagières chez les sujets aphasiques, notamment des troubles de la fluence, des déviations phonétiques, une apraxie bucco-faciale, une anomie, des paraphasies, des mots de prédilection, du jargon, des troubles syntaxiques, des troubles de la prosodie et de l'hypophonie. La description détaillée de ces perturbations figure en annexe p. 2.

2.4. Les différents types d'aphasie

De nombreuses classifications des aphasies ont été proposées. Elles se sont fondées sur différentes distinctions destinées à classer les syndromes aphasiques. Il a été proposé une distinction entre les aphasies sensorielles et les aphasies motrices ou encore entre les aphasies réceptives et expressives. Cependant, ces distinctions ne sont pas toujours opérantes par rapport à la complexité du fonctionnement langagier et des symptômes cliniques observés. La distinction la plus largement admise repose actuellement sur la notion de fluence. Cela permet de distinguer deux grandes catégories d'aphasies, les aphasies fluentes et les aphasies non fluentes. Cette distinction a été à l'origine du plus grand nombre de classifications, notamment la classification classique établie par l'école de Boston dans les années 1960. Cette classification a ensuite été révisée de nombreuses fois par Geschwind, Benson, Goodglass et Kaplan, afin d'y intégrer les nouvelles avancées dans le domaine de l'aphasiologie. Cette classification a initialement identifié neuf grands syndromes, parfois regroupés sous le terme d' « aphasies corticales » : les aphasies de type Broca, Wernicke, de conduction, anomique, transcorticales (motrice, sensorielle, mixte), globale et non fluente mixte. On a depuis identifié d'autres types d'aphasie, notamment les

aphasies sous-corticales (Chomel-Guillaume et al., 2010).

Le protocole que nous allons présenter dans le cadre de ce mémoire permet de recueillir des données qui peuvent être analysées et comparées à d'autres données regroupées au sein d'une base de données informatisée appelée AphasiaBank. Celle-ci regroupe des données langagières de personnes aphasiques et de sujets contrôles. Les corpus langagiers des sujets aphasiques y sont regroupés selon les différents syndromes aphasiques. Les types d'aphasie retenus dans le cadre d'AphasiaBank sont les suivants : l'aphasie de Broca, l'aphasie globale, l'aphasie transcorticale motrice, l'aphasie de Wernicke, l'aphasie anomique, l'aphasie de conduction et l'aphasie transcorticale sensorielle.

Nous allons présenter ces différents types d'aphasie en nous référant à la classification proposée par Gil (1999), qui présente l'avantage de regrouper et synthétiser différentes classifications. Nous ne décrivons cependant pas l'ensemble de cette classification mais seulement les aphasies retenues dans la classification de la base de données AphasiaBank. L'ensemble de la classification présentée par Gil figure en annexe p. 7.

Pour désigner les aphasies non fluentes et les aphasies fluentes, Gil (1999) emploie les termes d'aphasies à langage réduit et d'aphasies à langage fluide. Il précise que la description des aphasies dites corticales ne signifie pas que ces aphasies ne peuvent pas dans certains cas être associées à des lésions sous-corticales. Il ajoute également que des lésions purement sous-corticales peuvent donner un tableau clinique identique à celui des aphasies corticales. Nous avons choisi de ne pas détailler ici les aphasies sous-corticales car celles-ci ne sont pas mentionnées dans la classification retenue par AphasiaBank.

2.4.1. Les aphasies à langage réduit (d'après Gil, 1999)

Dénominations	Symptomatologie et localisation
<ul style="list-style-type: none"> - Aphasie de Broca - Aphasie d'expression ou aphasie motrice corticale (Déjérine) - Aphasie motrice efférente ou aphasie motrice cinétique (Luria) - Aphasie de réalisation phonématique (Hecaen) 	<p>Perturbations langagières</p> <p><u>Expression orale</u></p> <ul style="list-style-type: none"> - Réduction quantitative du langage, pouvant se limiter à la production de stéréotypies - Agrammatisme - Désintégration phonétique rendant la production orale lente, laborieuse et hachée (suppressions de phonèmes, simplifications de groupes consonantiques, tendance à utiliser des consonnes sourdes plutôt que sonores) - Peut être associée à une apraxie bucco-faciale - Dissociation automatico-volontaire (possible préservation du langage automatique) - Prosodie émotionnelle préservée, mais prosodie linguistique altérée (parfois syndrome de l'accent étranger) - Manque du mot (facilitation phonologique efficace) <p><u>Compréhension orale</u> : peu ou pas altérée</p> <p><u>Langage écrit</u> : souvent altéré</p> <ul style="list-style-type: none"> - Paragraphies de type dysorthographique et agrammatisme - Alexies ou de dyslexies de formes variables

	<p>Troubles associés</p> <ul style="list-style-type: none"> - Hémiplégié droite totale ou à prévalence brachio-faciale - Parfois hémianopsie latérale homonyme - Etat dépressif ou réactions de catastrophe (du fait de la conscience des troubles ou de la topographie lésionnelle frontale) <p>Localisation lésionnelle</p> <ul style="list-style-type: none"> - Partie triangulaire (aire 45 de Brodmann) et partie operculaire (aire 44 de Brodmann) du gyrus frontal inférieur gauche - Peut impliquer des lésions des aires corticales proches et, plus en profondeur, de la capsule interne
<ul style="list-style-type: none"> - Aphasie globale - Aphasie totale (Déjérine) - Grande aphasie de Broca 	<p>Perturbations langagières</p> <ul style="list-style-type: none"> - Suspension totale ou quasi-totale du langage (oral et écrit) - Absence de compréhension (oral et écrit) <p>Localisation lésionnelle et troubles associés</p> <ul style="list-style-type: none"> - Lésions massives pré et rétro-sylviennes gauches : généralement associée à une hémiplégié sensitivo-motrice importante - Lésions non contiguës des aires de Broca et de Wernicke, épargnant le faisceau pyramidal : pas de déficit moteur
<ul style="list-style-type: none"> - Aphasie transcorticale motrice - Adynamie de la parole (Kleist et Pick) - Aphasie dynamique (Luria) 	<p>Perturbations langagières</p> <p><u>Expression orale</u></p> <ul style="list-style-type: none"> - Aspontanéité de l'expression verbale, pouvant aller jusqu'au mutisme (aspontanéité levée par des épreuves de répétition) - Tendance à l'écholalie - Manque du mot (facilité par l'ébauche orale) <p><u>Compréhension orale et écrite</u> : préservée</p> <p>Localisation lésionnelle</p> <ul style="list-style-type: none"> - Lésions de l'aire motrice supplémentaire - Atteinte profonde de la substance blanche sous-corticale reliant l'aire de Broca à l'aire motrice supplémentaire

Tableau n°1 : Les aphasies à langage réduit (d'après Gil, 1999)

2.4.2. Les aphasies à langage fluide (d'après Gil, 1999)

Dénominations	Symptomatologie et localisation
<ul style="list-style-type: none"> - Aphasie de Wernicke - Aphasie sensorielle 	<p>Perturbations langagières</p> <p><u>Expression orale</u></p> <ul style="list-style-type: none"> - Expression orale fluide, voire logorrhéique - Nombreuses paraphasies phonémiques et verbales formelles, mais également sémantiques et verbales, et néologismes - Discours parfois incompréhensible, parfois jargon - Dyssyntaxie <p><u>Compréhension orale</u></p> <ul style="list-style-type: none"> - Trouble majeur de la compréhension verbale <p><u>Langage écrit</u></p> <ul style="list-style-type: none"> - Alexie - Nombreuses paralexies et paragraphies <p>Troubles associés</p> <ul style="list-style-type: none"> - Anosognosie - Hémianopsie latérale homonyme <p>Localisation lésionnelle</p> <ul style="list-style-type: none"> - Aire de Wernicke (aire associative auditive située à la partie postérieure de la face externe de la circonvolution temporale supérieure au niveau de l'aire 22) - Partie adjacente du lobe pariétal constituée du gyrus supramarginal (aire 40) et du gyrus angulaire (aire 39)

<ul style="list-style-type: none"> - Aphasie anomique - Aphasie amnésique de Pitres 	<p>Perturbations langagières</p> <p><u>Expression orale</u></p> <ul style="list-style-type: none"> - Manque du mot isolé <p><u>Compréhension orale</u> : préservée</p> <p><u>Langage écrit</u> : préservé</p> <p>Localisation lésionnelle</p> <ul style="list-style-type: none"> - Circonvolution temporale inférieure
<ul style="list-style-type: none"> - Aphasie de conduction - Aphasie centrale décrite (Goldstein) - Aphasie motrice afférente 	<p>Perturbations langagières</p> <p><u>Expression orale</u></p> <ul style="list-style-type: none"> - Atteinte de la structure phonologique des mots : production de nombreuses paraphasies phonémiques et verbales formelles, fréquentes conduites d'approche <p><u>Compréhension orale et écrite</u> : préservée</p> <p>Localisation lésionnelle</p> <ul style="list-style-type: none"> - Lésions du faisceau arqué, reliant les aires de Wernicke et de Broca, localisées au niveau de la substance blanche pariétale - Atteinte associée du cortex du gyrus supra-marginal et du faisceau arqué, le plus souvent dans sa partie postérieure proche du cortex pariétal
<ul style="list-style-type: none"> - Aphasie transcorticale sensorielle - Aphasie de Wernicke de type II (Roch-Lecours et Lhermitte) 	<p>Perturbations langagières</p> <p><u>Expression orale</u></p> <ul style="list-style-type: none"> - Expression fluente - Capacités de répétition intactes <p><u>Compréhension orale et écrite</u> : atteinte massive</p> <p>Localisation lésionnelle</p> <ul style="list-style-type: none"> - Lésions postérieures à l'aire de Wernicke dans une zone qualifiée de bordante (Benson) incluant les aires 37 et 39 (gyrus angulaire), pouvant s'étendre au cortex visuel associatif, alors que l'aire de Broca, l'aire de Wernicke et le faisceau arqué ne sont pas touchés

Tableau n°2 : Les aphasies à langue fluide (d'après Gil, 1999)

2.5. Handicap et communication dans l'aphasie : mieux évaluer le discours déviant

La notion de handicap a récemment évolué avec la révision de la Classification Internationale du Fonctionnement du handicap, publiée en 2001 par l'OMS.

La précédente classification du handicap datait de 1980. Elle définissait trois niveaux d'atteinte, établis à partir des travaux de Wood : la déficience, l'incapacité et le désavantage.

La nouvelle classification adopte une terminologie plus positive et ajoute les dimensions personnelles et environnementales à la conception du handicap. D'après cette classification : « *l'état de fonctionnement et de handicap d'une personne est le résultat de l'interaction dynamique entre son problème de santé (maladies, troubles, lésions, traumatismes) et les facteurs contextuels qui comprennent à la fois des facteurs personnels et environnementaux* » (Crevier-Buchman, 2005, p. 4). Cette classification définit cinq niveaux : les fonctions organiques, les structures anatomiques, les activités, les participations ou implications dans la vie réelle et les facteurs environnementaux (Crevier-Buchman, 2005).

Le handicap n'est donc plus conçu comme un état figé mais comme un processus dynamique, fruit de l'interaction entre un trouble de santé, des facteurs individuels et environnementaux. Ainsi, deux personnes ayant un trouble de santé similaire pourront être confrontées à des limitations et à

des restrictions d'activité différentes.

A partir de cette nouvelle conception, le handicap généré par les troubles de la communication de la personne aphasique recouvre deux dimensions : une dimension individuelle, liée aux capacités de l'individu, et une dimension sociale, liée aux situations sociales auxquelles l'individu est confronté.

Le dictionnaire d'orthophonie définit la communication comme « *tout moyen verbal ou non verbal utilisé par un individu pour échanger des idées, des connaissances, des sentiments, avec un autre individu* » (Brin-Henry et al., 2004, p. 57). La communication est fondamentale pour l'être humain, elle lui permet en effet d'entrer en interaction avec son environnement social. Lorsque cette disposition naturelle à communiquer est perturbée suite à un trouble du langage acquis, le handicap généré altère la qualité de vie de la personne aphasique.

Si la communication repose en partie sur le non verbal, elle fait également appel aux moyens d'expression verbale dont dispose l'individu. Cette expression verbale s'effectue par le langage oral, et la mise en discours de pensées et d'émotions revêt ainsi un rôle fondamental dans les interactions entre locuteurs, dans la cadre d'une communication « normale » (communication non dégradée) ou dans le cas d'une communication altérée (suite à un déficit neurologique chez des patients aphasiques). En effet, le discours est défini comme l'expression d' « *un groupe d'énoncés visant à véhiculer un message entre interlocuteurs* » (Ska et al., 2004, p. 302, cité par Wright, 2011).

Ainsi, le discours est fréquemment perturbé, en surface, chez les patients aphasiques, et présente différentes déviations linguistiques pouvant affecter les niveaux d'organisation phonétique, phonologique, lexico-sémantique et syntaxique. Ces altérations discursives entravent donc la communication et perturbent l'interaction entre la personne aphasique et son environnement, générant ainsi un handicap.

Dans son article, publié dans la revue *Aphasiology*, Wright insiste sur l'intérêt de l'étude du discours en aphasiologie (Wright, 2011). Elle montre que l'analyse du discours est essentielle car les adultes aphasiques présentent très souvent des difficultés à communiquer du fait de la perturbation de leurs habiletés discursives. Elle affirme que l'analyse des compétences discursives des patients peut permettre d'objectiver les aptitudes à communiquer des individus et que les variations observées dans la production du discours peuvent être utilisées pour évaluer l'efficacité et l'adéquation des traitements proposés.

L'évaluation du discours présente donc un intérêt fondamental pour caractériser les troubles de la communication dont souffrent les personnes aphasiques, et permettre ainsi de proposer des traitements adaptés afin d'atténuer le handicap.

3. L'ÉVALUATION EN CLINIQUE ORTHOPHONIQUE

3.1. L'évaluation orthophonique de l'aphasie

L'évaluation orthophonique a pour but d'évaluer les capacités linguistiques mais également les capacités communicationnelles de la personne aphasique. Cette évaluation vise à caractériser de manière fine les capacités préservées et déficitaires d'un patient, afin de déterminer son profil sémiologique en référence à une classification syndromique des aphasies, et afin d'établir un programme de rééducation le plus adapté possible au profil du patient.

L'évaluation peut être réalisée à différents moments de la prise en charge de la personne aphasique. Selon le délai entre la survenue de l'aphasie et le moment de l'évaluation, la stratégie d'évaluation ne sera pas la même. En phase aiguë, l'administration de tests standardisés est parfois difficile, l'évaluation revêt alors parfois un caractère informel visant à déterminer les capacités de communication de façon plus générale, ce qui n'empêche pas l'orthophoniste d'effectuer un bilan plus ciblé sur certains aspects du langage en fonction du profil du patient, s'il le juge nécessaire. En phase post-aiguë et chronique, des batteries de tests standardisées peuvent être utilisées, ce qui permet une évaluation plus détaillée et plus approfondie (Chomel-Guillaume et al., 2010).

3.1.1. Domaines de l'évaluation et outils existants

Pour l'évaluation de l'aphasie, les orthophonistes disposent de différents outils. Les outils les plus couramment utilisés en France sont :

- *l'échelle d'évaluation de l'aphasie (HDAE)*, qui est une adaptation française du *Boston Diagnostic of Aphasia Examination (BDAE)*, publié en 1972 par Goodglass et Kaplan. Cette adaptation a été réalisée par Mazaux et Orgogozo en 1981. (Mazaux et Orgogozo, 1981).
- *le protocole Montréal-Toulouse d'examen linguistique de l'aphasie (MT86)*, conçu par Nespoulous et Roch-Lecours en 1986, et révisé en 1992. Il est constitué de deux versions, la version *Alpha*, qui est un examen clinique standardisé, et la version *Bêta* qui permet d'établir un examen linguistique détaillé du langage (Nespoulous et al., 1986).
- *le test pour l'examen de l'aphasie*, élaboré par Ducarne de Ribaucourt en 1965, et révisé en 1975 et en 1989 (Ducarne de Ribaucourt, 1989).
- *Le test de langage élaboré pour adultes (TLE)*, coordonné par Rousseaux et Dei Cas en 2012 (Rousseaux et Dei Cas, 2012).

Nous allons d'abord présenter les principaux axes d'évaluation proposés par ces outils d'évaluation globale des fonctions langagières, puis nous aborderons des outils d'évaluation plus spécifiques à certains domaines, notamment l'impact de l'aphasie sur la vie quotidienne.

• Les outils d'évaluation globale des fonctions langagières du patient aphasique

Dans la plupart de ces batteries, l'évaluation orthophonique de capacités langagières de l'aphasie s'articule autour de cinq axes principaux : l'expression orale, la compréhension orale, l'expression écrite, la compréhension écrite et les praxies bucco-faciales. Certaines batteries proposent en outre l'évaluation du langage élaboré.

L'**expression orale** est évaluée selon différents domaines. On évalue généralement le langage spontané et le discours narratif, les capacités de dénomination, la disponibilité lexicale, les automatismes linguistiques et les capacités de répétition.

Le langage spontané est évalué lors d'un entretien semi-dirigé au cours duquel l'examineur incite le patient à produire un discours libre, en lui posant différentes questions (*HDAE, MT86, test pour l'examen de l'aphasie*). Le discours narratif est évalué à partir d'une description d'image (*HDAE, MT86, test pour l'examen de l'aphasie*), ou parfois par la restitution d'un texte lu ou entendu. Ces deux épreuves permettent d'évaluer les capacités d'élaboration langagière, de disponibilité lexicale, d'encodage syntaxique, la cohérence du discours mais également la qualité des productions sur les plans phonétique, phonologique et sémantique.

La dénomination d'image permet d'évaluer l'accès lexical à partir d'un support iconographique, ainsi que la qualité de la production orale. Les images proposées peuvent être des objets (*HDAE, MT86, test pour l'examen de l'aphasie*), des symboles, des nombres, des couleurs, des parties du corps (*HDAE*) ou encore des actions (*HDAE, MT86, test pour l'examen de l'aphasie*). Mais il existe également des épreuves de dénomination par le contexte, ne faisant pas appel à un support imagé (*HDAE*). Les items sont généralement choisis selon des critères de fréquence, de classe grammaticale, d'opposition entre le nom générique et le nom spécifique (ex : meuble/table dans le *MT86*), ainsi que selon le critère manipulable ou non manipulable (ex : échelle/village dans le *MT86*) (Viader et al., 2010).

Une épreuve de disponibilité lexicale est généralement proposée. Elle vise à évaluer les capacités d'évocation lexicale sans support iconographique. Le patient est généralement invité à donner un maximum de mots appartenant à un même champ sémantique (fluence catégorielle) (*HDAE, MT86*) ou commençant par une même lettre (fluence formelle), en un temps limité. Mais certains tests ne proposent ni critère sémantique, ni critère formel, le patient est seulement invité à dire le plus de mots possible (*test pour l'examen de l'aphasie*). Cette épreuve fait également intervenir les fonctions exécutives car elle nécessite la mise en place d'une stratégie efficace de recherche de la part du patient.

Les automatismes linguistiques sont évalués par la production de séries automatiques telles que les chiffres, les jours de la semaine, les mois de l'année (*HDAE, MT86, test pour l'examen de l'aphasie*) ou le chant (*HDAE, MT86*), ainsi que par la complétion de phrases (*HDAE*).

L'évaluation des capacités de répétition permet d'évaluer les capacités expressives mais également audiophonatoires. On peut proposer une répétition de phonèmes, de syllabes (*test pour l'examen de l'aphasie*), de mots simples et complexes (*HDAE, MT86, test pour l'examen de l'aphasie*), de non mots (*MT86*) et de phrases simples et complexes (*HDAE, MT86, test pour l'examen de l'aphasie*). Les mots et les phrases proposés sont sélectionnés selon des critères de complexité, de longueur et de concrétude (*HDAE*). La répétition de mots permet d'évaluer les capacités articulatoires du patient. La répétition de phrases évalue également la mémoire de travail et les capacités phonologiques.

La **compréhension orale** peut être évaluée par des épreuves de désignation, d'exécution d'ordres simples et complexes ou par des épreuves de logique et de raisonnement.

Les épreuves de désignation peuvent concerner des objets, des images et des parties du corps (*HDAE, MT86*). La désignation d'images consiste à apparier un mot ou une phrase entendue à sa représentation iconographique. L'appariement d'un mot entendu à une image permet d'évaluer la compréhension lexicale. La désignation est effectuée parmi un choix multiple de différents items pouvant contenir un ou plusieurs distracteurs. Les distracteurs peuvent être phonologiques (bouton/mouton dans le *MT86*), sémantiques (bouton/fermeture éclair dans le *MT86*) ou visuels (bouton/roue) dans le *MT86*. L'appariement d'une phrase entendue à une image permet d'évaluer la compréhension lexicale, syntaxique et morphologique. Les distracteurs proposés peuvent concerner le niveau lexical (« *la fille court/le garçon court* » dans le *MT86*) et le niveau morphosyntaxique (ex : « *la petite fille montre la dame qui pousse le bébé/ la petite fille qui montre la dame pousse le bébé* » dans le *test pour l'examen de l'aphasie*)(Viader et al., 2010).

Par rapport à l'épreuve de désignation d'images, l'exécution d'ordres simples et complexes (*HDAE, MT86, test pour l'examen de l'aphasie*) implique davantage d'habiletés pratiques et gestuelles.

Enfin, la compréhension orale peut être évaluée par des épreuves consistant à répondre à des questions orales par oui et par non, comme l'épreuve de logique et de raisonnement (*HDAE*) ou l'épreuve de logique verbale du *TLE*.

L'**expression écrite** peut être évaluée à travers différentes épreuves notamment l'expression écrite spontanée (*HDAE, MT86, test pour l'examen de l'aphasie*), la narration écrite (*HDAE, MT86*), la dénomination écrite (*HDAE, MT86, test pour l'examen de l'aphasie*), la production écrite d'automatismes (*HDAE*), la copie de mots et de phrases (*HDAE, MT86*), ainsi que l'écriture sous dictée de mots et de phrases (*HDAE, MT86, test pour l'examen de l'aphasie*).

La **compréhension écrite** peut être évaluée par des appariements de mots et de phrases écrits à des images (*HDAE, MT86, test pour l'examen de l'aphasie*). De même que dans les épreuves de compréhension orale, différents types de distracteurs peuvent être proposés. La compréhension écrite peut également être évaluée par des tâches de discrimination littérale ou verbale (*HDAE, MT86*) et des tâches d'association de segments de phrases et de textes à partir

d'un choix multiple de phrases. Si le patient ne présente pas de troubles de l'expression orale trop importants, on peut également lui proposer de réaliser un récit oral à partir d'une histoire lue (*test pour l'examen de l'aphasie*). Enfin, l'épreuve de lecture à haute voix teste les capacités visuo-phonatoires (*HDAE, MT86, test pour l'examen de l'aphasie*).

Un examen des **praxies bucco-faciales** est également réalisé afin d'apprécier la mobilité des organes bucco-faciaux (*MT86, test pour l'examen de l'aphasie*). Cette évaluation est indispensable et constitue un pré-requis à l'évaluation arthrique et phonétique.

Certaines batteries proposent également une évaluation du **langage élaboré**. Le *test pour l'examen de l'aphasie* propose par exemple des définitions de mots, de la concaténation de phrases ou des définitions de formules métaphoriques. Le *Test de langage élaboré* pour adultes (*TLE*) évalue la modalité orale du langage élaboré, en particulier les aspects du langage qui ont des rapports avec la sémantique, la syntaxe et la pragmatique. Pour tester le langage élaboré, il propose entre autres des définitions de mots, de la concaténation de phrases, une recherche de synonymes, une épreuve de discours procédural, des recherches d'intrus ou encore des explications de proverbes.

• **Les outils d'évaluation spécifiques**

À ces échelles globales s'ajoutent des outils d'évaluation plus spécifiques permettant d'évaluer de façon plus approfondie certains domaines. Nous pouvons par exemple citer la *DO80* de Deloche et Hannequin (Deloche et Hannequin, 2007) ou le *DVL 38* (Hammelrath, 1999) pour la dénomination, et le *Lexis* de Partz et al. (De Partz et al., 2002) pour l'accès lexical et la compréhension lexico-sémantique.

Il existe également des outils permettant d'évaluer les capacités de communication de façon plus écologique et l'impact de l'aphasie sur la vie quotidienne. Nous pouvons par exemple citer *l'échelle de communication verbale de Bordeaux* (Darrigrand et Mazaux, 2002), le *test lillois de la communication* (Rousseaux et al., 2001), la *Grille d'Analyse Linguistique d'Interaction libre* (Sainson et Guyou, 2016) ou encore le *protocole Montréal d'Évaluation de la Communication* (Joanette et al., 2005). (Viader et al., 2010). Ce dernier concerne davantage les sujets cérébro-lésés droits mais peut s'avérer intéressant à proposer aux sujets aphasiques dans certains cas.

Certains de ces outils évaluent à la fois les aspects structurels et pragmatiques du langage, d'autres n'abordent que l'un ou l'autre de ces deux aspects. L'outil que nous souhaitons proposer dans le cadre de ce mémoire ne concerne que l'analyse des aspects structurels du langage et n'aborde pas les aspects pragmatiques.

3.1.2. L'analyse et l'interprétation des résultats

La passation des épreuves précédemment citées donne lieu à des résultats d'ordre quantitatif et qualitatif. Ces résultats peuvent généralement être reportés sur des grilles d'analyse permettant de visualiser le profil du patient et d'orienter vers un syndrome aphasique particulier.

Les résultats quantitatifs permettent d'établir un score chiffré, sous forme de points ou de pourcentages, pour chaque épreuve, et de déterminer si le score obtenu est considéré comme normal ou pathologique, par rapport aux scores obtenus par des sujets contrôles ou d'autres sujets aphasiques. Ces résultats permettent également d'effectuer des comparaisons entre les scores obtenus aux différentes épreuves ou entre différentes variables étudiées au sein d'une même épreuve.

Les résultats quantitatifs sont complétés par des résultats qualitatifs permettant de mettre en évidence le type d'erreurs observées et les modes de facilitations efficaces pour le patient (Viader et al., 2010).

L'interprétation finale des résultats doit tenir compte de caractéristiques propres au patient, notamment son niveau socio-culturel et professionnel, son niveau scolaire, son comportement linguistique habituel et son état psychologique au moment de la passation du test. L'interprétation des résultats doit également tenir compte de caractéristiques inhérentes aux modalités de test, notamment les canaux sensoriels impliqués dans la passation des épreuves. Il faut en effet pouvoir déterminer si l'échec à une épreuve particulière est réellement imputable à un déficit de la fonction linguistique supposée être testée, ou au canal sensoriel utilisé, pouvant être altéré par l'existence de troubles associés (Viader et al., 2010).

3.2. L'évaluation orthophonique du discours continu oral en aphasiologie

Il existe de nombreux outils permettant d'évaluer le discours dans le cadre de différentes pathologies, notamment dans les troubles du langage oral chez l'enfant ou des personnes atteintes de maladies neuro-dégénératives. Le discours peut également être évalué dans le cadre de neurochirurgies pour soigner des patients atteints de tumeurs cérébrales (Fleurier, 2013). Une étude a montré que seule l'évaluation du langage spontané permettait de souligner l'apparition de déficits langagiers chez le patient qui n'étaient pas mis en évidence par une tâche de dénomination seule (Fleurier, 2013). Cela a ainsi montré l'intérêt fondamental d'évaluer le discours au même titre que les aspects lexicaux du langage afin d'avoir un aperçu global du langage du patient et de rendre compte de tous les paramètres impliqués dans une situation naturelle de conversation.

Nous n'aborderons pas en détail l'ensemble des outils existants mais nous nous centrerons sur l'évaluation orthophonique du discours en aphasiologie.

3.2.1. Les paramètres d'évaluation du discours

L'analyse du discours permet d'appréhender l'expression orale d'un point de vue transphrastique, dans des conditions de production plus proches des situations de communication réelles. En effet, en situation naturelle, l'individu s'exprime généralement par la production d'un discours, et non par des mots ou des phrases isolées.

Cependant, l'analyse du discours se révèle complexe. En effet, celui-ci met en jeu différents processus impliquant des habiletés linguistiques mais également d'autres habiletés cognitives plus larges, ainsi que des compétences pragmatiques et sociales. Certains des tests d'évaluation évoqués précédemment ne traitent que des aspects linguistiques, d'autres intègrent les aspects cognitifs, pragmatiques et sociaux. Dans le cadre de ce mémoire, nous avons choisi de traiter uniquement l'évaluation des aspects linguistiques formels. Nous ne traiterons donc pas des autres aspects cognitifs impliqués et des compétences pragmatiques et sociales.

L'évaluation du discours doit prendre en compte la pluralité des modes de discours. En effet, comme Wright l'a relevé, les différents processus linguistiques sont sensibles à la variété des tâches proposées (Wright, 2011). L'évaluation du discours doit donc proposer différentes tâches destinées à évaluer les différents modes de discours, que nous avons définis précédemment p. 6.

Le discours énonciatif peut être évalué à travers des tâches de production de discours spontané, sous forme de discussion libre, ou éventuellement induit par des questions posées par l'examineur.

Le discours narratif et descriptif peut être évalué par des tâches consistant à décrire ou raconter une histoire connue ou à partir d'un support imagé.

Le discours argumentatif peut être évalué par une discussion autour d'un thème, le patient étant incité à se positionner et à donner son point de vue sur un thème particulier.

Enfin, le discours procédural peut être évalué par des tâches nécessitant de décrire la réalisation d'une action.

Le recueil des différents types de discours doit ensuite faire l'objet d'une analyse linguistique précise. Cette analyse doit reposer sur une étude des différents niveaux du langage : les niveaux phonétiques, phonologiques, lexico-sémantiques et syntaxiques. Une analyse plus large impliquant les compétences pragmatiques doit également être menée, mais ce point ne sera pas traité dans le cadre de ce mémoire.

3.2.2. Les outils d'évaluation orthophoniques du discours en aphasiologie

Nous allons présenter les épreuves visant à évaluer le discours proposées dans les principales batteries de tests destinées à l'évaluation de l'aphasie.

- **L'échelle d'évaluation de l'aphasie (HDAE)** (Mazaux et Orgogozo, 1981)

Dans l'échelle d'évaluation de l'aphasie (HDAE), le discours oral est évalué à travers trois types d'épreuves. Une épreuve de conversation, constituée de sept questions, est proposée dans le but de faire parler le patient sous forme de discussion libre. Une autre épreuve de langage spontané, dans laquelle l'examineur doit inciter le patient à parler pendant une dizaine de minutes en relançant si nécessaire la conversation avec des questions sur sa profession et l'histoire de sa maladie, est également proposée. Ces deux épreuves visent à évaluer le discours énonciatif. Enfin, une dernière épreuve, destinée à évaluer le discours narratif et descriptif, consiste à décrire une image.

Le discours ainsi obtenu est analysé selon une échelle de gravité de l'aphasie et un profil de l'expression orale spontanée. L'échelle de gravité de l'aphasie repose sur un score allant de 0 (aucune expression intelligible et aucune compréhension orale) à 5 (handicap linguistique à peine perceptible, ou seulement ressenti par le malade, sans que l'auditeur puisse l'objectiver). Le profil de l'expression orale spontanée propose une cotation allant de 1, pour les altérations les plus importantes, à 7, pour l'absence d'anomalie. Six domaines sont évalués : la prosodie, la longueur des phrases (plus longues séries de mots produites au cours d'une même émission), l'articulation (facilité d'expression au niveau phonologique), la syntaxe (constructions grammaticales, même partielles), les paraphasies et le contenu informatif. On obtient ainsi un score mais celui-ci repose sur des données essentiellement qualitatives issues des cotations établies par le clinicien.

Parmi les résultats obtenus à ces épreuves, seul le score de l'échelle de gravité de l'aphasie peut être comparé à celui d'autres sujets aphasiques car il figure sur la feuille de profil z-score et on dispose de la moyenne et de l'écart-type obtenus par d'autres sujets aphasiques. Le profil d'expression orale spontanée ne peut être comparé ni à des sujets témoins, ni à d'autres sujets aphasiques.

- **Le protocole Montréal-Toulouse d'examen linguistique de l'aphasie (MT86)**
(Nespoulous et al., 1986)

La version Bêta du *protocole Montréal-Toulouse d'examen linguistique de l'aphasie* évalue le discours grâce à deux types d'épreuves : une interview dirigée et une description d'image.

L'interview dirigée est composée de douze stimuli destinés à faire produire un discours libre au patient sous forme de discussion. Les questions proposées appellent à une réponse par oui ou par non (ex : « êtes-vous marié ? »), d'autres sont ouvertes et incitent à élaborer un bref discours. Certains stimuli permettent d'évaluer le discours énonciatif (ex : *raconter ses loisirs, un voyage ou l'histoire de la maladie,...*), d'autres évaluent le discours descriptif (ex : « *décrivez moi votre appartement* »). Le discours argumentatif est également évalué à travers des questions portant sur le travail des femmes.

Le discours narratif oral est évalué à travers une épreuve incitant le patient à raconter une histoire présentée sur une image.

Les différents types de discours ainsi obtenus sont évalués au niveau quantitatif et qualitatif. Au niveau qualitatif, pour l'ensemble de ces épreuves, l'examineur est invité à noter l'absence ou la présence discrète, modérée ou sévère, des anomalies suivantes : réduction quantitative, agrammatisme, manque du mot, déviations phonétiques, déviations phonémiques (et/ou néologismes), déviations verbales. La compréhension orale est également évaluée lors de l'épreuve d'interview dirigée, l'examineur indiquant pour chaque stimuli si le patient semble avoir compris ou pas la question.

Au niveau quantitatif, l'interview dirigée permet d'établir un score de 0 à 18. Ce score correspond au nombre d'item-stimuli compris par le patient. Cette cotation quantitative permet donc d'évaluer la compréhension mais pas la production.

La cotation quantitative de l'épreuve de discours narratif oral repose sur une double analyse : lexico-sémantique et syntaxique. Pour l'analyse lexico-sémantique, l'examineur procède à un relevé des éléments de l'image verbalisés par le patient. Il précise également si ces éléments ont été restitués sous la forme d'un lexème, d'un morphème grammatical ou d'un syntagme plus complexe. Pour l'analyse syntaxique, l'examineur relève les types de structures syntaxiques suivantes : phrase complète stéréotypée, phrase minimale complète, phrase minimale lacunaire, phrase étendue, phrase avec proposition(s) enchâssée(s). Il précise également si ces structures véhiculent un contenu référentiel ou modalisateur.

Les résultats obtenus par un patient à ces épreuves ne peuvent pas être comparés à ceux d'autres sujets contrôles ou aphasiques car on ne dispose pas de moyenne et d'écart-type des sujets contrôles ou aphasiques.

- **Le test pour l'examen de l'aphasie** (Ducarne de Ribaucourt, 1989)

Le *test pour l'examen de l'aphasie* évalue le discours oral selon quatre types d'épreuves. Une première épreuve propose des questions sur l'état civil du patient (*nom, âge, profession,...*). Une deuxième épreuve est un dialogue contraint constitué de questions (ex : « *que fait-on quand le réveil sonne ?* » « *pourquoi y a-t-il beaucoup de touristes à Paris ?* »), évaluant le discours argumentatif. Une troisième épreuve évalue le discours énonciatif en incitant le patient à produire un récit personnalisé sur sa vie familiale, sa maladie ou son métier. Enfin, le discours narratif et descriptif est évalué à partir d'une description d'image. Cette dernière épreuve est regroupée avec l'épreuve de dénomination.

Ces épreuves font ensuite l'objet d'une cotation à la fois qualitative et quantitative. Pour toutes les épreuves d'expression orale, une cotation qualitative, reposant sur l'appréciation des signes sémiologiques aphasiques suivant trois degrés de gravité, est proposée. Elle concerne les signes

suivants : anosognosie, incitation, persévération, défaut d'accès lexical, réduction syntaxique, déviations phonémiques et graphémiques, déviations sémantiques, dyssyntaxie, troubles arthriques.

Une cotation quantitative est également proposée pour chacune de ces épreuves. Les réponses aux questions sur l'état civil sont notées sur un point selon l'adéquation de la réponse. Pour les épreuves de dialogue contraint et de récit personnalisé, chaque réponse est notée sur deux points si la réponse est exacte avec un lexique et une syntaxe adéquats et sur un point si l'information transmise est parcellaire. Enfin, la description d'image est notée sur cinq points, un point étant attribué pour chaque action verbalisée dans son intégralité.

Cette analyse quantitative évalue donc simultanément la compréhension et l'expression sur un plan formel et signifiant. L'analyse ne permet pas de distinguer ce qui relève de l'un ou l'autre de ces aspects.

Les résultats obtenus par un patient ne peuvent être situés par rapport à une population de sujets contrôles ou aphasiques, en l'absence de données permettant d'effectuer cette comparaison.

- **Le test de langage élaboré pour adultes (TLE)** (Rousseaux et Dei Cas, 2012)

Le *test de langage élaboré pour adultes* évalue le discours à travers trois types d'épreuves. Il propose tout d'abord une épreuve de discours procédural dans laquelle il est demandé au patient de détailler une activité de la vie quotidienne (se laver les mains par exemple). Il propose ensuite une épreuve de discours « déclaratif », qui correspond au mode d'organisation du discours narratif, où le sujet doit compléter une histoire dont le début est lu par l'examineur. Cette épreuve présente l'intérêt de vérifier si les patients sont capables de produire une suite de phrases cohérente et correcte sur le plan syntaxique. Il propose enfin une épreuve de discours argumentatif où le patient est invité à se positionner sur un fait de société.

Une cotation à la fois quantitative et qualitative est proposée pour chacune des épreuves.

La cotation quantitative de l'épreuve de discours procédural permet d'établir un score pouvant aller de zéro à trois points pour chaque item selon le nombre d'étapes citées par le patient. Pour l'épreuve de discours déclaratif, la cotation évalue la capacité du patient à raconter une histoire en cohérence par rapport aux détails donnés et la complexité syntaxique du récit. Chaque réponse peut être notée sur trois points si l'histoire est cohérente, la syntaxe correcte et le sujet d'énonciation respecté, deux points après une incitation, un point si le sujet modifie le sujet d'énonciation et zéro point si le récit est incohérent, que le sujet a besoin d'être incité à plusieurs reprises ou s'il ne répond pas. Pour le discours argumentatif, sur le même modèle, la cotation va de trois à zéro points selon le caractère cohérent de la réponse de la personne et sa capacité à argumenter ses idées. Ces cotations quantitatives sont donc issues d'une analyse qualitative

effectuée par le clinicien.

Il est également possible d'effectuer une analyse qualitative, selon différents paramètres tels que le type, la qualité, ou encore la précision des réponses.

Au niveau quantitatif, le score obtenu à chaque épreuve de discours peut être renseigné dans une grille. Le score total est calculé en additionnant le score obtenu à chacune des épreuves. Celui-ci permet ensuite situer les résultats du patient par rapport à une norme de sujets contrôles tenant compte du niveau d'éducation.

- **Le Bilan Informatisé d'Aphasie (BIA)** (Gatignol et al., 2012)

Le *Bilan Informatisé d'Aphasie (BIA)* est une évaluation informatisée des capacités verbales des sujets atteints de lésions cérébrales acquises (Gatignol et al., 2012) (cf description détaillée p.39).

Ce bilan évalue le discours oral à partir de deux épreuves. Une épreuve de langage oral spontané permet, grâce à une série de six questions, d'apprécier la qualité du discours énonciatif. Une épreuve de description d'image permet quant à elle d'évaluer le discours narratif et descriptif.

L'épreuve de langage oral spontané est cotée sur six points. Chaque réponse formulée correctement à la fois au niveau sémantique et syntaxique est comptabilisée comme correcte. Cette cotation quantitative est donc basée sur une analyse qualitative effectuée par le clinicien.

La cotation de l'épreuve de description d'image est à la fois qualitative et quantitative. Le patient dispose de deux minutes pour décrire l'image avec un maximum de détails, le clinicien comptabilisant chacune des réponses cohérentes produites et chaque persévération. Il relève également le nombre et le type de paraphasies produites par le patient.

Pour chacune de ces épreuves, il est possible de comparer les résultats à ceux obtenus par des sujets contrôles. Pour une analyse qualitative, l'examineur peut par ailleurs noter dans une partie « commentaires » ou sur un cahier de passation les productions ou le type d'erreurs produites par le sujet.

3.2.3. Synthèse des outils actuels

Les épreuves de discours proposées par les outils actuels sont variées et permettent d'évaluer la majeure partie des aspects constitutifs du discours.

L'analyse du discours proposée par ces différents tests est à la fois qualitative et quantitative. Les cotations quantitatives, permises par ces batteries, sont effectuées grâce à l'intuition linguistique, l'expérience et la pratique clinique de l'orthophoniste qui priment pour mener ce type d'évaluation, et sont donc tributaires du jugement du clinicien. Bien que ce jugement soit indispensable à l'analyse clinique, la question de la fiabilité inter-examineur peut se poser car ces systèmes de cotation peuvent parfois permettre des cotations différentes pour une même

production.

Ces cotations quantitatives permises par certains de ces outils permettent une comparaison des résultats obtenus avec un groupe de sujets contrôles (*échelle gravité de l'aphasie du HDAE, BIA, TLE*), mais pas avec un groupe de sujets aphasiques.

Cependant ces outils ne proposent pas d'analyse détaillée et objective des aspects formels du discours produit par le patient. Concernant cet aspect spécifique, on ne dispose donc d'aucun référentiel objectivé permettant de situer les résultats du patient par rapport à d'autres sujets, ce qui rend difficile l'appréciation précise de la sévérité des altérations linguistiques formelles par rapport à la production de discours pathologiques ou nonpathologiques.

De même, en l'absence de critères formels précis et objectifs, la comparaison des résultats obtenus par un même patient à deux temps différents risque d'être rendue imprécise.

Ainsi, une analyse du discours plus détaillée au niveau formel et reposant sur des critères objectifs pourrait permettre d'améliorer la fiabilité des résultats obtenus et faciliterait la comparaison entre les résultats obtenus par un patient et un autre groupe de sujets et l'évaluation d'un patient à différents moments de la prise en charge.

4. L'UTILISATION DES OUTILS INFORMATIQUES DANS L'ÉTUDE DU DISCOURS EN APHASIOLOGIE

4.1. Outils informatisés actuels dédiés à l'évaluation orthophonique de l'aphasie en France

Les ressources informatisées pour la prise en charge sont nombreuses. Le site Ortho & Co propose ainsi des liens vers des outils technologiques comme des logiciels ou des applications qui peuvent être utilisés avec les patients (Site Ortho & Co). Le site Aphasia Software Finder propose également des liens vers de nombreux logiciels ou applications en langue anglaise plus particulièrement destinés à la rééducation des patients aphasiques (Site Aphasia Software Finder).

S'il existe, à ce jour, un certain nombre d'outils informatisés pour la prise en charge de la personne aphasique, il en existe peu dédiés à l'évaluation de l'aphasie en France (Gatignol et al., 2012, p. 7). Nous pouvons néanmoins citer le *BIA* (Gatignol et al., 2012), la *G.A.L.I.*, (Sainson et Guyou, 2016), la *BIMM* (Gatignol et Marin Curtoud, 2007) et l'*AndroidBAT* (Cook et al., 2013).

Le *Bilan Informatisé d'Aphasie* (Gatignol et al., 2012) propose une version courte et une version longue. La version longue évalue les domaines suivants : expression orale, compréhension orale, expression écrite, compréhension écrite, mémoire et langage élaboré. Chaque domaine d'évaluation se compose de plusieurs épreuves. Le module « expression orale »

propose par exemple une épreuve de langage oral spontané, de description d'image, de fluence verbale ou encore de dénomination sur entrées visuelles, auditives et tactiles.

Le *BIA* permet une évaluation à la fois qualitative et quantitative du langage du sujet. Pour chaque épreuve, l'examineur valide ou invalide la réponse du patient en cliquant sur la touche correspondante à l'écran. Pour certaines épreuves, c'est le sujet lui-même qui indique sa réponse en cliquant sur l'icône appropriée. Cela permet une correction et un calcul automatique des résultats du sujet aux différentes épreuves. Cela permet également de chronométrer le temps de chaque épreuve et d'enregistrer le temps de réponse du sujet pour chaque item. Pour chaque épreuve figurent une moyenne et un écart-type, calculés à partir des moyennes et écarts-types des sujets contrôles. Pour une analyse qualitative, l'examineur peut par ailleurs noter dans une partie « commentaires » ou sur un cahier de passation les productions ou le type d'erreurs produites par le sujet.

Il s'agit ainsi, à la différence du protocole que nous proposons, d'une batterie « classique » d'évaluation assistée par ordinateur. L'outil informatique ne fournit pas de mesures supplémentaires concernant le langage du patient, mais sert à assister le thérapeute dans la passation du bilan afin de simplifier le calcul des scores.

La *Grille d'Analyse Linguistique d'Interaction libre (G.A.L.I)* constitue également un outil informatisé au service de l'évaluation du discours du patient cérébro-lésé (Sainson et Guyou, 2016). Cependant, l'utilisation informatique ne sert pas les mêmes objectifs que le protocole que nous proposons. En effet, l'outil informatique est simplement utilisé pour comptabiliser les réponses du patient et rendre la passation plus aisée pour le clinicien.

Nous pouvons également citer la *Batterie Informatisée du Manque du Mot (BIMM)* qui est un outil d'évaluation informatisé du manque du mot (Gatignol et Marin Curtoud, 2007). Grâce à l'outil informatique, les résultats sont présentés dans un compte rendu comprenant, les scores sous forme de notes et de profils, les erreurs qualitatives avec leurs pourcentages d'apparition et les résultats items par items. L'outil informatique apporte ainsi des informations concernant les erreurs qualitatives, mais uniquement dans le cadre d'une épreuve de dénomination et non de discours.

Enfin, le projet *AndroidBAT* a permis de transformer le *Test d'Aphasie chez les bilingues (BAT)* en un outil virtuel (Cook et al., 2013). Il s'agit d'une application sur tablette qui permet une collecte de données directement intégrée au test, comme le temps de réponse du participant ou encore la variabilité de vitesse des productions orales ou écrites de celui-ci.

Ainsi, les outils informatisés existant actuellement présentent l'intérêt de faciliter le recueil de données langagières par le thérapeute et le calcul des scores. Ils apportent néanmoins peu de précisions quantitatives sur les aspects formels du discours oral du patient aphasique.

4.2. Les bases de données et les projets de développement collaboratif dans l'étude du discours en aphasiologie

4.2.1. Les principes d'une base de données

Une base de données est un ensemble de moyens permettant de collecter, regrouper, stocker et classer des informations. Les informations stockées sur une base de données peuvent ainsi être facilement accessibles à l'ensemble des utilisateurs de la base de données, selon des conditions d'accès définies par les concepteurs du système. Les bases de données peuvent être plus ou moins ouvertes concernant la possibilité par les utilisateurs de rajouter des données à la base de données existante et concernant l'accès aux informations qui y sont stockées.

Les bases de données présentent donc de nombreux intérêts. Elles permettent de stocker des données, généralement en ligne sur internet, et ainsi de conserver des informations. Elles permettent également de mettre en commun différentes informations, provenant de plusieurs sources, à un ensemble d'utilisateurs. Les bases de données facilitent donc la diffusion des informations et des connaissances. Certaines bases de données sont collaboratives. Elles permettent à leurs utilisateurs d'ajouter des informations à la base de données, cette dernière s'enrichissant ainsi par les contributions de chacun.

Elles présentent donc un certain nombre d'avantages qui ont incité certains chercheurs ou cliniciens à mettre en place des bases de données collaboratives à des fins de recherche et de pratique clinique.

• **L'Outil d'Aide au Diagnostic Orthophonique (OADO)**

En France, c'est dans cette perspective qu'a été élaborée la base de données intitulée « Outil d'Aide au Diagnostic Orthophonique » (OADO). Cette base de données est hébergée sur le site internet de l'Union Nationale pour le Développement de la Recherche et de l'Évaluation en Orthophonie (UNADREO). Elle vise à regrouper les profils de patients cérébro-lésés souffrant de pathologies acquises afin d'aider les orthophonistes à établir un diagnostic et mettre en place un plan de rééducation adapté au profil de leur patient. C'est une base de données collaborative, tous les orthophonistes qui le souhaitent peuvent s'y inscrire gratuitement et bénéficier d'un nom d'utilisateur et d'un mot de passe leur permettant d'avoir accès à la base de données et d'y contribuer. Les orthophonistes qui souhaitent ajouter le profil d'un de leur patient à la base de données sont invités à entrer, dans le respect de l'anonymat, différentes informations concernant leur patient telles que son identité, sa latéralité, l'étiologie et les manifestations de sa pathologie, l'histoire de sa maladie, les hospitalisations, les résultats d'imagerie, les zones anatomiques affectées et ses résultats obtenus au bilan orthophonique. Des calculs statistiques automatiques peuvent être mis en œuvre afin de connaître les résultats statistiques des patients en fonction des

lésions et des subtests d'outils utilisés. Cet outil permet de comparer le profil d'un nouveau patient aux profils présentant les mêmes caractéristiques déjà contenus dans le logiciel. On obtient ainsi une projection des fonctions à examiner pour ébaucher un diagnostic et mettre en place des pistes de rééducation (UNADREO, 2013).

D'autres initiatives du même type ont été mises en œuvre au niveau international. Nous allons à présent présenter la base de données collaborative internationale AphasiaBank, dans laquelle s'inscrit le protocole d'évaluation informatisée du discours que nous présentons dans ce mémoire.

Comme l'OADO, la base de données AphasiaBank est une base de données collaborative à laquelle chacun peut contribuer librement. Ces deux bases de données permettent de regrouper les profils de plusieurs patients afin d'effectuer des comparaisons entre ces différents profils contenus sur la base de données. Cependant l'OADO ne permet de stocker, d'effectuer des calculs et des comparaisons que sur des résultats chiffrés obtenus à des tests orthophoniques alors que la base de données AphasiaBank propose de stocker, d'effectuer des calculs et des comparaisons sur des échantillons de discours recueillis auprès des sujets. De plus, la base de données AphasiaBank stocke des données issues de sujets aphasiques et non aphasiques, alors que l'OADO ne contient que des profils de sujets cérébro-lésés.

4.2.2. La base de données internationale AphasiaBank : intérêt pour la recherche fondamentale

- **CHILDES, TalkBank, AphasiaBank**

AphasiaBank est une base de données informatisée regroupant des échantillons de discours provenant de personnes aphasiques et de sujets contrôles. Elle fait partie d'une plus large base de données appelée TalkBank. Elle est accessible sur le site internet AphasiaBank, à l'adresse suivante : www.talkbank.org/AphasiaBank/.

Cette base de données a été conçue en 2005, afin de répliquer et d'étendre le projet CHILDES (*Child Language Data Exchange System*) à l'étude de la communication de la personne aphasique. Au départ, le projet CHILDES, conçu dans les années 1980, a mis à la disposition des chercheurs et des cliniciens une base de données et des outils de transcription et d'analyse informatisés dans le domaine de l'acquisition du langage de l'enfant. Ce système a ensuite été étendu à l'étude de l'aphasie grâce au projet AphasiaBank. Les projets CHILDES et AphasiaBank sont principalement financés par des subventions publiques provenant des *National Institutes of Health* (NIH).

La création de ces bases de données s'inscrit dans une approche fonctionnelle, et plus particulièrement dans la théorie émergentiste et le modèle de compétition développés par différents auteurs, tels que MacWhinney, l'un des concepteurs des projets CHILDES et

AphasiaBank. La théorie émergentiste cherche à rendre compte du rôle primordial du contexte dans l'émergence et l'acquisition du langage chez l'enfant. Le langage est une fonction de communication, il s'apprend donc en usage, en contexte, à travers des situations de communication. Le langage est considéré comme étant indissociable de son contexte d'énonciation. L'analyse du langage doit ainsi se faire à partir de situations de communication authentiques, et c'est dans cette perspective que ces bases de données ont été constituées. Elles permettent en effet d'effectuer des analyses linguistiques sur des données langagières recueillies dans un contexte d'interaction « naturelle ». Nous n'avons cependant pas développé la théorie émergentiste dans notre partie théorique car elle traite de l'émergence du langage chez l'enfant, ce qui n'est pas notre objet d'étude.

• Le protocole AphasiaBank

Les échantillons de discours stockés sur AphasiaBank sont récoltés à partir d'un protocole de recueil de données standard spécifiquement conçu pour le projet AphasiaBank.¹ Ce protocole est en langue anglaise. Pour le moment, seule une adaptation en espagnol a été proposée. Il est constitué de différentes tâches de discours et d'épreuves complémentaires visant à évaluer les autres aspects du langage oral des sujets aphasiques. Les tâches de discours comportent des tâches de discours spontané, le sujet étant amené à raconter l'histoire de sa maladie, sa récupération et un événement important de sa vie, des tâches de narration d'histoire à partir de supports imagés, une tâche de narration d'un conte de fée, et une tâche de discours procédural. Les batteries de test précédemment citées proposent des tâches de discours similaires. Le discours énonciatif y est également évalué à travers des tâches de discours spontané, comprenant aussi l'histoire de la maladie. Des tâches de narration d'histoire à partir de supports imagés sont également proposées, mais une seule image est présentée au sujet alors que trois supports différents sont proposés au sujet dans le protocole AphasiaBank. Parmi les autres tests que nous avons précédemment évoqués, seul le *TLE* dispose d'une épreuve de discours procédural. La narration d'un conte de fée n'est pas proposée dans les batteries précédemment citées. Le *TLE* propose cependant une tâche de discours « déclaratif » dans laquelle le sujet est amené à compléter une histoire. En outre, le discours argumentatif n'est pas analysé par le protocole AphasiaBank alors qu'il est évalué dans le *test pour l'examen de l'aphasie*, le *MT86* et le *TLE*.

Le protocole AphasiaBank comprend également des épreuves complémentaires à réaliser auprès des sujets aphasiques afin de préciser leur profil langagier, telles que des épreuves de répétition, de dénomination de noms, de dénomination de verbes, de compréhension orale de phrases, de logique et de raisonnement, ainsi que des épreuves issues de la *Western Aphasia Battery-Revised (WAB)* (Kertész, 2007, cité par MacWhinney et al., 2011) destinées à évaluer le

¹ Le protocole AphasiaBank en langue anglaise est disponible sur le site <http://talkbank.org/APhasiaBank/> à la rubrique « Protocols ».

type et le niveau de sévérité de l'aphasie.

Les données langagières issues des épreuves de discours sont recueillies à travers des enregistrements audio ou vidéo et sont ensuite retranscrites dans le format de transcription standard CHAT. Les transcriptions peuvent être reliées aux fichiers audio et vidéo, ce qui permet de lire les transcriptions sur écran tout en écoutant les entretiens enregistrés ou filmés. Les données retranscrites peuvent ensuite être analysées par différents programmes d'analyse automatique.

• **AphasiaBank : mode d'accès, constitution et objectifs**

La base de données AphasiaBank est une base de données collaborative, elle croît donc sans cesse au fur et à mesure des corpus de données contrôles et aphasiques déposés par ses membres. Cette base de données, les programmes d'analyse qui y sont reliés, et toutes les informations les concernant, sont accessibles gratuitement sur le site AphasiaBank. L'accès à certaines données, notamment aux échantillons de discours, est protégé par un nom d'utilisateur et un mot de passe, dont disposent tous les membres contributeurs d'AphasiaBank. Mais l'adhésion est automatiquement accordée à tous les chercheurs et cliniciens travaillant sur l'aphasie qui en font la demande. L'accès aux programmes, aux manuels et aux autres ressources est gratuit et ouvert à toute personne intéressée.

En mai 2016, la base de données en langue anglaise contenait des échantillons issus de 471 personnes aphasiques et 201 sujets contrôles, ce qui en fait la plus grande base de données informatisée d'échantillons de discours aphasique au monde. Les critères d'inclusion permettant d'inclure des sujets aphasiques dans la base de données sont pour l'instant limités aux personnes dont l'aphasie résulte d'un accident vasculaire cérébral qui peut être vérifié par neuro-imagerie ou par un diagnostic médical clair.

La base de données est principalement constituée d'échantillons en langue anglaise, mais des échantillons sont également recueillis dans d'autres langues, notamment en cantonais, en mandarin, en allemand et en suédois (MacWhinney et al., 2011). En langue française, une première contribution a été faite à la base de données par les travaux de Sahraoui (Sahraoui, 2009) et Sahraoui et Nespoulous (Sahraoui et Nespoulous, 2012), mais hors du cadre du protocole AphasiaBank.

Des données démographiques approfondies, ainsi que les résultats obtenus à des épreuves complémentaires, sont recueillis pour chaque participant inclus dans la base de données, afin de compléter le profil du patient ou du sujet contrôle. Toutes ces données sont accessibles aux membres d'AphasiaBank sur le site internet (MacWhinney et al., 2011).

L'objectif général d'AphasiaBank est de mettre en place un système permettant d'accumuler et de partager des données, de proposer des analyses automatiques précises et fiables afin

d'améliorer la connaissance du langage des personnes aphasiques et de contribuer à l'amélioration des traitements de l'aphasie. Cet objectif répond à des besoins importants tant dans le domaine de la recherche que dans le domaine de la clinique. En effet, au niveau de la recherche, cette base de données constitue un apport majeur. L'archivage et la mutualisation des données est une condition essentielle pour le progrès scientifique. Pourtant un partage et un échange de données libres et ouverts n'ont pas toujours été la norme dans l'étude du langage. Grâce à la mise en place de cette base de données, contrairement aux travaux antérieurs, les données, les procédures, et les résultats de ces analyses sont désormais intégralement rendus publics et peuvent donc être reproduits et même discutés par la communauté scientifique (MacWhinney et al., 2011).

Le projet AphasiaBank repose sur trois outils complémentaires et intégrés : la base de données, le format de transcription et de codage CHAT et le logiciel d'analyse CLAN. Nous allons présenter ces deux derniers outils, qui permettent de traiter et d'analyser les données langagières de la base de données de façon automatique.

4.3. Présentation des outils utilisés pour le traitement automatique informatisé du langage oral au sein du projet AphasiaBank

Le traitement automatique informatisé des échantillons de discours de la base de données AphasiaBank est rendu possible grâce à l'utilisation du système de transcription et de codage CHAT et du logiciel d'analyse CLAN.

4.3.1. Le système CHAT

Le système CHAT (*Codes for the Human Analysis of Transcripts*) est un système de conventions de transcription et de codage qui permet d'uniformiser le format des transcriptions de la base de données AphasiaBank. Ce système est conçu pour permettre la transcription et le codage informatique de tous types d'interactions conversationnelles, impliquant des locuteurs avec ou sans trouble du langage. Le système CHAT fournit également des codages permettant de signaler la présence d'erreurs phasiques telles que les paraphasies, les néologismes ou encore les erreurs morphologiques et grammaticales. Les fichiers transcrits et codés en format CHAT peuvent ensuite être analysés par le logiciel CLAN. Ils peuvent également être convertis en format XML (langue utilisée pour les documents textes sur le web) et ainsi être compatibles avec d'autres programmes d'analyse puissants tels que ELAN² ou Praat³ (MacWhinney, 2000a).

2 ELAN est un logiciel d'annotation linguistique de fichiers audio ou vidéo, développé à l'Institut Max Planck de psycholinguistique. Ce logiciel permet de créer, d'éditer, de visualiser et de rechercher les annotations effectuées sur les fichiers audio et/ou vidéo. (Hellwig, 2016)

3 PRAAT est un logiciel permettant de manipuler, de traiter et de synthétiser les sons vocaux. Il a été conçu à l'Institut de sciences phonétiques de l'Université d'Amsterdam par Paul Boersma et David Weenink. (Site PRAAT)

4.3.2. Le logiciel CLAN

Le logiciel CLAN (*Computerised Language Analysis*) est un logiciel libre qui permet d'effectuer de multiples analyses des données langagières transcrites en format CHAT. Le logiciel propose un nombre très important de programmes permettant d'effectuer des analyses précises très diverses sur les aspects phonologiques, lexicaux, morphosyntaxiques, discursifs et gestuels des interactions retranscrites (MacWhinney, 2000a). Les commandes CLAN sont très nombreuses et peuvent répondre à divers objectifs de recherche. Dans le cadre de ce mémoire, nous n'utiliserons qu'un des programmes proposés par le logiciel CLAN, le programme EVAL.

Le logiciel CLAN peut être téléchargé librement sur internet à l'adresse suivante : <http://childes.talkbank.org/clan/>. Les instructions d'installation du logiciel figurent en annexe p.10. Le logiciel CLAN est très fréquemment remis à jour, ce qui nécessite d'en réinstaller très souvent une nouvelle version. Ces changements sont fondamentaux pour la recherche car ils permettent d'améliorer constamment le codage et l'analyse des données langagières.

4.4. Analyse informatisée du discours en aphasiologie : fonctionnement du programme EVAL

4.4.1. EVAL : généralités

Le programme EVAL est l'un des nombreux programmes proposés par le logiciel CLAN. EVAL permet d'analyser les productions langagières d'un sujet et de dresser un profil de ses performances linguistiques par la production d'un tableau regroupant 30 mesures linguistiques calculées automatiquement par CLAN. Des exemples de tableaux de résultats obtenus par EVAL sont disponibles en annexe p. 26 et p. 36-39.

Le profil ainsi obtenu est affiché dans un tableau excel. EVAL permet également de comparer, si on le souhaite, les résultats obtenus par un sujet à ceux de patients similaires ou de sujets contrôles non aphasiques sélectionnés à partir de la base de données AphasiaBank. Cette comparaison n'est possible qu'en utilisant des tâches de discours issues du protocole AphasiaBank, auxquelles ont été soumis l'ensemble des sujets de la base de données. Il est également possible d'utiliser EVAL à partir d'autres tâches de discours et sans comparaison avec la base de données. Cela permet d'obtenir des scores bruts. Cependant, l'intérêt de la base de données est de pouvoir comparer les résultats obtenus par un patient à d'autres sujets, afin de mieux situer ses résultats. EVAL permet en outre de comparer les résultats obtenus par un même patient à deux temps différents, et peut ainsi contribuer à évaluer l'évolution d'un patient au fur et à mesure de la rééducation.

Un programme similaire, appelé KIDEVAL, a été conçu pour étudier l'acquisition du langage de l'enfant. Il permet d'effectuer une analyse proche de celle d'EVAL mais avec des indicateurs

plus adaptés à l'analyse du langage de l'enfant.

4.4.2. Préparation d'une transcription pour l'analyse EVAL

Une transcription doit passer par plusieurs étapes avant de pouvoir être analysée par EVAL. La transcription doit d'abord être transcrite et codée selon les règles du système CHAT. Elle doit ensuite passer par une étape de vérification et une étape d'étiquetage morphologique. La description technique détaillée de ces différentes étapes figure en annexe p. 11-22.

• **Règles de transcription et de codage pour l'analyse EVAL**

Les transcriptions destinées à être analysées par EVAL peuvent être transcrites et codées selon un système simplifié de transcription et de codage abrégé créé spécifiquement pour ce programme (cf annexe Coding Cheat Sheet, p. 28). Cependant EVAL fonctionne également sur des transcriptions créées en utilisant l'ensemble des normes et des codes CHAT (EVAL Manual, 2016). Il n'est donc pas nécessaire d'utiliser l'ensemble des règles du système CHAT.

Un fichier de transcription doit respecter un certain formalisme afin d'être correctement analysé par le programme.

Le fichier doit ainsi comporter un certain nombre de lignes ou d'en-têtes obligatoires. Ces en-têtes permettent de préciser le nom du participant⁴, ses données démographiques (*âge, sexe, type d'aphasie si le participant est aphasique*), la langue utilisée dans la transcription, la date de la passation, sa durée, ainsi que l'intitulé de chaque tâche de discours (*Gem*) administrée. La mention des tâches de discours au sein de la transcription permet ensuite d'effectuer des comparaisons spécifiquement entre certaines tâches (EVAL Manual, 2016).

Au sein de la transcription, chaque énoncé⁵ est inscrit sur sa propre ligne distincte et doit se terminer par un signe de ponctuation ou d'autres symboles permettant d'indiquer que l'énoncé a été laissé inachevé par le locuteur.

Un point important de la transcription est la question de la segmentation du discours en énoncé. Le manuel EVAL précise qu'il est parfois difficile de décider à quel endroit segmenter les énoncés. En cas de doute, le manuel recommande de préférer des énoncés courts plutôt que trop longs. Il est également conseillé de s'appuyer sur les indicateurs suivants, énoncés par ordre d'importance : la syntaxe, l'intonation, les pauses et la sémantique. Concernant la syntaxe, à moins qu'il n'y ait de fortes contre-indications prosodiques, une phrase bien formée est considérée comme un énoncé. Cependant, un énoncé ne doit pas nécessairement être grammaticalement

4 Le nom du participant est transcrit de la façon suivante : le nom de l'examineur, un nombre à deux chiffres indiquant que c'est la première (01), la deuxième (02), etc. que l'examineur a testée, suivi de la lettre a pour le test initial, b pour une deuxième passation, et ainsi de suite pour les passations suivantes auprès du même participant.

5 Dans le manuel d'utilisation d'EVAL, un énoncé est défini comme « un segment de discours en cours d'exécution qui semble former une unité cohérente » (Saffran et al, 1989, cité dans EVAL Manual, 2016).

correct pour être considéré comme un énoncé. Une intonation descendante, ou ascendante dans le cas d'une question, indique généralement la fin d'un énoncé. Les pauses et les critères sémantiques se révèlent être des indicateurs moins fiables que la syntaxe et l'intonation pour indiquer la fin d'un énoncé, en particulier chez les personnes aphasiques (EVAL Manual, 2016). Les difficultés rencontrées dans la décision de segmentation en énoncé peuvent être expliquées par les particularités du discours oral précédemment évoquées (cf p. 10).

• **Codages particuliers**

Les codages permettent au logiciel CLAN d'analyser correctement les transcriptions. Si certains codages ne sont pas nécessaires, d'autres sont particulièrement importants car leur utilisation a une influence sur les calculs effectués par EVAL. C'est le cas des codages concernant les répétitions, les reformulations, l'exclusion de mots ou d'énoncés, les énoncés inintelligibles, les abréviations et les codages d'erreurs (cf annexe p. 14-17).

De nombreux autres codages existent, tels que le codage des pauses, mais ils n'ont pas d'influence sur les calculs effectués par EVAL. Ils peuvent cependant être utiles pour le clinicien qui souhaite affiner l'analyse du discours recueilli. Le transcripteur est donc libre de choisir les codages qu'il souhaite utiliser et de décider ainsi du degré de finesse de son analyse.

• **Vérification de la transcription**

Une fois la transcription et le codage effectués, il est nécessaire de lancer le programme CHECK qui réalise une vérification de la transcription (cf annexe p. 17). Ce programme permet de mettre en évidence les erreurs concernant le respect des normes de transcription CHAT. Toutes les erreurs mises en évidence par ce programme doivent être corrigées afin d'obtenir le message suivant : « *Success ! No errors found* » (EVAL Manual, 2016).

• **Lien entre la transcription et les fichiers média**

Lorsque la transcription est terminée et vérifiée, il est possible de la relier au fichier audio ou vidéo correspondant, bien que cela ne soit pas nécessaire au fonctionnement d'EVAL.

L'alignement du fichier média à la transcription permet de fournir une information précise de la durée de chaque énoncé.

• **Étiquetage morphologique : génération de la ligne %mor**

L'étape suivante consiste à lancer le programme MOR qui va effectuer un étiquetage morphologique des parties du discours. Mais avant de pouvoir lancer le programme MOR, il est nécessaire de vérifier que le programme reconnaît tous les mots présents dans la transcription. En effet, le programme ne pourra attribuer la catégorie grammaticale au mot que si celui-ci figure dans

les fichiers lexiques de la grammaire MOR.

La grammaire MOR regroupe trois fichiers définissant les règles grammaticales de la langue et des fichiers lexiques regroupant les mots et affixes de la langue classés par catégories grammaticales. Il existe trois fichiers définissant la grammaire de la langue. Le fichier des règles d'allomorphe (ar.cut) répertorie la façon dont les morphèmes peuvent changer de forme. Le fichier des règles de concaténation (cr.cut) précise la façon dont les morphèmes peuvent se combiner dans la langue. Enfin, le fichier étiqueteur de formes spéciales (sf.cut) permet d'identifier les types lexicaux spéciaux tels que les néologismes ou les onomatopées. Ces fichiers permettent au programme MOR de retrouver la plupart des mots de la langue sans qu'ils soient nécessairement tous entrés dans les fichiers lexiques. Il suffit en effet d'entrer les radicaux et les affixes dans les fichiers lexiques, leurs formes fléchies seront ensuite identifiées par MOR par la mise en oeuvre des fichiers grammaires (MacWhinney, 2000b, p. 163-164). Les fichiers lexiques sont constitués par catégories grammaticales. Il existe ainsi un fichier regroupant tous les noms, un fichier regroupant tous les déterminants, un fichier regroupant tous les adjectifs et ainsi de suite. Au sein de chacun de ces fichiers lexiques, les mots sont classés par ordre alphabétique et leur catégorie grammaticale est précisée entre accolades.

Il est possible que certains mots de la transcription ne soient pas dans la grammaire MOR et ne puissent donc être reconnus par le programme. Cela peut se produire pour plusieurs raisons : le mot peut avoir été mal orthographié dans la transcription ou peut n'avoir jamais été utilisé dans d'autres transcriptions et donc ne pas avoir été inclus dans les fichiers lexiques. Afin que MOR puisse reconnaître toutes les structures morphologiques des mots de la transcription, certains codages doivent être utilisés pour indiquer le mot cible, lorsqu'il est connu, des formes phonologiques déviantes et des abréviations utilisées par le locuteur. Le programme MOR prendra ainsi en compte la forme standard du mot (MacWhinney et al., 2011).

Afin de s'assurer que tous les mots de la transcription figurent dans les fichiers lexiques de la grammaire MOR et peuvent donc être étiquetés par le programme d'étiquetage morphologique en tant que parties du discours, il faut lancer une commande spécifique du programme CLAN (cf annexe p. 19).

Si certains mots ne sont pas reconnus, il faut soit corriger les fautes de frappe ou d'orthographe qui peuvent expliquer que le programme ne les aient pas identifiés, soit les ajouter à la grammaire existante en spécifiant leur catégorie grammaticale. Si aucun mot n'est listé, MOR reconnaît tous les mots et on peut lancer le programme d'étiquetage morphologique (cf annexe p. 20) (EVAL Manual, 2016).

Ce programme va générer une ligne supplémentaire, la ligne %mor, sous chaque énoncé du locuteur en indiquant à quelle catégorie grammaticale appartiennent les mots employés par le locuteur (cf annexe p.21). Ces codes morphologiques peuvent ensuite être utilisés pour calculer automatiquement une grande variété d'analyses linguistiques (MacWhinney et al., 2011).

Le programme MOR liste toutes les catégories grammaticales et les analyses morphologiques possibles pour chaque mot. Un même mot peut appartenir à différentes catégories grammaticales selon le contexte dans lequel il est utilisé, et donc être listé dans plusieurs catégories grammaticales dans le lexique de la grammaire MOR. Par exemple, le mot « que » est listé plusieurs fois dans les fichiers lexiques MOR, il figure dans la liste des pronoms, dans celle des conjonctions et dans celle des prépositions. Le programme POST, lancé automatiquement après MOR, permet de résoudre ces ambiguïtés en déterminant la catégorie appropriée en fonction du contexte d'utilisation du mot, c'est-à-dire selon les mots situés avant et après lui (MacWhinney et al., 2011).

4.4.3. Les mesures linguistiques calculées par EVAL

Lorsque la transcription est passée par les étapes de vérification et d'étiquetage morphologique, on peut lancer l'analyse EVAL afin d'obtenir un profil des performances linguistiques du participant.

Toutes les mesures d'EVAL sont calculées à partir de la ligne %mor sauf les mesures FREQtupes et FREQtokens qui sont calculées à partir de la ligne principale. Par défaut, les phrases marquées d'un [+ exc] sont exclues de l'analyse. Il est cependant possible de les inclure dans l'analyse en tapant une commande spécifique. (EVAL Manual, 2016).

Nous allons décrire les trente mesures calculées automatiquement par le programme EVAL (EVAL Manual, 2016).

Mesure	Définition de la mesure
Duration	Cette mesure fournit la durée totale de l'échantillon sous le format heures:minutes:secondes. Si la transcription est reliée au fichier média (audio ou vidéo), EVAL calcule cet item automatiquement. Sinon, pour obtenir cette information, il faut chronométrer la session et ajouter une ligne @Time Duration aux en-têtes de la transcription.
Total Utts	Cette mesure calcule le nombre total d'énoncés. Elle comprend tous les énoncés utilisés dans le calcul de la longueur moyenne des énoncés (MLU), ainsi que les énoncés inintelligibles (codés xxx). Elle exclut les énoncés qui ne sont pas des mots, par exemple les énoncés codant seulement des gestes (&=geste).
MLU Utts	Cette mesure précise le nombre d'énoncés utilisé pour calculer la longueur moyenne des énoncés. Elle exclut les énoncés inintelligibles et les énoncés qui ne sont pas des mots. Il faut ajouter une commande spécifique (+s) si l'on souhaite intégrer les énoncés et les mots des énoncés avec xxx.
MLU Words	Cette mesure calcule la longueur moyenne des énoncés en mots. Elle ne prend pas en compte les mots qui sont dans des énoncés avec les codes xxx, yyy ou www.
MLU Morphemes	Cette mesure calcule la longueur moyenne des énoncés en morphèmes. Elle ne prend pas en compte les morphèmes qui sont dans des énoncés avec les codes xxx, yyy ou www.
FREQ types	Cette mesure précise le nombre total de types de mots, c'est-à-dire le nombre total de mots différents utilisés. Par défaut, cette mesure ne comprend pas les répétitions et les reformulations. Pour les inclure, il faut ajouter +r6 à la commande EVAL.

FREQ tokens	Cette mesure précise le nombre total de tokens, c'est-à-dire d'entités lexicales utilisées. Par défaut, cette mesure n'inclut pas les répétitions et les reformulations. Pour les inclure, il faut ajouter +r6 à la commande EVAL.
FREQ TTR	Cette mesure donne le ratio type/token. Ce ratio est obtenu en divisant le nombre de types de mots différents (types) par le nombre total de mots prononcés (tokens). Ce ratio type/token constitue donc une mesure de la diversité lexicale.
Verbs/Utt	Cette mesure donne le nombre de verbes par énoncé, ce qui correspond à peu près au nombre de propositions par énoncé. Le calcul de cette mesure inclut les verbes, les copules ⁶ , et les auxiliaires suivis de participes passé ou présent, mais n'inclut pas les modaux.
% Word Errors	Cette mesure fournit le pourcentage de mots qui sont codés comme des erreurs [*]. Cet item peut être affiché en nombre brut plutôt qu'en pourcentage en ajoutant +o4 à la commande EVAL.
Utt Errors	Cette mesure fournit le nombre d'énoncés codés comme des erreurs [+ *]. Ce code est destiné à marquer les erreurs pour un examen plus approfondi. Il ne permet pas une comparaison précise avec la base de données contrôle, car les erreurs au niveau de l'énoncé des contrôles ne sont pas codées.
Mor Words	Cette mesure indique le nombre de mots utilisés pour l'analyse par MOR. Contrairement aux mesures FREQ types, FREQ tokens et FRQ TTR, par défaut, MOR compte comme mots les clitiques. ⁷
Density	Cette mesure indique la densité des idées propositionnelles. Elle a été adaptée de la troisième version du « Computerized Propositional Idea Density Rater » (CPIDR3), développé sous la direction de Michael A. Covington. Cette mesure repose sur un ratio entre le nombre de propositions sémantiques et le nombre de mots et vise à refléter la qualité informative des propositions langagières.

Tableau n°3 : Mesures EVAL – Mesures générales (d'après EVAL Manual, 2016)

EVAL calcule également le pourcentage du nombre total de mots des catégories de mots suivantes :

Mesure	Partie du discours mesurée
% Nouns	Noms
% Plurals	Pluriels
% Verbs	Verbes (<i>inclut ceux étiquetés par MOR comme verbe, participe, copule et modal</i>)
% Aux	Auxiliaires
% 3S	Troisième personne du singulier
% 1S/3S	Formes identiques pour la première et la troisième personne du singulier
% Past	Formes verbales conjuguées au passé
% Past P	Participes passé
% PresP	Participes present
% Prep	Prépositions
% Adv	Adverbes

6 Un verbe copule est un marqueur du rapport prédicatif que l'attribut du sujet entretient avec le sujet. Le verbe être, ainsi que certains verbes d'état comme *devenir*, *rester*, *sembler* sont fréquemment utilisés comme des verbes copules. Le verbe copule est un élément purement relationnel et référentiellement vide (contrairement au verbe être lexical, synonyme d'exister) (Riegel et al., 2014, p. 422-423).

7 Un clitique est un mot qui ne peut être accentué car il fait corps avec le mot suivant. Ce sont surtout des déterminants simples, les pronoms clitiques antéposés à la forme verbale, les prépositions, les conjonctions et le premier élément *ne* de la négation (Riegel et al., 2014, p. 108).

% Adj	Adjectifs
% Conj	Conjonctions
% Det	Déterminants
% Pro	Pronoms

Tableau n°4 : Mesures EVAL – Calcul des différentes catégories de mots (d'après EVAL Manual, 2016)

Ces scores peuvent être affichés en nombres bruts en ajoutant +o4 à la commande EVAL.

Retracing [//]	Cette mesure calcule le nombre d'autocorrections
Repetition [/]	Cette mesure fournit le nombre de répétitions

Tableau n°5 : Mesures EVAL - Nombre de répétitions et de reformulations (d'après EVAL Manual, 2016)

4.4.4. Le lancement du programme EVAL

4.4.4.1. Comparaison d'un sujet à la base de données

Lors du lancement du programme EVAL, il est possible de choisir les caractéristiques du groupe de comparaison de la base de données auquel nous souhaitons comparer les résultats du participant.

Nous pouvons choisir un groupe de comparaison constitué de participants contrôles et/ou de participants aphasiques. Dans ce dernier cas, il est possible de déterminer le ou les types d'aphasie des participants que nous souhaitons inclure dans le groupe de comparaison. La classification des aphasies proposée par le programme EVAL est celle de la *Western Aphasia Battery-Revised (WAB)* (Kertész, 2007, cité par MacWhinney et al., 2011). Nous pouvons également préciser la tranche d'âge et le sexe des participants à inclure dans le groupe de comparaison, ainsi que la ou les tâches de discours AphasiaBank que nous souhaitons comparer à notre transcription.

Les résultats de l'analyse sont ensuite affichés dans un tableau excel présentant les résultats du participant ainsi que les scores moyens du groupe de comparaison sélectionné et l'écart de performance (en écart-types) entre le participant et le groupe de comparaison pour chacune des mesures calculées par EVAL (cf annexe p. 26 et 36-39) (EVAL Manual, 2016).

4.4.4.2. Comparaison d'un même sujet à deux temps différents

Il est également possible de comparer plusieurs transcriptions différentes sans effectuer de comparaison avec la base de données AphasiaBank. On peut par exemple comparer le discours d'un même patient à différents moments, pour analyser l'évolution du patient au cours du temps. Pour cela, il est possible d'utiliser des tâches de discours issues ou non du protocole AphasiaBank.

Les fichiers à analyser doivent avoir été préalablement préparés à l'analyse EVAL par les programmes CHECK et MOR.

Le programme EVAL effectue alors une analyse sur chacun des fichiers sélectionnés et affiche les résultats côte-à-côte dans la feuille de calcul afin qu'ils puissent être comparés (cf annexe p. 27) (EVAL Manual, 2016).

PARTIE 2 : PROBLÉMATIQUE ET HYPOTHÈSES

1. PROBLÉMATIQUE

1.1. Raisons du choix de cette problématique : Intérêts potentiels de l'analyse informatisée par EVAL du discours aphasique dans l'évaluation clinique orthophonique

1.1.1. L'importance de l'évaluation du discours en aphasiologie

L'évaluation du discours oral aphasique représente un enjeu majeur en clinique orthophonique. En effet, comme nous l'avons vu précédemment, le discours est central dans la communication et son altération constitue un handicap majeur altérant la qualité de vie de la personne aphasique. L'évaluation spécifique du discours revêt donc une place fondamentale dans l'évaluation du langage et de la communication de la personne aphasique car le discours incarne l'utilisation du langage, en situation, dans des conditions réelles de communication.

Cependant l'analyse du discours est particulièrement complexe car la mise en discours met en jeu de multiples niveaux d'organisation linguistique formels (*phonétique, phonologique, lexico-sémantique, morpho-syntaxique*) ainsi que pragmatiques, et l'aphasie peut perturber ces différents niveaux. Son analyse nécessite donc une étude fine et précise de ces différentes composantes, d'autant plus que l'analyse des erreurs en discours peut donner de précieuses indications permettant d'identifier les mécanismes déficitaires sous-jacents. Ainsi, l'évaluation du discours peut permettre de mieux caractériser les troubles de la communication du patient et de préciser le profil sémiologique, ce qui permettra de proposer une prise en charge la plus adaptée possible aux besoins du patient afin de le guider et l'accompagner dans son parcours de récupération, et de l'aider à compenser son niveau de handicap.

1.1.2. Apport potentiel d'EVAL par rapport aux autres outils d'évaluation du discours

Il existe différents outils destinés à évaluer le discours aphasique en orthophonie (cf p.34). Ces outils permettent de recueillir les différents types de discours (*énonciatif, narratif, descriptif, argumentatif, procédural*), mais, dans ces batteries, la finesse des analyses permise dans les autres subtests ne se retrouve pas toujours de façon aussi approfondie dans les épreuves de discours. L'analyse du discours proposée par ces outils repose sur des données quantitatives et qualitatives dont les cotations se révèlent pour beaucoup intuitives et subjectives. Se pose alors la question de la fiabilité inter-examineur dans la procédure de cotation et la question de la sensibilité à la détection de changements lors de l'analyse des productions d'un même sujet à

deux temps différents, sur les plans qualitatif et quantitatif.

Certains de ces outils permettent de comparer les scores issus des cotations quantitatives du discours d'un patient à ceux d'un groupe de sujets contrôles. Cependant, ils n'envisagent pas de comparaison entre les données discursives brutes (avant cotation) recueillies auprès d'un sujet aphasique et celles recueillies, dans une même tâche, à la fois par une population de sujets témoins et de sujets aphasiques présentant un profil de perturbation similaire.

Le discours lui-même et ses composants ne peuvent donc pas être directement comparés avec d'autres discours recueillis dans les mêmes conditions.

Or, une comparaison précise et directe du discours oral, notamment de ses aspects formels, pourrait permettre de mieux caractériser les perturbations langagières des patients aphasiques. En effet, en discours oral continu, on peut fréquemment observer des erreurs qui sont en lien avec les processus normaux d'élaboration du discours et qui ne relèvent pas forcément de processus déficitaires. Il est donc important de pouvoir comparer le discours d'un sujet aphasique à ceux d'une population de sujets contrôles afin de mieux discerner quelles erreurs relèvent de processus normaux et lesquelles relèvent de véritables processus pathologiques en lien avec l'aphasie.

Ainsi, afin d'améliorer l'évaluation du discours, il pourrait être utile de disposer d'outils permettant d'évaluer de façon objective et détaillée ses aspects formels, proposant une comparaison du discours produit par un patient à ceux d'une population de référence, constituée de sujets contrôles et de sujets aphasiques.

1.1.3. L'intérêt de l'utilisation d'EVAN dans l'évaluation du discours

Les outils informatisés comme CLAN et EVAN sont déjà très utilisés dans la recherche en aphasiologie. Une partie des études réalisées à partir des données AphasiaBank est présentée dans un article publié par MacWhinney dans le volume 25 de la revue *Aphasiology* (MacWhinney et al., 2011). L'utilisation de ces outils a beaucoup apporté à la recherche en améliorant le stockage, la transcription et l'analyse des données langagières aphasiques. Mais EVAN a également été conçu pour être utilisé par des cliniciens et il est actuellement utilisé à des fins cliniques aux Etats-Unis. Pour obtenir des mesures utiles en clinique, les cliniciens n'utilisent généralement pas l'ensemble des tâches du protocole AphasiaBank, mais uniquement une seule tâche, préférentiellement la tâche de narration à partir de l'image « Window ». La taille conséquente de la base de données anglaise (201 sujets contrôles en mai 2016) permet en effet d'obtenir des informations suffisamment pertinentes à travers l'analyse d'une seule tâche de discours.⁸

L'apport de l'utilisation de tels outils en clinique orthophonique n'a pour le moment pas encore

⁸ Communication personnelle avec Margie Forbes (mail du 20 mai 2016), membre fondateur d'AphasiaBank, et Jessica Richardson (mail du 21 mai 2016), clinicienne et chercheuse basée aux Etats-Unis.

été évalué en France. Les différents outils proposés par le projet AphasiaBank nous ont cependant amenés à réfléchir à cette question. Nous allons ainsi présenter ce qui peut nous amener à nous interroger sur l'intérêt potentiel de l'utilisation d'EVAL en orthophonie en France, en abordant les utilisations possibles de la base de données AphasiaBank, du système de transcription CHAT et de l'analyse informatisée du discours.

- **L'utilité de la base de données AphasiaBank**

La base de données AphasiaBank vise à regrouper des transcriptions d'échantillons de discours issus d'une population de sujets contrôles et aphasiques. Il est possible d'effectuer des comparaisons entre un échantillon de discours d'un patient et ceux contenus dans la base de données. Ceci pourrait permettre de mieux situer le discours d'un patient par rapport à d'autres sujets, soit contrôles, soit présentant le même type d'aphasie. De plus, la comparaison avec des corpus produits par des sujets non aphasiques pourrait permettre de déterminer le niveau d'erreurs indiquant que les perturbations observées relèvent de la pathologie et non plus de la normalité.

- **L'apport de l'utilisation du système de transcription CHAT**

Le système CHAT a été conçu pour transcrire des interactions langagières à partir d'enregistrements audio ou vidéo. Ces enregistrements permettent de contourner un certain nombre de difficultés posées par la transcription manuelle effectuée au cours de l'interaction. En effet, il est très difficile de noter avec précision les productions langagières d'un sujet étant donné le rythme rapide de l'écoulement de la parole. De plus, le fait de prendre des notes peut perturber l'interaction en cours. L'enregistrement audio ou vidéo semble donc indispensable à une évaluation précise du discours oral.

Les données recueillies doivent ensuite être retranscrites. Les transcriptions manuelles sont souvent très dépendantes du transcripateur. Chaque chercheur ou clinicien a souvent sa propre façon de transcrire et de coder les déviations aphasiques. De plus, les transcriptions manuelles sont plus difficilement partageables que les transcriptions informatisées qui peuvent être stockées informatiquement, mises en ligne sur une base de données et ainsi être partagées avec d'autres chercheurs ou cliniciens.

L'utilisation du système CHAT permet d'uniformiser les transcriptions et le codage des erreurs et ainsi de faciliter la lisibilité de la transcription par d'autres chercheurs et cliniciens. De plus, l'utilisation de ce système permet une double analyse du corpus retranscrit, par les cliniciens et les chercheurs d'une part et par les programmes d'analyse informatique d'autre part. En effet, l'utilisation de ce système rend possible la mise en œuvre de tout un ensemble de programmes d'analyse informatique du discours.

• L'apport de l'analyse informatisée

L'analyse informatisée du discours proposée par EVAL permet d'utiliser des programmes informatiques puissants afin d'analyser un grand nombre de variables linguistiques avec objectivité et précision. La cotation des résultats n'est ainsi plus dépendante du jugement de l'examineur mais calculée automatiquement par le programme informatique sur la base de ce qui a été produit par le patient puis transcrit.

La précision des mesures calculées par EVAL peut permettre d'effectuer des comparaisons précises entre différents sujets, mais également entre différentes productions d'un même sujet et ainsi de détecter des changements obtenus au cours de la prise en charge.

De plus, le fait que les mesures linguistiques soient calculées automatiquement plutôt que manuellement permet des calculs plus précis, plus rapides et autorise le calcul de certains indices qu'il aurait été impossible de réaliser manuellement.

En outre, ces analyses informatisées fournissent au clinicien un écart-type calculé automatiquement et avec précision, en prenant en compte plusieurs critères tels que le sexe, l'âge ou le type d'aphasie. Un écart-type est fourni pour chacune des mesures linguistiques étudiées, ce qui permet une analyse détaillée des différents paramètres du discours.⁹

L'analyse proposée par EVAL permet également d'affiner l'évaluation du discours en autorisant l'analyse de chaque type de discours de façon distincte. Ceci peut permettre de mettre en évidence des différences dans l'utilisation des différents types de discours par le patient.

Cette analyse informatisée pourrait donc constituer un outil intéressant pour le clinicien en vue d'affiner l'évaluation du discours. Cependant, cette analyse n'est pas suffisante en elle-même. Les mesures linguistiques donnent des indications sur les manifestations de surface mais l'analyse de ces manifestations et la mise en évidence des mécanismes sous-jacents pouvant les expliquer ne peuvent être effectuées que par le clinicien. Ces analyses informatisées ne prennent sens que si elles sont interprétées par le clinicien au sein d'une évaluation plus globale du langage et de la communication prenant en compte la personnalité du patient.

1.2. Problématique résumée

Problématique : Notre propos est d'explorer la possibilité d'utiliser le protocole AphasiaBank et le programme EVAL sur des sujets contrôles et aphasiques francophones, et ainsi de nous interroger sur la possibilité et la pertinence de l'utilisation d'EVAL en clinique orthophonique. Ainsi, le protocole anglo-saxon AphasiaBank est-il adaptable à la langue française et peut-il être administré à des sujets francophones ? Peut-on envisager la création d'une base de données en langue française, et rendre le programme EVAL opérationnel afin de pouvoir analyser des corpus langagiers francophones ?

⁹ Un exemple d'analyse effectuée par le programme EVAL figure en annexe p. 26 et p.36-39.

2. HYPOTHÈSES

Les hypothèses s'orientent sur trois domaines : l'adaptation du protocole AphasiaBank, l'utilisation du protocole sur des sujets francophones et l'adaptation et l'utilisation du programme d'analyse EVAL.

En effet, pour répondre à cette problématique, plusieurs étapes sont nécessaires. Le protocole AphasiaBank constitué d'épreuves de discours spécifiques et d'épreuves complémentaires évaluant le profil langagier oral du patient aphasique doit d'abord être adapté à la langue française. Le protocole doit ensuite être administré à des sujets contrôles et à des sujets aphasiques.

Les données langagières recueillies par le biais du protocole AphasiaBank doivent enfin être retranscrites en format CHAT sur le logiciel CLAN. De plus, afin de pouvoir utiliser toutes les fonctionnalités d'EVAL sur nos corpus, notamment la comparaison avec un groupe de référence, il est ensuite nécessaire d'initier la création d'une base de données en langue française en intégrant nos transcriptions à la base de données AphasiaBank. Le but n'est pas de créer une véritable base de données mais seulement d'initier sa création, en proposant une version française du protocole de recueil de données, ce qui permettra d'alimenter la base de données par la suite, afin de pouvoir utiliser EVAL et d'évaluer ses potentiels apports pour l'évaluation orthophonique. Nous nous centrerons dans ce mémoire sur les aspects linguistiques formels et n'aborderons pas les aspects pragmatiques car ceux-ci ne peuvent être analysés par EVAL mais relèvent de l'analyse du clinicien ou alors d'autres programmes de CLAN qui nécessiteraient des codages particuliers d'ordre pragmatique.

Au cours de la conception de notre problématique et de notre protocole, trois hypothèses ont progressivement émergé qui relèvent soit de l'adaptation du protocole, soit de son utilisation, soit de l'adaptation et l'utilisation d'EVAL.

Hypothèses

Hypothèse 1 : Le protocole AphasiaBank est adaptable à la langue française.

Sous-hypothèse 1 : Le protocole de recueil de données démographiques AphasiaBank est adaptable à la langue française.

Sous-hypothèse 2 : Le protocole de recueil de données discursives AphasiaBank est adaptable à la langue française.

Sous-hypothèse 3 : Le protocole de recueil de données issues des épreuves complémentaires AphasiaBank est adaptable à la langue française.

Hypothèse 2 : Le protocole AphasiaBank est utilisable avec des sujets contrôles et des sujets aphasiques francophones.

Sous-hypothèse 1 : Le protocole de recueil de données démographiques AphasiaBank est utilisable avec des sujets contrôles et des sujets aphasiques francophones.

Sous-hypothèse 2 : Le protocole de recueil de données discursives AphasiaBank est utilisable avec des sujets contrôles et des sujets aphasiques francophones.

Sous-hypothèse 3 : Le protocole de recueil de données des épreuves complémentaires AphasiaBank est utilisable avec des sujets contrôles et des sujets aphasiques francophones.

Hypothèse 3 : EVAL est opérationnel avec des corpus issus de sujets contrôles et de sujets aphasiques francophones.

Sous-hypothèse 1 : Les corpus francophones peuvent être transcrits et codés selon les normes CHAT.

Sous-hypothèse 2 : Les corpus francophones peuvent être vérifiés par le programme CHECK et les erreurs détectées peuvent être corrigées.

Sous-hypothèse 3 : Le programme d'étiquetage morphologique peut être lancé sur les corpus francophones.

Sous-hypothèse 4 : Les corpus francophones peuvent être mis en ligne sur la base de données française d'AphasiaBank.

Sous-hypothèse 5 : Le programme EVAL peut être lancé de façon opérationnelle sur les corpus francophones.

PARTIE 3 : PARTIE EXPÉRIMENTALE

Nous allons aborder les trois étapes successives de notre travail en expliquant dans un premier temps notre démarche méthodologique et dans un second temps les résultats obtenus suite à cette démarche.

Ainsi, nous présenterons tout d'abord la méthodologie adoptée concernant l'adaptation du protocole de recueil de données AphasiaBank, suivie des résultats concernant cette adaptation. Nous présenterons ensuite la méthodologie que nous avons suivie dans l'utilisation de ce protocole en vue d'initier la création d'une base de données française, puis les résultats concernant cette utilisation. Enfin, nous présenterons notre méthodologie concernant l'opérationnalisation des données pour l'utilisation d'EVAl, pour ensuite présenter les résultats de ce travail.

1. ADAPTATION DU PROTOCOLE DE RECUEIL DE DONNÉES APHASIABANK

1.1. Méthodologie dédiée à l'adaptation du protocole de recueil de données AphasiaBank

Le protocole AphasiaBank comporte deux versions, une version destinée aux participants contrôles et une version destinée aux participants aphasiques.

Ces deux protocoles contiennent des documents permettant de recueillir des données démographiques sur les participants et des instructions pour administrer les tâches de discours. Ces documents sont spécifiques à chacune des populations. Le protocole destiné aux participants aphasiques comporte en outre une liste d'épreuves complémentaires à administrer aux participants afin de préciser leur profil de langage oral.

L'ensemble de ces documents est disponible en langue anglaise sur le site d'AphasiaBank, à l'exception de certaines épreuves complémentaires du protocole des participants aphasiques.

Afin que ce protocole soit également accessible aux chercheurs et cliniciens français, nous avons effectué un travail de traduction et d'adaptation des documents proposés dans ce protocole. Ce travail d'adaptation a consisté en une réflexion afin de traduire les documents permettant le recueil des données démographiques, les instructions d'administration des tâches de discours, pour les participants contrôles d'une part et pour les participants aphasiques d'autre part, ainsi que les épreuves complémentaires proposées aux participants aphasiques, afin d'en proposer une version française.

Ainsi, notre démarche d'adaptation du protocole et notre réflexion concernant cette adaptation servent un double objectif : nous permettre d'utiliser le protocole AphasiaBank, d'initier la création

d'une base de données en langue française pour pouvoir utiliser EVAL, mais également de permettre par la suite aux cliniciens et aux chercheurs francophones d'utiliser le protocole AphasiaBank, EVAL et de pouvoir alimenter la base de données AphasiaBank avec d'autres corpus francophones.

1.1.1. Adaptation des documents permettant le recueil des données démographiques

Des données démographiques précises sur chaque participant inclus dans la base de données sont disponibles sur le site d'AphasiaBank. Ces données ne sont accessibles qu'aux membres d'AphasiaBank disposant d'un nom d'utilisateur et d'un mot de passe. Ces données sont présentées sous forme de tableurs excel. Un tableau regroupe les données démographiques de l'ensemble des participants contrôles de la base de données, un autre celles de l'ensemble des participants aphasiques. L'anonymat des participants est respecté puisque que les participants sont identifiés par le nom de l'examineur, suivi d'un nombre à deux chiffres indiquant que c'est la première (01), deuxième (02), etc. personne testée par l'examineur, et d'une lettre minuscule indiquant que c'est la première passation effectuée auprès de ce participant (a), la deuxième (b), etc.

Des formulaires permettant de recueillir les données démographiques auprès des participants ont été conçus et sont disponibles sur le site AphasiaBank. Il en existe un exemplaire pour les participants contrôles et un exemplaire pour les participants aphasiques.

Dans notre démarche d'adaptation du protocole AphasiaBank, nous avons proposé une version française des formulaires de recueil de données démographiques et des tableaux permettant de saisir les données ainsi recueillies.

1.1.1.1. Description des formulaires originaux à faire remplir par les participants

Les formulaires à destination des participants contrôles et aphasiques contiennent un certain nombre de rubriques similaires. Ils se distinguent néanmoins par l'ajout de questions plus spécifiques à l'une ou l'autre des populations.

Ainsi, dans le **formulaire destiné aux participants contrôles**, des questions supplémentaires sont posées afin de vérifier que les participants répondent aux critères d'inclusion de la base de données AphasiaBank. Des questions permettent de détecter d'éventuels troubles de la mémoire, des antécédents d'AVC, de traumatisme crânien ou d'autres problèmes neurologiques et enfin d'éventuels antécédents de troubles de la parole, du langage ou de la communication.

Le **formulaire destiné aux participants aphasiques** comporte également des questions spécifiques. Les participants aphasiques n'étant pas toujours en mesure de remplir le formulaire eux-mêmes, une question concerne les personnes qui ont aidé à remplir le formulaire. Les autres

questions permettent de connaître le profil du participant, notamment les troubles associés à l'aphasie (troubles moteurs, héminégligence), son ancienneté, ainsi que la durée et le contenu de la rééducation orthophonique suivie par le participant.

1.1.1.2. Adaptation des formulaires à faire remplir par les participants

Lors de l'adaptation des formulaires, nous avons essayé de rester le plus proche possible des formulaires originaux. Nous avons cependant dû adapter certaines formulations afin de rendre le formulaire plus approprié aux particularités culturelles françaises ou plus facile à remplir par les participants.

Nous avons choisi de traduire le terme « Race » par le terme « Nationalité » car si le terme de race est couramment utilisé aux Etats-Unis, il n'est culturellement pas acceptable en France, tous comme les choix multiples qui sont proposés à la suite (*White / African American / American Indian/Alaska Native / Asian / Hispanic / Latino / Native Hawaiian / Pacific Islander / Mixed / Other*), que nous avons décidé de ne pas garder dans la version française.

Dans un souci de clarté, nous avons également traduit la rubrique « Years of education » par « Nombre d'années d'étude effectuées depuis le CP ». Nous avons en effet pensé que le terme « Années d'étude » n'aurait pas été assez clair pour les participants concernant le niveau scolaire à partir duquel comptabiliser les années de scolarisation, ce qui aurait empêché d'obtenir des données fiables et comparables entre tous les participants.

Dans le **formulaire destiné aux participants aphasiques**, nous avons également dû ajouter une question concernant les antécédents de troubles de la parole, du langage et de la communication avant la survenue de l'AVC, puisque cela constitue un critère d'exclusion de la base de données AphasiaBank. Les formulaires tels que nous les avons adaptés figurent en annexe p. 40-42 (pour les participants contrôles) et p. 43-45 (pour les participants aphasiques).

1.1.1.3. Description des tableaux de recueil de données démographiques originaux de la base de données AphasiaBank

Ces tableaux, disponibles en langue anglaise sur le site AphasiaBank, se présentent sous forme de tableaux excel avec des menus déroulants permettant de remplir certaines catégories plus aisément.

Les tableaux comportent des rubriques générales et certaines plus spécifiques à l'une ou l'autre des populations.

Ainsi, dans le **tableau de recueil de données démographiques des participants contrôles** figurent des colonnes supplémentaires permettant de spécifier les scores obtenus à certains tests (cf p. 79) permettant de vérifier l'absence des critères d'exclusion des sujets contrôles.

Le **tableau de recueil de données démographiques des participants aphasiques**

comporte beaucoup plus de colonnes que celui des participants contrôles. En effet, il contient de nombreuses rubriques supplémentaires permettant de spécifier l'étiologie de l'aphasie, son ancienneté, son type, les troubles associés (atteinte motrice, apraxie de la parole, dysarthrie), ainsi que de nombreuses informations sur la lésion cérébrale à l'origine de l'aphasie (date, étiologie, localisation, description,...) et les lésions cérébrales antérieures, ainsi que sur la durée de la rééducation orthophonique suivie par le participant. Enfin, des rubriques permettent de préciser les années d'expérience de l'examineur dans le domaine de l'aphasie, le type et la durée de la relation entre le participant et l'examineur.

1.1.1.4. Adaptation des tableaux de recueil de données démographiques de la base de données AphasiaBank

Nous avons donc créé deux tableaux de recueil de données démographiques afin de proposer une version française des tableaux de recueil de données démographiques.

Pour adapter le **tableau concernant les participants contrôles** nous avons dû modifier l'intitulé de la colonne « Race » que nous avons traduit par « Nationalité ».

Le tableau tel que nous l'avons adapté figure en annexe p. 46-49.

Pour adapter le **tableau concernant les participants aphasiques**, nous avons décidé d'ajouter trois colonnes qui n'étaient pas dans la version originale du tableau : le nombre actuel de séances d'orthophonie hebdomadaire, la présence ou non d'antécédent de troubles de la parole, du langage ou de la communication et une colonne permettant de préciser lesquels, car cela fait partie des critères d'exclusion que nous avons retenus.

Nous avons également supprimé une colonne concernant la sélection du type d'aphasie selon le système de classification de Luria. Nous avons choisi de ne pas conserver ce système de classification dans la version française du tableau car cette classification n'est pas la classification « classique » utilisée par les cliniciens français (Chomel-Guillaume et al., 2010), il ne nous apparaissait donc pas pertinent de la conserver. Nous avons donc fait le choix de ne conserver qu'un type de classification, celui du système de Boston, qui est la classification utilisée par la base de données AphasiaBank.

Nous avons cependant noté quelques différences entre la classification proposée dans le menu déroulant du tableau de recueil de données démographiques et la classification proposée par EVAL lors de la sélection des paramètres de l'échantillon de la base de données. En effet, l'aphasie transcorticale mixte et l'aphasie optique apparaissent dans le tableau de recueil de données démographiques mais pas dans les choix de sélection proposés par EVAL. Nous n'avons cependant pas modifié ce paramètre en laissant les choix « aphasie transcorticale mixte » et « aphasie optique » dans le tableau.

Concernant le choix des termes, dans la sélection de l'étiologie de l'aphasie, nous avons

choisi de traduire le terme de « semantic dementia » par « variante sémantique de l'aphasie primaire progressive » en laissant le terme de démence sémantique entre parenthèses. Le terme de « démence sémantique » est l'ancienne terminologie attribuée à la variante sémantique de l'aphasie primaire progressive. Nous avons pensé que le terme de démence n'était pas approprié car un sujet peut être atteint d'une variante sémantique de l'aphasie primaire progressive, et ne pas être atteint de démence. Nous avons cependant laissé le terme de « démence sémantique » entre parenthèses afin que les cliniciens ou les chercheurs qui connaissent cette pathologie sous ce nom puissent l'identifier dans le tableau et de ne pas trop nous éloigner de la version originale.

Le tableau tel que nous l'avons adapté figure en annexe p. 50-58.

1.1.2 Adaptation du protocole de recueil de données langagières AphasiaBank

1.1.2.1. Adaptation des épreuves de discours

1.1.2.1.1. Description du protocole original destiné aux participants contrôles

Le protocole original, disponible sur le site AphasiaBank, est divisé en quatre sections correspondant aux quatre épreuves de production de discours suivantes :

Le « récit personnel » qui se trouve dans la section I correspond aux tâches de **discours spontané**. Les participants contrôles sont interrogés à propos d'une maladie ou d'une blessure, leur rétablissement suite à cette maladie ou blessure puis au sujet d'une expérience qu'ils ont vécue avec des personnes ayant des difficultés pour communiquer, et enfin à propos d'un événement important de leur vie.

La section II propose des tâches de **narration d'histoire à partir d'images**. Les images ainsi que les informations relatives aux droits d'auteur sont disponibles en annexe p. 59-60. Trois dessins en noir et blanc sont présentés aux participants. Ils sont invités à regarder l'image et à raconter une histoire avec un début, un milieu et une fin. La première image-stimulus est une image de quatre cases d'un enfant jouant avec un ballon de football et brisant une fenêtre, la seconde est une image de six cases d'un enfant refusant un parapluie et étant pris sous la pluie, et le troisième est l'image d'un chat coincé dans un arbre.

La section III propose une tâche de **narration d'un conte**. Un livre d'images de poche de Cendrillon est présenté aux participants, le texte étant caché¹⁰. On leur propose de feuilleter le livre afin de se remémorer l'histoire. Ces images illustrent les événements centraux de l'histoire de Cendrillon pour encourager un récit comprenant à la fois une structure narrative et des détails descriptifs. Ensuite, le livre est enlevé et ils sont invités à en dire autant qu'ils peuvent sur l'histoire.

La section IV propose une tâche de **discours procédural** où les participants sont invités à décrire la façon dont ils feraient un sandwich au beurre de cacahuète et à la confiture. Dans le

10 Communication personnelle avec Audrey Holland (octobre 2015), membre d'AphasiaBank

protocole original, il est précisé que les utilisateurs du test situés en dehors des Etats-Unis peuvent le remplacer par une autre préparation avec des aliments simples.

Un script a été développé pour harmoniser les interventions des examinateurs. Ceux-ci doivent rester le plus proche possible du script et veiller à être le plus silencieux possible durant l'administration du test en privilégiant les encouragements non-verbaux. Ce script comprend deux niveaux d'incitations à utiliser si le participant ne répond pas à la première incitation dans un délai de dix secondes.

Le protocole doit être administré en une seule fois et filmé selon des règles précises disponibles sur le site AphasiaBank (Site AphasiaBank).

1.1.2.1.2. Adaptation du protocole destiné aux participants contrôles

Lors de la phase d'adaptation, nous avons essayé de rester le plus fidèle possible au protocole original. Certains éléments n'ayant plus de sens une fois traduits en français de façon « littérale », nous avons cependant dû faire des choix concernant la traduction de certains termes ou phrases afin de proposer un protocole cohérent et compréhensible en français.

Au sein de la **section I** (discours spontané) partie A, « Histoire d'une maladie et récupération », la phrase "*We're really interested in knowing about your language*" présente dans le protocole original a été supprimée car cela n'apportait pas de précision supplémentaire pour la compréhension de la consigne et n'était pas du tout naturelle à l'oral.

Dans la partie B « Événement important », nous avons choisi de traduire « *can you tell me a story about something important that happened to you in your life?* » par « *Pourriez-vous me raconter un événement marquant de votre vie ?* ».

Au sein de la **section II** (Narration d'histoire à partir d'images), nous avons remplacé le titre original « Picture Description » par « Discours narratif à partir d'images ».

Dans la partie B, « Le parapluie », nous avons rajouté les éléments soulignés ci-dessous, pour plus de naturel à l'oral. « Voici d'autres images qui racontent aussi une histoire. Regardez-les toutes et ensuite, comme pour la série d'images précédentes, je vous demanderai de me raconter l'histoire avec un début, un milieu et une fin. Vous pourrez aussi continuer à regarder les images quand vous raconterez l'histoire. »

Au sein de la **section III** (Narration d'un conte), nous avons dû faire un choix concernant le livre présenté aux participants. Le protocole original se base sur le livre d'images publié par Walt Disney en 1993 (Grimes, 1993). Le livre américain n'étant pas facilement accessible en France, nous nous sommes procurés la version originale afin de trouver un équivalent français qui soit proche du livre original et facilement accessible. Après avoir effectué une sélection de livres potentiels et une comparaison de tous les livres sélectionnés, nous avons dû choisir un livre équivalent en français.

Au sein de la **section IV** (Discours procédural), comme précisé dans le protocole original, nous avons fait le choix d'un sandwich culturellement français en choisissant le sandwich jambon beurre. Pour effectuer ce choix, il était nécessaire de trouver un sandwich simple et connu de tous, comme dans la version originale. Ainsi, le sandwich jambon beurre nous a paru être celui le plus approprié pour cette épreuve.

Les instructions du protocole d'épreuves de discours à destination des participants contrôles tel que nous l'avons adapté figure en annexe p. 61-65.

1.1.2.1.3. Description du protocole destiné aux participants aphasiques

Le protocole destiné aux participants aphasiques est identique à celui présenté aux participants contrôles à l'exception de la section I : Les récits personnels sont débutés en demandant au participant de parler de son langage oral, de son AVC et de son rétablissement. Il leur est ensuite demandé, comme dans le protocole destiné aux participants contrôles, de raconter un événement important de leur vie.

L'autre différence réside dans le fait qu'il existe un questionnaire d'aide appelé « *Troubleshooting* », disponible sur le site AphasiaBank. Il est destiné aux participants qui ne peuvent pas répondre au deuxième niveau d'incitation du protocole normal et qui ont besoin d'incitations supplémentaires avec des questions simplifiées. Pour chaque partie, une question supplémentaire est posée afin d'aider le participant à produire une réponse.

Pour l'épreuve de discours procédural, une image-stimulus avec des photographies de beurre de cacahuète, de pain, et de confiture est disponible pour être utilisée avec les participants qui ont besoin d'une aide supplémentaire.

Il est possible de fournir une carte avec OUI/NON au participant si l'atteinte langagière est trop sévère pour qu'il puisse répondre. Si, malgré les sollicitations supplémentaires, le participant n'est pas capable de répondre, l'examineur met fin à la passation. Ce questionnaire d'aide permet ainsi au clinicien de terminer l'entretien de façon plus simple si la personne aphasique a trop de difficulté pour continuer.

1.1.2.1.4. Adaptation du protocole destiné aux participants aphasiques

Comme pour le protocole destiné aux participants contrôles, nous avons traduit le protocole destiné aux participants aphasiques en conservant toutes les épreuves présentes dans la version originale.

Nous avons supprimé la phrase traitant des « Exit Questions » formulée ainsi dans le protocole original « *If the participant does not respond to the Broken Window picture, and you believe these tasks are too difficult, you can use the "Exit Questions" at the end of the Troubleshooting document to end the session or you can discontinue as you see fit* ». En effet,

après confirmation des concepteurs, ces questions ont finalement été supprimées de la version définitive.

Nous avons également adapté au français le questionnaire d'aide (cf annexe p. 71-72), ainsi que la carte OUI/NON (cf annexe p. 73), qui peut être utilisée tout au long de ce questionnaire.

Pour l'aide de l'épreuve du discours procédural, nous avons dû choisir des photos correspondant aux ingrédients nécessaires à la fabrication d'un sandwich jambon beurre car les images originales correspondaient aux ingrédients nécessaires à la fabrication d'un sandwich au beurre de cacahuète. Les images du protocole original n'étant pas disponibles sur le site AphasiaBank, nous avons donc dû faire des choix afin de sélectionner celles qui nous semblaient les plus appropriées. Afin que n'importe quel clinicien puisse se les procurer, nous avons acheté via internet des photos libres de droits. Elles figurent en annexe p. 74-75.

Les instructions du protocole d'épreuves de discours à destination des participants aphasiques tel que nous l'avons adapté figure en annexe p. 66-70.

1.1.2.2. Adaptation des épreuves complémentaires

1.1.2.2.1. Description des épreuves complémentaires originales

Dans le protocole original il est également préconisé d'administrer des tests complémentaires aux patients aphasiques. Nous allons à présent détailler chacun de ces tests, qui figurent, pour certains, sur le site AphasiaBank.

• ***AphasiaBank Repetition Test*** (Holland et al.,2007)

Il s'agit d'un test de répétition non standardisé, développé pour évaluer les capacités de répétition de mots et de phrases, conçu par Holland, MacWhinney, Fromm et Forbes. Ce test évalue la mémoire auditivo-verbale et non l'expression orale.

La **section I** évalue la répétition de mots en proposant des listes de mots de plus en plus longues. La première épreuve comprend jusqu'à huit mots. Le participant doit répéter la liste à chaque fois que celle-ci s'allonge d'un mot supplémentaire. Il ne dispose que d'un essai à chaque ajout de mot.

La passation se poursuit jusqu'à ce que le participant échoue à répéter 2 séries de mots consécutives. On obtient alors un score sur huit qui correspond à la plus longue série répétée par le participant. L'ordre des mots ne compte pas puisque la réponse est comptée comme correcte même s'il y a des changements dans l'ordre des mots.

La seconde épreuve comprend des listes qui contiennent de un à sept mots. Ces listes sont qualifiées « d'ouvertes » car les mots sont différents dans chaque liste. Le participant dispose de trois essais à chaque nouvel item.

La passation est poursuivie tant que le participant parvient à répéter tous les mots, sans tenir

compte de l'ordre, dans au moins deux des trois listes à une longueur donnée.

Pour cette épreuve, on calcule deux empan : un empan est calculé pour l'ordre sériel et un autre pour l'ordre aléatoire. On compte la plus longue liste dans laquelle le patient a répété au moins deux ou trois séquences de mots correctement.

Pour ces deux épreuves, les approximations où le mot-cible est reconnaissable (différent d'un phonème, mais ne formant pas un autre mot), les déformations dues à des troubles dysarthriques ou apraxiques sont considérées comme correctes.

La **section II** de l'*AphasiaBank Repetition Test* évalue la répétition de phrases.

Pour la première épreuve, le patient doit répéter une phrase de longueur croissante. On trouve six phrases en tout, allant de quatre mots pour la première à dix-neuf mots pour la dernière. L'examineur peut répéter une fois la phrase si le participant le demande ou s'il ne répond pas au bout de 10 secondes. On continue la passation jusqu'à ce que le participant échoue à répéter tous les mots ou tous les mots principaux dans l'ordre correct pour deux phrases consécutives.

Le score obtenu correspond à la phrase la plus longue répétée correctement, c'est-à-dire la répétition de tous les mots ou de tous les mots principaux dans le bon ordre. On compte un point pour chaque mot que le patient répète correctement dans chaque phrase.

Pour la deuxième épreuve, on présente une série de douze phrases au patient : quatre sans erreurs, quatre avec des erreurs sémantiques et quatre avec un potentiel effet d'interférence. L'examineur peut répéter une fois chaque item si le participant le demande ou s'il ne répond pas dans les 10 secondes. La passation est interrompue si le participant n'obtient que deux ou moins de deux mots corrects à trois phrases consécutives. Le score est de un point pour chaque mot répété correctement dans chaque phrase.

On retrouve des sous-scores pour les phrases qui contiennent des erreurs sémantiques et avec un effet d'interférence.

Les approximations sont également acceptées tant que la cible est reconnaissable. Les changements morphologiques ne sont pas considérés comme des erreurs, ni l'ajout d'un mot tant que cela ne change pas le sens de la phrase. Les paraphasies sémantiques et les changements dans l'ordre des mots sont considérés comme des erreurs et doivent être mentionnés sur la feuille de cotation.

- **Verb Naming Test** - tiré de *Northwestern Assessment of Verbs and Sentences-Revised*, Field Test Version (Thompson, 2011, cité sur le site AphasiaBank).

Ce test évalue la dénomination de vingt-deux verbes à partir d'images. On compte un point pour chaque réponse correcte. Les sous-scores classent les verbes selon leur structure argumentale, c'est-à-dire selon le nombre d'arguments obligatoires ou optionnels avec lesquels ils

peuvent s'utiliser.¹¹

- **Boston Naming Test** - seconde édition, version courte (Kaplan et al., 2001, cité sur le site AphasiaBank)

Ce test évalue la dénomination de mots à partir d'images. Les items sont uniquement des objets et la cotation se fait en écart-type. Il existe deux versions courtes selon l'âge du patient. La version A comprend trente-quatre images pour les moins de 70 ans et la version C comprend vingt images pour les plus de 70 ans. Il n'est pas disponible sur le site AphasiaBank.

- **Western Aphasia Battery-Revised** – Les subtests de l'Aphasia Quotient uniquement (Kertesz, 2007, cité par MacWhinney et al., 2011)

L'objectif de ces épreuves est de fournir une mesure normalisée du type d'aphasie et de sa gravité grâce à l'Aphasia Quotient (AQ). Ainsi, seules les tâches de la *Western Aphasia Battery* (WAB) liées au langage sont réalisées et non ce qui concerne le QI.

Ce test n'est pas disponible sur le site AphasiaBank.

- **Complex Ideational Material** - Version courte – tiré du *Boston Diagnostic Aphasia Examination* (Goodglass et al., 2001, cité sur le site AphasiaBank)

Il s'agit de questions de logique et de raisonnement auxquelles le participant doit répondre par oui ou par non. Dans la version courte, on trouve six paires de questions. Les deux premières paires sont des questions courtes. Puis l'examineur lit un petit texte à la suite duquel il pose de nouveau deux paires de question à chaque fois. Chaque paire contient une réponse « oui » et une réponse « non ». On compte un point pour chaque paires de questions correctes.

- **Sentence Comprehension** – Adapté de *Philadelphia Comprehension Battery* (Saffran et al, 1988, cité sur le site AphasiaBank)

Il s'agit d'une épreuve de compréhension orale de phrases. Dans cette épreuve, le participant doit désigner l'une des deux images qui correspond à la phrase énoncée oralement par l'examineur. Cette épreuve comporte vingt phrases classées selon leur caractéristiques grammaticales. On retrouve les catégories suivantes : ORC (Object relative clause), SRC (Subject Relative clause), LOC (phrase locative), ACT(phrase active), PAS (phrase passive). On trouve deux phrases de chaque catégorie, chaque phrase étant qualifiée soit de « *lexicale* » (L) soit de

11 En syntaxe, le terme d'argument est généralement utilisé pour regrouper les fonctions syntaxiques de sujet et de complément (complément direct, attribut, complément indirect,...) (Site Sémanticlopédie). Un verbe intransitif est un verbe à un argument, un verbe transitif est un verbe à deux arguments, un verbe ditransitif est un verbe à trois arguments (Site Agre-ink).

« réversible » (R). On retrouve alors 10 phrases lexicales et 10 phrases réversibles.¹²

Concernant tous ces tests, le site AphasiaBank précise que l'*AphasiaBank Repetition Test*, le *Verb Naming Test* et le *Boston Naming Test* doivent être enregistrés. Il n'est cependant pas nécessaire de retranscrire les réponses de ces tests. Il est également précisé de ne faire passer le *Boston Naming Test* et la *Western Aphasia Battery* que si l'examineur ne dispose pas de scores datant de moins de 6 mois.

Tous les résultats aux tests doivent être inscrits dans le tableau excel intitulé « Test Results.xls » et sont disponibles pour les membres AphasiaBank dans un tableur sur le site AphasiaBank.

1.1.2.2.2. Adaptation des épreuves équivalentes : prise en compte du contexte clinique en France

La version française d'un certain nombre de tests utilisés dans la version anglo-saxonne du protocole n'existant pas, il a été nécessaire de proposer des épreuves équivalentes en français.

L'objectif initial était de rester fidèle au protocole d'origine dans un souci d'uniformisation des données recueillies et présentes dans la base de données AphasiaBank.

Cependant, nous n'avons pas traduit directement les tests utilisés dans la version originale car traduire un test standardisé et validé dans une autre langue n'aurait pas de validité et présenterait de ce fait, peu d'intérêt.

De plus, il était nécessaire d'adapter les tests à la réalité clinique française. Nos choix ont donc été également motivés par le fait qu'il fallait proposer, tant que cela était possible, des épreuves déjà connues et utilisées par les cliniciens et les orthophonistes français, dans un souci de facilité d'utilisation, d'accessibilité et de connaissance des tests proposés.

• **Adaptation de l'*AphasiaBank Repetition Test*** (Holland et al., 2007)

Il ne nous a pas paru opportun de simplement traduire l'*AphasiaBank Repetition Test*.

Après réflexion, nous avons finalement choisi l'**épreuve de répétition de mots et de phrases** du *HDAE* (Mazaux et Orgogozo, 1981). Si cette épreuve évalue la mémoire auditivo-verbale, elle ne calcule pas d'empan comme c'est le cas dans l'épreuve originale.

Comme dans le test proposé dans le protocole AphasiaBank, les difficultés articulatoires ne sont pénalisées que si le mot n'est plus reconnaissable.

Cependant, plusieurs éléments diffèrent par rapport au test original. Tout d'abord le nombre

12 La notion de « lexical » et « réversible » renvoie à la nature des distracteurs proposés sur les images. La phrase est qualifiée de « réversible » lorsque le distracteur représente une image comportant une inversion entre les deux agents impliqués dans l'action. Si le distracteur n'est pas réversible, la phrase est alors qualifiée de « lexicale ». (communication personnelle avec Margie Forbes, mail du 07/06/16).

d'items n'est pas le même car l'épreuve de répétition du protocole AphasiaBank comprend huit mots pour la première liste et des listes allant jusqu'à sept mots pour la seconde. La façon dont les items sont présentés diffère également. En effet, ils sont présentés un à un dans le test français et par liste croissante dans le test anglais.

Concernant l'épreuve de répétition de phrases, il n'existe pas de phrase avec effet potentiel d'interférence dans le test français, ni de phrases avec des erreurs sémantiques, comme c'est le cas dans le test original.

A la différence du test de répétition de phrases d'AphasiaBank, les sous-scores de l'épreuve de répétition du *HDAE* ne prennent pas non plus en compte la répétition des « mots principaux » de la phrase.

Nous avons également proposé une **épreuve d'empan numérique** issue de la *Batterie d'Efficiences Mnésiques - BEM 144* (Signoret, 1991). L'idéal, pour rester proche du test original, aurait été de trouver une épreuve de span de mots et de phrases, mais nous n'avons pas trouvé d'épreuve de ce type lorsque nous avons effectué le travail d'adaptation.

Comme dans le test original, on calcule la série la plus longue répétée correctement. Cependant, à la différence du test original, ce sont des chiffres qui sont répétés et non des mots et cette épreuve calcule un empan de chiffres dans l'ordre mais pas dans un ordre aléatoire, comme c'est le cas dans la version originale.

- **Adaptation du *Verb Naming Test*** - tiré de *Northwestern Assessment of Verbs and Sentences-Revised*, Field Test Version (Thompson, 2011, cité sur le site AphasiaBank).

Il existe plusieurs tests proposant des dénominations de verbes. Le *MT86* et le *HDAE* proposent une dénomination de six verbes d'action. Nous n'avons cependant pas choisi ces tests car le nombre d'items était trop réduit par rapport à la version originale.

Nous avons ainsi choisi, le test de *Dénomination de Verbes Lexicaux (DVL 38)* (Hammelrath, 1999). Celui-ci comporte 38 verbes alors que le *Verb Naming Test (VNT)* n'en compte que 22. Les verbes sont classés selon leur fréquence lexicale (haute, moyenne et basse fréquence) et non selon leur structure argumentale, comme c'est le cas dans le *VNT*.

Par rapport au test original nous n'aurons donc pas de sous-scores relatifs à la structure argumentale des verbes.

- **Adaptation du *Boston Naming test*** - seconde édition, version courte (Kaplan et al., 2001, cité sur le site AphasiaBank)

Le test proposé dans le protocole AphasiaBank, le *Boston Naming Test (BNT)*, existe dans une version adaptée au français. Nous avons pu nous procurer la version longue et non la version courte adaptée au français de ce test. La version longue française comporte soixante items, ce qui

nous a paru trop élevé par rapport à ce qui était proposé dans le test original.

Les batteries plus spécifiques telles que la *DO80* (Deloche et Hannequin, 2007) ou *Lexis* (De Partz et al., 2002) nous ont également paru trop longues par rapport au test proposé dans la version originale.

L'épreuve la plus proche concernant le type et le nombre d'items était donc, selon nous, l'épreuve de dénomination du *MT86* (Nespoulous et al., 1986). Cependant, à la différence du test original, la cotation ne se fait pas en écart-type mais avec un simple score.

Nous avons choisi de ne pas garder la dénomination de verbes car cela ne fait pas partie du protocole initial, qui ne propose de ne dénommer que des objets, et qu'il est par ailleurs proposé une épreuve de dénomination de verbes plus spécifique.

- **Adaptation de l'*Aphasia Quotient*** - Tiré de la *Western Aphasia Battery-Revised* (Kertesz, 2007, cité par MacWhinney et al., 2011)

Cette batterie n'existe pas en Français et nous n'avons pas pu nous procurer la version anglaise. Il ne semble par ailleurs pas exister d'équivalent de l'*Aphasia Quotient* en français.

Le *HDAE* (Mazaux et Orgogozo, 1981) propose une échelle de gravité de l'aphasie, qui peut être remplie grâce aux épreuves de conversation et de langage spontané. Nous avons donc choisi de nous baser sur cette échelle. Cependant, certaines questions de l'épreuve de conversation et langage spontané du *HDAE* étant identiques à ce que l'on propose dans l'épreuve de discours EVAL, nous avons décidé d'utiliser cette échelle mais de la coter à partir de nos épreuves de conversation et de langage spontané.

Il n'existe pas non plus d'équivalent français pour la classification du type d'aphasie que la *WAB* propose. Après avoir demandé confirmation aux concepteurs d'EVAL, nous avons décidé de ne nous référer au diagnostic clinique du thérapeute, car c'est ce qui est pratiqué en France.

- **Adaptation du *Complex Ideational Material*** - Version courte – tiré du *Boston Diagnostic Aphasia Examination* (Goodglass et al., 2001, cité sur le site *AphasiaBank*)

L'épreuve originale, « *Complex Ideational Material* », existe dans la traduction française du *BDAE*. Il s'agit du subtest logique et raisonnement (opérations intellectuelles complexes) du *HDAE* (Mazaux et Orgogozo, 1981).

La version française comprend les mêmes questions et petits textes que l'épreuve originale avec, en plus, quatre questions courtes et deux autres petits textes. Elle est donc composée de douze paires de questions, alors que la version anglaise en comporte six. Pour rester au plus proche du protocole original, nous avons utilisé ce subtest en ne reprenant que les six paires de questions de l'épreuve originale en langue anglaise. Nous avons choisi de les énoncer dans le même ordre que celui proposé dans l'épreuve originale. On obtient alors un score mais il n'est pas

possible de calculer un écart-type.

- **Adaptation du Sentence Comprehension** – Adapté de *Philadelphia Comprehension Battery* (Saffran et al, 1988, cité sur le site AphasiaBank)

L'*HDAE* ne proposant pas d'épreuve semblable, nous avons choisi d'utiliser l'épreuve de compréhension de mots et de phrases du *MT86* (Nespoulous et al., 1986).

La question de la passation de l'épreuve dans sa globalité s'est posée. Celle-ci comporte trente-huit items en tout avec neuf items de compréhension de mots, ce qui nous a semblé long dans un premier temps.

Afin de respecter la progression proposée de l'épreuve du *MT86* et de proposer le même nombre d'items que l'épreuve du *Sentence Comprehension*, nous avons finalement choisi de ne proposer que les vingt premiers items de la compréhension de phrases du *MT 86*.

Nous avons veillé à ce qu'il existe bien deux items correspondants à chaque type de phrase évalué dans le test anglais original. Le test comporte donc bien deux phrases relatives objet, deux phrases relatives sujet, deux phrases locatives et deux phrases passives. Cependant, le nombre de phrases actives diffère car nous avons choisi de compter toutes les phrases restantes comme des phrases actives. On aboutit alors à un total de douze phrases actives au sein de notre épreuve, alors que le test original en compte seulement quatre.

Seule la distinction phrase lexicale ou réversible, n'a pas été prise en compte dans ce test, car nous n'avons pas trouvé d'équivalent français.

Nous avons également choisi de proposer aux participants l'épreuve de compréhension orale de mots, qui n'est pas présente dans le test original, afin de vérifier que la compréhension de phrases ne soit pas perturbée par la compréhension lexicale.

La liste des épreuves complémentaires du protocole AphasiaBank adapté à la langue française figure en annexe p. 76.

1.1.2.2.3. Adaptation du tableau de résultats aux épreuves complémentaires

Le tableau original, disponible sur le site AphasiaBank, reprend tous les scores aux différentes épreuves proposées en anglais.

Les épreuves complémentaires proposées n'étant pas les mêmes que dans le protocole original nous avons dû modifier l'ensemble du tableau pour que cela corresponde aux épreuves que nous avons proposées, en veillant à garder la même mise en page. Il a donc fallu choisir, pour chaque épreuve, le nombre de sous-scores et de colonnes de résultats à entrer dans le tableau.

Concernant l'épreuve de dénomination du *MT86* nous avons choisi de mettre les sept sous-scores suivants :

– Score sans Hors Temps : total de réponses produites dans un délai de cinq secondes

- Nombre de mots produits Hors Temps : dans un délai supérieur à cinq secondes
- Nombre de mots produits avec facilitation formelle ou contextuelle
- Nombre de circonlocutions
- Nombre de néologismes
- Nombre de gestes référentiels adéquats

Pour l'épreuve de compréhension de phrases du *MT86* nous avons choisi de mettre quatre colonnes pour calculer les sous-scores relatifs aux phrases locatives, passives, relatives objet et relatives sujet même si cela ne correspond pas exactement à la description du test donnée dans le *MT86*.

Nous avons choisi de rajouter une colonne pour les phrases actives et de compter un total sur douze. Ceci implique que pour le calcul de ce score, les phrases qui ne sont pas seulement actives, mais qui évaluent un autre paramètre (*phrases locatives, relatives,...*) sont exclues.

Le tableau de résultats aux épreuves complémentaires que nous avons adapté figure en annexe p. 77-80.

1.2. Résultats concernant l'adaptation du protocole de recueil de données AphasiaBank

Nous allons maintenant présenter les résultats obtenus suite à l'adaptation des différents documents du protocole AphasiaBank que nous venons de présenter. Nous exposerons tout d'abord les résultats concernant l'adaptation des documents permettant le recueil de données démographiques puis ceux concernant l'adaptation des documents permettant le recueil de données langagières.

Tous les documents tels que nous les avons adaptés sont désormais disponible en ligne sur le site AphasiaBank depuis le 13 mai 2016.

1.2.1. Résultats concernant l'adaptation des documents permettant le recueil de données démographiques

1.2.1.1 Résultats concernant l'adaptation du formulaire destiné aux participants

Le formulaire de recueil de données démographiques destiné aux participants contrôles, comme celui destiné aux participants aphasiques a pu être traduit en entier.

Pour adapter ces formulaires, nous avons essayé de rester le plus proche possible des formulaires originaux. Nous avons cependant dû modifier ou supprimer certaines formulations qui n'étaient plus adaptées une fois traduites en français ou encore rajouter certains éléments afin d'apporter plus de précision dans le recueil des données.

Les formulaires tels que nous les avons adaptés figurent en annexe p. 40-42 (pour les participants contrôles) et p. 43-45 (pour les participants aphasiques).

1.2.1.2. Résultats concernant l'adaptation des tableaux de recueil de données démographiques de la base de données AphasiaBank

Nous avons pu adapter le tableau de recueil de données démographiques des participants contrôles sans y apporter de modification mis à part l'intitulé de la colonne « Race ».

Pour le tableau de recueil de données démographiques destiné aux participants aphasiques nous avons dû faire davantage de modifications, toujours dans un souci de précision du recueil de données démographiques.

Les deux tableaux tels que nous les avons adaptés et conçus figurent en annexe p. 46-49 (pour les participants contrôles) et p. 50-58 (pour les participants aphasiques).

1.2.2 Résultats concernant l'adaptation des épreuves de discours AphasiaBank

1.2.2.1 Résultats concernant l'adaptation des épreuves de discours destinées aux participants contrôles

Nous avons pu effectuer une adaptation en conservant toutes les épreuves présentes dans la version anglo-saxonne.

Le travail d'adaptation a cependant nécessité la modification de quelques formulations par rapport à la version originale.

Concernant l'épreuve de narration d'un conte, à partir du livre de Cendrillon, nous avons pu proposer un livre équivalent en français. Nous avons choisi le livre d'images Disney, édité chez Hachette Jeunesse (Cendrillon, 2014). Notre choix s'est orienté vers celui-ci pour plusieurs raisons. Au niveau du nombre de pages, l'original en comporte 25 et celui que nous avons utilisé en comporte 32. Si les images ne sont pas toutes les mêmes, certaines sont identiques et le graphisme est similaire. Il s'agit d'un livre en couleur comme dans la version originale et de même format, souple et de poche. Tout comme dans le protocole original, nous avons choisi de cacher le texte du livre « Cendrillon » français avec des étiquettes blanches autocollantes.

D'autre part, nous avons dû modifier l'intitulé de l'épreuve de discours procédural en remplaçant le sandwich au beurre de cacahuète par le sandwich jambon beurre.

De plus, nous avons choisi de ne pas adapter les tests optionnels destinés aux participants contrôles car la passation de ceux-ci était facultative et que nous avons préféré vérifier l'absence des critères d'exclusion par le biais des formulaires d'information.

Le script du protocole tel que nous l'avons adapté et proposé aux participants contrôles est désormais disponible sur le site AphasiaBank et joint en annexe p. 61-65.

1.2.2.2 Résultats concernant l'adaptation des épreuves de discours destinées aux participants aphasiques

Nous avons pu adapter les épreuves de discours destinées aux participants aphasiques sans faire de modification par rapport à la version originale, à l'exception des modifications déjà citées précédemment dans la traduction du script destiné aux participants contrôles. Nous avons utilisé le même livre pour la narration de Cendrillon.

Concernant le choix des images du sandwich nous avons acheté les photos sur le site Fotolia afin que celles-ci soient utilisables par tous car libres de droit. Il s'agit de photos en couleur décontextualisées sur fond blanc, présentées sur la même page. Vous trouverez ces images en annexe p. 74-75.

La carte OUI/NON que nous avons créée est disponible en annexe p. 73.

Le script du protocole tel que nous l'avons traduit et proposé aux participants aphasiques a également été mis en ligne sur le site AphasiaBank et joint en annexe p. 66-70.

Le questionnaire d'aide tel que nous l'avons traduit est également disponible en annexe p. 71-72.

1.2.2.3 Résultats concernant l'adaptation des épreuves complémentaires

Voici le tableau récapitulatif des épreuves équivalentes choisies :

TESTS ORIGINAUX	EPREUVES EQUIVALENTES PROPOSEES
AphasiaBank Repetition Test	HDAE : Répétition de mots et de phrases BEM144 : Span auditivo-verbal
Verb Naming Test	DVL-38 : Dénomination de verbes
Boston Naming Test	MT86 : Dénomination de noms
<i>Western aphasia Battery – Revised (2007)</i>) les subtests de l'aphasia quotient uniquement	HDAE : Echelle de gravité de l'aphasie (pour la détermination du type d'aphasie, nous nous référons au diagnostic clinique)
Complex ideational material	HDAE : Subtest logique et raisonnement (opérations intellectuelles complexes)
Sentence comprehension	MT86 : Compréhension orale de mots et de phrases

Tableau n°6 : Adaptation des épreuves complémentaires du protocole AphasiaBank en français

- **Résultats concernant l'adaptation de l'*AphasiaBank Repetition Test*** (Holland et al.,2007)

Nous avons choisi d'utiliser l'épreuve de répétition de mots et de phrases du *HDAE* (Mazaux et Orgogozo, 1981).

L'épreuve de répétition de mots comporte vingt items avec un demi point accordé par bonne répétition. On obtient alors un score de 0 à 10.

L'épreuve de répétition de phrases comporte seize items comprenant huit phrases concrètes

et huit phrases abstraites. On attribue un score sur 8 pour les phrases concrètes et un score sur 8 pour les phrases abstraites.

Nous avons également choisi de proposer une épreuve d'empan numérique tirée de la *BEM 144* (Signoret, 1991). Il s'agit d'une classique épreuve de répétition immédiate de séries de chiffres. Le sujet est invité à « répéter dans l'ordre » les séries de chiffres qui vont être énoncées par l'examineur. La feuille de notation comporte des séquences croissantes de chiffres. On donne toujours les deux séries d'une longueur donnée, même si la première a été réussie. L'arrêt de l'épreuve se fait après un échec à deux séries d'une même longueur. La notation s'effectue en notant la série la plus longue répétée correctement même une seule fois.

- **Résultats concernant l'adaptation du *Verb Naming Test*** - tiré de *Northwestern Assessment of Verbs and Sentences-Revised*, Field Test Version (Thompson, 2011, cité sur le site AphasiaBank)

Pour adapter le *Verb Naming Test* (VNT) proposé dans la version originale, nous avons choisi de faire passer le *test de Dénomination de Verbes Lexicaux (DVL-38)* (Hammelrath, 1999). Il s'agit de dénommer 38 verbes. On obtient un score total sur 114 au test et des sous-scores pour les verbes de haute, moyenne et basse fréquence.

- **Résultats concernant l'adaptation du *Boston Naming Test*** - seconde édition, version courte (Kaplan et al., 2001, cité sur le site AphasiaBank)

Pour adapter cette épreuve nous avons choisi l'épreuve de dénomination de noms du *MT86*. (Nespoulous et al., 1986).

Cette épreuve comporte vingt-six items dont vingt noms communs, avec les catégories fruits, meubles et outils. La notation se fait par un score sur vingt-cinq qui correspond au nombre de réponses attendues produites dans un délai maximal de cinq secondes, avec ou sans déviations phonétiques ou phonémiques. On trouve ensuite des sous-scores relatifs au nombre de bonnes réponses au-delà de cinq secondes, au nombre de déviations verbales, de circonlocutions, de néologismes ou de gestes référentiels adéquats produits par le patient. Le test propose également de comptabiliser le nombre de fois où la facilitation formelle ou contextuelle a permis au patient de produire la réponse attendue.

- **Résultats concernant l'adaptation de l'*Aphasia Quotient*** - tiré de la *Western Aphasia Battery-Revised* – Les subtests de l'Aphasia Quotient uniquement (Kertesz, 2007, cité par MacWhinney et al., 2011)

Pour déterminer la sévérité de l'aphasie nous avons choisi de nous référer à l'échelle de gravité de l'aphasie du *HDAE* (Mazaux et Orgogozo, 1981). Le score est sur sept et évalué à partir

de l'épreuve de conversation et de langage spontané. Il s'agit de faire parler le patient pendant dix minutes en lui posant des questions. Nous avons cependant décidé de coter à partir de nos épreuves de conversation et de langage spontané.

Pour la détermination du type d'aphasie, nous avons choisi de nous référer au diagnostic clinique du thérapeute.

- **Résultats concernant l'adaptation du *Complex Ideational Material*** - Version courte – tiré du *Boston Diagnostic Aphasia Examination* (Goodglass et al., 2001, cité sur le site AphasiaBank)

Nous avons choisi le subtest logique et raisonnement (opérations intellectuelles complexes) du *HDAE* (Mazaux et Orgogozo, 1981). Il s'agit de la même épreuve que celle proposée dans le protocole anglo-saxon.

- **Résultats concernant l'adaptation du *Sentence Comprehension*** – Adapté de *Philadelphia Comprehension Battery* (Saffran et al, 1988, cité sur le site AphasiaBank).

Nous avons choisi de faire passer l'épreuve de compréhension orale de mots et de phrases du *MT86* (Nespoulous et al., 1986) en ne gardant que les 20 premiers items de l'épreuve de compréhension de phrases.

On évalue tout d'abord la compréhension des mots. On propose la désignation de cinq noms présentés sur des planches comportant six images avec un distracteur visuel, un distracteur phonologique, un distracteur sémantique et deux distracteurs qui ne sont pas reliés sémantiquement, visuellement ou phonologiquement à la cible. On trouve ensuite la désignation de quatre substantifs utiles pour la passation de l'épreuve de compréhension de phrases. Enfin, l'épreuve de compréhension de phrases teste la compréhension des prépositions locatives, des propositions relatives à « sujet clivé », des propositions relatives à « objet clivé », des phrases actives et passives. Les planches proposées au patient comportent quatre ou deux images. On obtient alors un score correspondant au total de phrases correctement désignées.

La liste des épreuves complémentaires du protocole AphasiaBank adapté à la langue française figure en annexe p. 76.

1.2.2.4. Résultats concernant l'adaptation du tableau de recueil des résultats aux épreuves complémentaires

Nous avons pu adapter le tableau de recueil des résultats dans son ensemble. Celui-ci diffère totalement du tableau original, puisque les épreuves que nous avons proposées dans l'adaptation n'étaient pas les mêmes que dans le protocole original. Le tableau que nous avons adapté figure en annexe p. 77-80.

2. POPULATION

2.1. Critères de sélection des sujets

2.1.1. Critères de sélection de la population contrôle

Afin de pouvoir participer à notre protocole destiné à initier la création d'une base de données de référence en langue française, les participants contrôles devaient répondre aux critères suivants :

- Pas d'antécédent ou de pathologie d'ordre neurologique (AVC, traumatisme crânien, etc.);
- Audition (corrigée ou non) suffisante pour effectuer la passation du protocole;
- Vue (corrigée ou non) suffisante pour effectuer la passation du protocole;
- Langue française parlée couramment ;
- Pas de pathologie entraînant une détérioration cognitive telle que les maladies d'Alzheimer ou de Parkinson ;
- Pas d'état dépressif au moment de la passation des épreuves du protocole;
- Pas d'antécédent de trouble(s) de la parole, du langage, ou de la communication.

Aux six premiers critères explicitement définis dans les instructions du protocole AphasiaBank, nous avons décidé, en concertation avec les concepteurs de la base de données, d'y ajouter l'absence d'antécédent de trouble(s) de la parole, du langage et de la communication, afin de pouvoir obtenir des données langagières reflétant le discours « normal », sans processus pathologique sous-jacent.

La présence de ces critères a été vérifiée par l'examineur qui a fait remplir à chaque participant un formulaire détaillé destiné à recueillir des données démographiques et médicales. Un exemplaire de ce formulaire figure en annexe p. 40-42. Afin d'évaluer ces critères de façon plus objective, il est possible d'administrer les tests suivants¹³ :

- Un test de dix phrases issues du CID Everyday Speech Sentences (Davis and Silverman, 1978) pour dépister les troubles de l'audition ;
- Une épreuve issue de *Augmentative and Alternative Communication* (Beukelman and Mirenda, 1998) pour dépister les troubles visuels ;
- La version courte de la Geriatric Depression Scale (Brink et al., 1982) pour évaluer l'état dépressif ;
- Le Mini-Mental Status Examination (Folstein, Folstein, & Fanjiang, 2002) pour l'évaluation des facultés cognitives (SiteAphasiaBank).

La passation de ces tests est facultative et le contrôle des critères d'inclusion peut être fait de façon purement clinique. Nous n'avons pas utilisé ces tests dans notre processus de sélection de

¹³ Seuls les trois premiers tests sont disponibles sur le site AphasiaBank.

la population contrôle. Les discussions ainsi que les formulaires que nous avons fait remplir aux participants nous ont donné suffisamment d'informations pour déterminer si les participants répondaient ou non à nos critères d'inclusion.

Les sujets contrôles ont été recrutés dans les départements de l'Hérault, de la Dordogne, des Côtes-d'Armor et du Morbihan.

2.1.2. Critères de sélection de la population aphasique

Afin de pouvoir participer à notre protocole destiné à initier la création d'une base de données française, les participants aphasiques devaient répondre aux critères suivants :

- Aphasie résultant d'un accident vasculaire cérébral de l'hémisphère gauche ;
- Absence de démence ou de comorbidités associées à des perturbations cognitives importantes ;
- Audition (corrigée ou non) suffisante pour effectuer la passation du protocole ;
- Vue (corrigée ou non) suffisante pour effectuer la passation du protocole ;
- Langue française parlée couramment ;
- Pas d'état dépressif au moment de la passation des épreuves du protocole ;
- Pas d'antécédent de trouble(s) de la parole, du langage, ou de la communication.

Aux deux premiers critères explicitement définis dans les instructions du protocole AphasiaBank, nous avons décidé, en concertation avec les concepteurs de la base de données, d'y ajouter l'absence d'antécédent de trouble(s) de la parole, du langage et de la communication, afin de pouvoir obtenir des données langagières reflétant les processus pathologiques dus à l'aphasie et non à d'autres pathologies de la parole, du langage ou de la communication.

Nous avons également ajouté des critères relatifs à l'audition, la vision, l'absence de dépression et la langue française parlée couramment afin de s'assurer que les performances langagières des participants ne soient pas altérées par des facteurs autres que l'aphasie. De plus, pour limiter les biais dans la comparaison entre les résultats de la population contrôle et de la population aphasique, il nous a semblé important de sélectionner les participants de façon à minimiser le plus possible les différences autres que la présence ou non d'une aphasie.

La présence de ces critères a été vérifiée par l'examineur qui a fait remplir à chaque participant le formulaire de recueil de données démographiques (cf annexe p. 43-45). Selon les cas, ce formulaire a été rempli par le participant lorsque celui-ci était en mesure de le faire par lui-même, par l'aidant principal du participant, par son orthophoniste ou par l'examineur à travers une discussion avec le participant, son aidant principal ou son orthophoniste. La survenue d'un AVC à l'origine de l'aphasie a été attestée par des examens de neuro-imagerie ou par un diagnostic médical précis.

Deux participants ont été exclus de notre protocole car ils ne répondaient pas aux critères

d'inclusion précédemment définis. Le premier présentait un antécédent de trouble du langage avant la survenue de l'AVC ainsi qu'un état dépressif. Le second était devenu aphasique à la suite d'un traumatisme crânien et non d'un accident vasculaire cérébral.

Les sujets aphasiques ont été recrutés par l'intermédiaire d'orthophonistes exerçant en libéral dans les départements de la Haute-Garonne et de l'Hérault.

2. 2. Description de l'échantillon de population

2.2.1. Description de la population contrôle

La population contrôle est constituée de 14 sujets. Tous ces sujets répondent aux critères d'inclusion mentionnés ci-dessus.

Les principales données démographiques concernant notre population de sujets contrôles sont les suivantes :

- Age moyen de la population : 61 ans et 9 mois
- Répartition selon le sexe : 43 % d'hommes, 57 % de femmes
- Répartition selon la latéralité : 86% de droitiers, 7% de gauchers, 7 % d'ambidextres
- Niveau d'étude moyen : 14 ans 6 mois
- Répartition selon le statut professionnel : 71,5% de personnes retraitées ou sans activité, 28,5% sujets actifs
- Répartition selon les langues parlées : 71,5 % de monolingues, 7 % de polyglottes, 21,5 % de bilingues tardifs
- Tous les sujets utilisent le français comme langue principale.

Tous les sujets sont de nationalité française bien que deux d'entre eux (14 % de l'échantillon) soient nés dans un autre pays. Ces deux sujets ont cependant évolué dans un milieu francophone et sont venus vivre en France avant l'âge de 3 ans. Leur niveau de maîtrise de la langue française est donc équivalent à celui des sujets qui sont nés en France.

ID du participant	Age au moment du test	Sexe	Latéralité	Années d'études	Statut professionnel	Pays de naissance	Nombre d'années vécues en France	Langue(s) parlée(s)
colin01a	47 ans 3 mois	M	G	16	T	FR	NA	POL
colin02a	63 ans	M	D	26	T	FR	NA	MON
colin03a	64 ans 4 mois	F	D	14	R	FR	NA	MON
colin04a	67 ans 9 mois	F	A	16	R	SN	65	BTA
colin05a	39 ans 10 mois	F	D	14	R	FR	NA	BTA
colin06a	75 ans	M	D	3	R	FR	NA	MON
colin07a	67 ans 8 mois	F	D	20	R	FR	NA	MON
lumeur01a	58 ans 3 mois	M	D	15	R	DE	57	MON
lumeur02a	52 ans 4 mois	M	D	12	T	FR	NA	MON
lumeur03a	60 ans 2 mois	F	D	14,5	R	FR	NA	MON
lumeur04a	62 ans 10 mois	F	D	14	R	FR	NA	MON
lumeur05a	84 ans 7 mois	M	D	8	R	FR	NA	MON
lumeur06a	75 ans 10 mois	F	D	14	R	FR	NA	MON
lumeur07a	45 ans 6 mois	F	D	17	T	FR	NA	BTA

Tableau n°7 : Description de la population contrôlée

M : Masculin – F : Féminin

G : Gaucher – D : Droitier – A : Ambidextre

Les années d'études sont exprimées en années effectuées à partir du CP.

T : Travailleur - R : Retraité ou sans activité

FR : France - SN : Sénégal - DE : Allemagne

NA : Non applicable

MON : Monolingue (une seule langue parlée) - BTA : Bilinguisme tardif (deux langues parlées, la seconde langue ayant été apprise après l'âge de 6 ans) - POL : Polyglotte (trois langues parlées ou plus)

2.2.2. Description de la population aphasique

La population aphasique est constituée de 11 sujets. Tous ces sujets répondent aux critères d'inclusion mentionnés ci-dessus.

Les principales **données démographiques générales** concernant notre population de sujets aphasiques sont les suivantes :

- Age moyen de la population : 72 ans et 6 mois
- Répartition selon le sexe : 64 % d'hommes, 36 % de femmes
- Répartition selon la latéralité : 64 % de droitiers, 27 % de gauchers, 9 % d'ambidextres
- Niveau d'étude moyen : 10 ans 3 mois
- Répartition selon le statut professionnel : 82 % de personnes retraitées ou sans activité, 18 % sujets actifs
- Répartition selon les langues parlées : 91 % de monolingues, 9 % de bilingues depuis l'enfance

- Tous les sujets utilisent le français comme langue principale
- Tous les sujets sont de nationalité française et sont nés en France

Les principales **données démographiques médicales** concernant notre population de sujets aphasiques sont les suivantes :

- Délai moyen post-AVC : 3 ans
- Répartition selon le type d'aphasie¹⁴ : 55 % de sujets fluents, 45 % de sujets non fluents
 - ° Aphasies fluentes : aphasie anomique (27,5%), aphasie de conduction (9%), aphasie de Wernicke (9%)
 - ° Aphasies non fluentes : aphasie globale (27,5%), aphasie de Broca (9%)
 - ° Type d'aphasie non précisé (18%)
- Troubles associés : apraxie de la parole (9 %), dysarthrie (9%)¹⁵
- Durée moyenne de la rééducation orthophonique : 2 ans 11 mois. Un des sujets (colin13a) ne suit plus de rééducation orthophonique, il avait mis fin à son suivi orthophonique 4 mois avant que nous le rencontrions pour la passation du protocole
- Répartition selon l'étiologie de l'AVC : ischémique (54,5 %), hémorragique (27,5 %), non renseigné (18%)
- Un des sujets (colin10a) présente une autre pathologie neurologique. Il est atteint de la maladie de Parkinson. Nous avons pu l'intégrer à notre protocole car cette maladie n'est pas associée à des troubles cognitifs importants.

14 Le type d'aphasie du patient a été déterminé à partir du bilan orthophonique et du dossier médical du patient, d'après la classification décrite p. 25.

15 Il est spécifié explicitement dans les critères d'inclusion de la base de données AphasiaBank que les participants présentant une apraxie de la parole et/ou une dysarthrie peuvent être inclus dans la base de données.

ID du participant	Age au moment du test	Sexe	Latéralité	Années d'études	Statut professionnel	Langue(s) parlée(s)
colin09a	48 ans 2 mois	M	G	11	T	MON
colin10a	80 ans 8 mois	M	A	5	R	BDE
colin11a	73 ans 1 mois	M	D	9	R	MON
colin12a	85 ans 1 mois	F	D	5	R	MON
colin13a	60 ans 10 mois	M	D	18	R	MON
colin14a	78 ans 7 mois	F	G	9	R	MON
colin15a	53 ans 6 mois	M	D	17	T	MON
lemeur09a	82 ans 6 mois	M	D	8	R	MON
lemeur10a	77 ans 10 mois	M	D	14	R	MON
lemeur11a	65 ans 3 mois	F	G	8	R	MON
colin16a	91 ans 9 mois	F	D	9	R	MON

Tableau n°8: Description de la population aphasique – Données générales

L'âge au moment du test est exprimé en années,mois (mois calculés des fractions d'une année, par exemple : 1 an et 6 mois = 1.5)

M : Masculin – F : Féminin

G : Gaucher – D : Droitier – A : Ambidextre

Les années d'études sont exprimées en années effectuées à partir du CP.

T : Travailleur - R : Retraité ou sans activité

MON : Monolingue (une seule langue parlée) – BDE : Bilingue depuis l'enfance (deux langues parlées, la second langue ayant été apprise avant l'âge de 6 ans)

ID du participant	Durée de l'aphasie	Type d'aphasie – impression clinique	Type d'aphasie – Classification Boston	Apraxie de la parole	Dysarthrie	Nombre d'années de rééducation orthophonique	Étiologie de la lésion	Antécédent ou présence d'autre pathologie neurologique
colin09a	4 ans 6 mois	NFL	BRO	N	N	4 ans 6 mois	ISC	N
colin10a	3 ans 1 mois	NFL	GLO	N	N	3 ans	ISC	Maladie de Parkinson
colin11a	2 ans 6 mois	NFL	GLO	O	N	2 ans 6 mois	ISC	N
colin12a	2 mois	FLU	ANO	N	N	1 mois	ISC	N
colin13a	1 an 4 mois	FLU	ANO	N	N	10 mois	HEM	N
colin14a	1 an 1 mois	FLU	ANO	N	N	1 an	HEM	N
colin15a	8 ans 7 mois	NFL	GLO	N	N	8 ans 7 mois	ISC	N
lemeur09a	4 ans 6 mois	NFL	NR	N	O	4 ans 6 mois	NR	N
lemeur10a	6 mois	FLU	NR	N	N	6 mois	NR	N
lemeur11a	2 ans	FLU	WER	N	N	2 ans	HEM	N
colin16a	4 ans 9 mois	FLU	CON	N	N	4 ans 9 mois	ISC	N

Tableau n°9: Description de la population aphasique – Données médicales

La durée de l'aphasie est exprimée en années,mois (mois calculés des fractions d'une année, par exemple : 1 an et 6 mois = 1.5)

NFL : Non fluent – FLU : Fluent

NR : Non Renseigné

BRO : Aphasie de Broca – GLO : Aphasie globale – ANO : Aphasie anomique – WER : Wernicke - CON : Aphasie de conduction

N : Non – O : Oui ISC : Lésion ischémique – HEM : Lésion hémorragique NA : Non applicable

Les tableaux complets détaillant l'ensemble des caractéristiques démographiques des participants contrôles et aphasiques figurent en annexe p. 46-49 (population contrôle) et p. 50-58 (population aphasique).

3. LE RECUEIL DE DONNÉES

Nous présenterons tout d'abord la méthodologie employée pour le recueil de données pour présenter ensuite les résultats obtenus suite au recueil de ces données.

3.1. Méthodologie dédiée au recueil de données

3.1.1. Création d'une lettre d'information et d'un formulaire de consentement éclairé

Il a été remis à tous les participants à notre protocole une lettre d'information précisant les modalités et les objectifs de l'étude ainsi qu'une lettre de consentement éclairé à remplir avant le début des passations. Un exemplaire de chacun de ces documents figure en annexe p. 81-84.

3.1.2. Le recueil de données démographiques

Avant la passation du protocole, chaque participant a dû remplir un formulaire d'information (cf annexe p. 40-45). Celui-ci pouvait être rempli par le participant lui-même ou bien, pour les participants aphasiques, avec l'aide de son conjoint, de l'orthophoniste, ou avec nous-mêmes le jour de la passation. Nous nous sommes également appuyées sur les dossiers médical et orthophonique du participant pour obtenir certaines informations supplémentaires qui ne figuraient pas le formulaire.

3.1.3. Le recueil de données langagières

Afin de maximiser la comparabilité entre les sessions, nous avons, comme indiqué dans les instructions de passation du protocole, essayé de rester le plus fidèle possible au script de départ, d'être aussi silencieux que possible pendant l'administration du protocole, en donnant un maximum d'encouragements non verbaux.

Toutes les passations ont été enregistrées en format vidéo ou audio. Nous avons choisi de ne pas imposer la vidéo si cela dérangeait le participant. L'entretien a donc été enregistré au format audio lorsque certains participants ont refusé l'enregistrement vidéo. Afin d'anonymiser les données transmises à la base de données, tous les fichiers ont ensuite été convertis au format audio.

Nous avons essayé d'effectuer l'enregistrement dans un endroit le plus calme et silencieux possible, afin que la qualité de l'audio ou de la vidéo ne soit pas trop altérée.

Nous avons effectué les passations en premier lieu auprès des sujets contrôles à leur domicile. Le protocole a pu être administré en une seule session pour tous les participants.

Puis, nous avons effectué les passations auprès des sujets aphasiques des épreuves de discours et des épreuves complémentaires. La passation s'est parfois déroulée en plusieurs fois car cela engendrait une trop grande fatigue pour certains patients. Nous avons suivi les indications du protocole préconisant de faire passer tous les tests complémentaires mêmes si ceux-ci ont été passés peu de temps avant, à l'exception du test de dénomination qui devait être réalisé uniquement si nous ne disposions pas de scores datant de moins de six mois.

3.2. Résultats concernant l'utilisation du protocole AphasiaBank avec des sujets contrôles et des sujets aphasiques francophones

3.2.1. Résultats concernant le recueil de données démographiques

3.2.1.1. Résultats concernant l'utilisation du formulaire à remplir par les participants contrôles

Lorsque nous avons fait remplir le formulaire par les participants contrôles, nous avons dû préciser certains éléments à l'oral.

Tout d'abord, nous avons dû préciser la définition du bilinguisme qui n'était pas assez claire. En effet, la formulation « langue parlée » ne précisait pas quel niveau de langue était requis pour être identifié comme bilingue. Ainsi, certains participants se disaient bilingues alors qu'ils ne l'étaient pas réellement. Pour la détermination du nombre d'années d'études effectuées à partir du CP, il a fallu préciser de ne pas compter les années redoublées et les années de formation professionnelle.

3.2.1.2. Résultats concernant l'utilisation du formulaire à remplir par les participants aphasiques

Les informations provenant du formulaire à remplir par le participant aphasique ne permettaient pas de remplir toutes les rubriques du tableau de données démographiques. Afin de ne pas surcharger le participant de questions, nous avons fait le choix de ne pas ajouter de questions supplémentaires au questionnaire, mais avons listé les informations qui nous manquaient et que nous devons rechercher dans le dossier du patient. Les informations que nous avons dû rechercher sont les suivantes :

- Etiologie de l'aphasie
- Type d'aphasie
- Apraxie de la parole
- Dysarthrie
- Côté de la lésion
- Etiologie de la lésion
- Localisation de la lésion
- Outils pour déterminer la localisation de la lésion
- Description de la lésion
- Côté de la lésion des autres AVC
- Antécédent ou présence d'autre(s) pathologie(s) neurologique(s)
- Antécédent ou présence de trouble(s) psychiatrique(s)
- Le patient suit-il toujours une rééducation?
- Fréquence des séances d'orthophonie
- Antécédent de trouble(s) du langage

3.2.2. Résultats concernant le recueil de données langagières

3.2.2.1. Résultats concernant le recueil de données langagières auprès des participants contrôles

3.2.2.1.1. Durée de la passation

Toutes les passations ont pu se dérouler en une seule fois.

En moyenne, la passation auprès des sujets contrôles a duré 16 minutes et 40 secondes. Le temps de chacune des passations est représenté dans le tableau suivant :

Nom du participant	Durée de la passation du protocole
colin01a	11 minutes 21 secondes
colin02a	18 minutes 15 secondes
colin03a	17 minutes 20 secondes
colin04a	15 minutes 15 secondes
colin05a	7 minutes 43 secondes
colin06a	13 minutes 7 secondes
colin07a	17 minutes 5 secondes
lemeur01a	27 minutes 45 secondes
lemeur02a	14 minutes 38 secondes
lemeur03a	11 minutes 33 secondes
lemeur04a	19 minutes 28 secondes
lemeur05a	17 minutes
lemeur06a	15 minutes 56 secondes
lemeur07a	26 minutes 59 secondes

Tableau n°10: Durée des passations auprès des sujets contrôles

3.2.2.1.2. Conditions de passation

Tous les participants contrôles ont accepté d'être filmés durant la passation.

Nous avons essayé d'effectuer l'enregistrement dans un endroit calme et silencieux, afin que la qualité de l'audio ou de la vidéo ne soit pas trop altérée mais cela n'a pas toujours été possible. En effet, l'enregistrement de colin06a a dû se dérouler en extérieur, ce qui a considérablement perturbé les conditions d'enregistrement.

La présence de personnes tierces pendant l'entretien a également perturbé certaines passations.

3.2.2.1.3. Utilisation du protocole

Si des précisions ont dû être apportées en cours de passation, suite à des questionnements des participants contrôles, le protocole a pu être utilisé dans son intégralité avec tous les participants contrôles.

La question concernant la récupération suite à la maladie a posé problème à un participant contrôle. En effet, celui-ci ne savait pas qu'il faudrait relater sa récupération après sa maladie et il n'en avait pas de souvenir. Il a donc choisi de raconter une autre maladie.

La question « **Avez-vous déjà rencontré des personnes ayant des difficultés pour communiquer ?** » a posé problème à plusieurs participants qui ont souvent demandé des précisions sur ce que cela signifiait. Certains ont même eu des difficultés à répondre à cette question. Par exemple, colin02a a demandé :

***PAR:** &euh difficultés par manque de [/] de communication c'est_à_dire d'ordre psychologique ou difficultés par manque de communication d'ordre &euh &euh disons lésion neurologique et [/] et autre ? ¹⁶

La question « **Pourriez-vous me raconter un événement marquant de votre vie ? Cela peut être un événement joyeux ou triste, ancien ou récent.** » a déstabilisé certains participants. Cela a engendré le récit d'événements très personnels et difficiles à raconter pour certains, qui n'ont de ce fait pas souhaité détailler davantage.

Concernant l'épreuve de discours narratif à partir d'images, plusieurs questions ont été posées afin d'obtenir des précisions sur les consignes. Par exemple, colin02a a demandé ceci :

***PAR:** l'histoire a [/] a trait à ce que je vois j'imagine ? [+ exc]

Le participant colin06a a demandé s'il fallait raconter une histoire « *en rapport avec ces images* » et lemeur02a a demandé, au sujet de l'histoire du sauvetage du chat, s'il fallait raconter une « *histoire globale* ». Les participants colin03a et lemeur05a ont quant à eux raconté des

16 L'explication détaillée de chacun des codages utilisés dans nos transcriptions figure p. 96.

histoires qu'ils avaient personnellement vécues à partir des images qu'ils avaient devant eux.

Concernant la narration de l'histoire de Cendrillon, seul le participant colin06a ne connaissait pas l'histoire.

Quatre participants n'ont pas eu besoin de regarder le livre d'images.

Nous avons parfois dû nous éloigner du script en précisant aux participants qu'il ne fallait pas raconter l'histoire avec les images sous les yeux, cela n'étant pas précisé dans le script original.

3.2.2.2. Résultats concernant le recueil de données langagières auprès des participants aphasiques

3.2.2.2.1. Résultats concernant le recueil de données discursives auprès des participants aphasiques

3.2.2.2.1.1. Durée de la passation

Toutes les passations des épreuves de discours ont pu se dérouler en une seule fois.

En moyenne, la passation du protocole des épreuves de discours auprès des sujets aphasiques a duré 17 minutes et 04 secondes allant de 9 minutes et 15 secondes pour la plus courte à 29 minutes et 20 secondes pour la plus longue.

Nous avons également calculé la durée moyenne de passation de chaque tâche de discours du protocole pour les participants aphasiques. Le tableau détaillant le temps de chaque épreuve pour chaque participant est disponible en annexe p. 85-86.

Nom de l'épreuve	Durée moyenne	Durée la plus courte	Durée la plus longue
<i>Histoire de l'AVC et récupération</i>	3 minutes 44 secondes	1 minute 33 secondes	7 minutes 26 secondes
<i>Événement important</i>	1 minute 22 secondes	27 secondes	2 minutes 43 secondes
<i>La fenêtre cassée</i>	1 minute 51 secondes	49 secondes	2 minutes 50 secondes
<i>Le parapluie</i>	2 minutes 49 secondes	50 secondes	5 minutes 9 secondes
<i>Le sauvetage du chat</i>	2 minutes 3 secondes	40 secondes	4 minutes 6 secondes
<i>Cendrillon</i>	4 minutes 45 secondes	2 minutes 18 secondes	9 minutes 30 secondes
<i>Sandwich</i>	1 minute et 4 secondes	21 secondes	2 minutes 5 secondes

Tableau n°11 : Durée moyenne de la passation de chaque tâche de discours du protocole chez les participants aphasiques

3.2.2.2.1.2. Concernant les conditions de passation

Trois participants aphasiques ont refusé l'enregistrement vidéo. Pour ces participants, les passations ont donc été enregistrées au format audio uniquement.

Les conditions d'enregistrement n'ont pas toujours été optimales. L'enregistrement « colin10a » est perturbé par la présence de la radio en bruit de fond en début d'enregistrement.

3.2.2.1.3. Utilisation du protocole de discours

La passation du protocole de discours a pu se dérouler dans l'ordre prévu par le script pour neuf participants sur onze. En effet, pour les participants colin11a et lemeur09a, l'épreuve « Sandwich » a été réalisée avant l'épreuve « Cendrillon ».

• **Résultats concernant la passation de l'épreuve « Histoire de l'AVC et récupération »**

A la question « *Comment trouvez-vous votre langage ces jours-ci ?* », les participants ont souvent répondu avec très peu de phrases et ont parfois eu du mal à trouver les mots pour le dire. Beaucoup nous ont répondu « *je ne sais pas* ». Voici par exemple la réponse du participant colin14a :

*PAR: &euh <c'est pas> [/] c'est pas très facile.
*PAR: mais enfin bon j'ai pas grand chose à vous dire là hein.

Le participant lemeur10a a lui demandé des précisions sur la consigne :

*PAR: pff &se@u qu'est_ce_que vous voulez dire ?

Nous avons ainsi dû relancer une grande majorité des participants sur cette question, en posant de nouveau la question « est-ce que vous pouvez me parler un peu de votre langage oral ? ».

Le questionnaire d'aide a ainsi dû être utilisé à deux reprises.

La question « Vous rappelez-vous quand vous avez eu votre AVC ? » a également posé beaucoup de difficultés à une majorité de participants, que nous avons souvent dû relancer. Voici par exemple un extrait de l'échange avec le participant colin10a :

*INV: est_ce_que est_ce_que vous pouvez me le raconter ?
(...)
*PAR: &soupir je sais pas le dire.
*INV: prenez votre temps.
*PAR: oh non.
*INV: non c'est trop difficile à dire ?
*PAR: oui [/] oui je sais pas.
(...)
*INV: et quels sont vos premiers souvenirs après votre A_V_C ?
*PAR: comment [/] &se@u comment +//.
*PAR: je sais pas non.

Le participant colin11a n'a pas pu raconter l'histoire de son AVC :

*INV: est_ce_que vous vous souvenez de votre A_V_C?
*PAR: non.
*INV: non ?
*PAR: non.
*INV: et quels sont vos premiers souvenirs après votre A_V_C?
*PAR: &euh après &euh &euh l'A_V_C &euh non &rho@u . [+ * gram]
*INV: non ?
*INV: d'accord.

• Résultats concernant la passation de l'épreuve « Événement important »

Au cours de cette épreuve, les participants aphasiques ont souvent eu des difficultés à trouver un événement important ou bien n'ont pas pu le raconter. Cependant, il n'existe pas de question d'aide pour cette épreuve, nous avons donc dû passer à l'épreuve suivante.

Les participants colin09a et lemeur10a ont dû être relancés plusieurs fois. Ils ont demandé des précisions sur la consigne à plusieurs reprises.

Un participant a refusé de raconter l'événement important, car le récit du décès d'un proche était trop dur à raconter.

@G: Important_Event
*INV: est_ce_que vous pourriez me raconter un événement important de votre vie ça peut être joyeux triste ancien récent.
*PAR: oh_ben j'ai perdu ma mère.
*PAR: ça a été terrible.
*PAR: voilà c'est tout.
*PAR: j'ai pas d'autre chose à dire.
*INV: d'accord.

Les participants colin11a et colin13a ne sont pas parvenus à le raconter.

Voici par exemple un extrait de l'échange avec le participant colin11a :

@G: Important_Event
*INV: pourriez-vous me raconter un événement important de votre vie ancien ou récent triste ou joyeux ?
*PAR: <triste ou joyeux> [/] &euh triste ou joyeux &euh. [+ * per] [+ * gram]
*PAR: &soupirs xxx non .
*INV: non ?
*PAR: non.
*INV: d'accord.

Les épreuves de la section I « Discours spontané » ont ainsi posé beaucoup de problèmes aux participants aphasiques. Cela leur a semblé assez difficile et certains ont pu l'exprimer en manifestant leur préférence pour des épreuves avec un support imagé :

*INV: je vais maintenant vous montrer des images.

*PAR: &m ah c'est bon. [+ exc]

*INV: vous préférez ?

*PAR: ouais [/] ouais c'est mieux. [+ exc]

*INV: je vais maintenant vous montrer des images.

*PAR: oui. [+ exc]

*PAR: ah ça sera peut_être plus facile. [+ exc]

• Résultats concernant la passation de l'épreuve « Narration d'histoire à partir d'images »

Bien que les cases soient numérotées sur les images, et même si cela n'est pas indiqué dans le script, nous avons dû préciser le sens de lecture des images pour quatre participants.

Le récit de la fenêtre cassée a pu être raconté par neuf participants. Le questionnaire d'aide a donc dû être utilisé deux fois pour cette épreuve.

Le récit du parapluie a également pu être raconté par neuf participants. Il a posé problème à deux participants qui ne savaient pas quoi raconter. Nous avons donc dû nous éloigner du protocole afin de préciser la consigne. Le questionnaire d'aide a dû être utilisé une seule fois.

Tous les participants ont pu faire le récit du « sauvetage du chat ».

• Résultats concernant la passation de l'épreuve « Cendrillon »

Seulement sept participants ont pu raconter l'histoire de Cendrillon. Sur les quatre participants qui n'ont pas pu raconter l'histoire de Cendrillon, deux ne connaissaient pas du tout l'histoire et deux n'ont pas réussi à la raconter. Nous avons donc proposé aux participants de raconter un autre conte de leur choix. Sur les quatre, un a finalement pu raconter le conte du « Petit Chaperon Rouge » et un autre l'histoire du « Petit Poucet ». Les deux autres participants ne sont pas parvenus à raconter un autre conte. Ainsi, l'échantillon de discours recueilli pour ces participants ne pourra pas être comparé à la base de données.

Un seul participant n'a pas voulu regarder le livre avant de raconter l'histoire.

Tous les autres participants ont pu feuilleter le livre une fois avant de raconter l'histoire, à l'exception du participant lemeur10a qui a pu feuilleter le livre une seconde fois, suite à sa demande.

Quatre d'entre eux ont fait beaucoup de commentaires en regardant les images, en

commentant les détails présents sur l'image ou bien ont commencé le récit avant d'avoir terminé de regarder les images. Cela n'étant pas prévu dans le script nous nous sommes autorisées à préciser au participant qu'il fallait attendre de ne plus avoir le livre sous les yeux avant de raconter.

Le questionnaire d'aide a dû être utilisé trois fois pour cette épreuve.

- **Résultats concernant la passation de l'épreuve « Sandwich »**

Dix participants ont pu décrire comment ils feraient un sandwich jambon beurre. Pour un participant l'utilisation des images a été nécessaire comme le préconise le questionnaire d'aide, ce dernier a donc dû être utilisé une seule fois.

- **Résultats concernant la passation du « questionnaire d'aide »**

Au total, le questionnaire d'aide a donc été utilisé neuf fois. Le tableau récapitulatif du nombre d'utilisation du questionnaire d'aide par épreuves est disponible en annexe p. 87.

3.2.2.2.2. Résultats concernant le recueil de données issues des épreuves complémentaires

3.2.2.2.2.1 Durée de la passation

La passation des épreuves complémentaires a pu se dérouler en une seule fois, à la suite des épreuves de discours, pour neuf participants. Les passations des participants colin09a et colin15a ont dû se dérouler en deux fois, du fait de la fatigue générée par la passation des différents tests.

La passation des épreuves complémentaires a duré en moyenne 32 minutes et 53 secondes.

3.2.2.2.2.2. Utilisation des épreuves complémentaires

Tous les tests ont pu être réalisés, dans l'ordre, pour tous les participants aphasiques.

Les résultats obtenus aux épreuves complémentaires par les participants aphasiques sont disponibles en annexe p. 77-80.

3.2.2.2.3. Résultats concernant le recueil de l'ensemble des données langagières

La passation de la totalité des épreuves langagières a duré en moyenne 49 minutes et 57 secondes.

Voici la durée de la passation du protocole et des épreuves complémentaires pour chaque participant aphasique :

Nom du participant	Durée de la passation des épreuves de discours	Durée de la passation des épreuves complémentaires	Durée totale de la passation
colin09a	12 minutes 4 secondes	23 minutes 22 secondes	35 minutes 26 secondes
colin10a	14 minutes 9 secondes	45 minutes 59 secondes	1 heure 8 secondes
colin11a	15 minutes 49 secondes	34 minutes 44 secondes	50 minutes 33 secondes
colin13a	14 minutes 16 secondes	17 minutes 52 secondes	32 minutes 8 secondes
colin14a	18 minutes 24 secondes	44 minutes 08 secondes	1 heure 2 minutes 32 secondes
colin16a	15 minutes 54 secondes	31 minutes 08 secondes	47 minutes 2 secondes
lemeur09a	21 minutes 57 secondes	27 minutes 26 secondes	49 minutes 23 secondes
lemeur10a	22 minutes 18 secondes	51 minutes 34 secondes	1 heure 13 minutes 52 secondes
lemeur11a	14 minutes 20 secondes	26 minutes 23 secondes	40 minutes 43 secondes
colin12a	9 minutes 15 secondes	23 minutes 07 secondes	32 minutes 22 secondes
colin15a	29 minutes 20 secondes	36 minutes 06 secondes	1 heure 5 minutes 26 secondes

Tableau n°12 : Durée de la passation des épreuves de discours et des épreuves complémentaires auprès des sujets aphasiques

4. L'OPÉRATIONNALISATION DES DONNÉES POUR L'UTILISATION D'ÉVAL

Nous allons d'abord présenter notre démarche méthodologique puis nous présenterons les résultats obtenus à l'issue de cette démarche.

4.1. Méthodologie de l'opérationnalisation des données pour l'utilisation d'ÉVAL

4.1.1. Sélection des conventions de transcription et des codages pertinents

Le système CHAT propose des conventions de transcription et de codage très approfondies permettant de transcrire tous types d'interactions conversationnelles. L'utilisation de l'ensemble de ces conventions n'est cependant pas nécessaire dans le cadre de l'utilisation d'ÉVAL. Un système de codage simplifié, conçu pour les utilisateurs d'AphasiaBank, est suffisant.

Si certains des codages CHAT sont nécessaires au bon fonctionnement de l'analyse ÉVAL car ils ont une influence sur les mesures calculées, d'autres ne sont pas indispensables et le transcripteur est libre de les utiliser ou non. Ainsi, chaque transcripteur peut choisir, dans une certaine mesure, les normes et les codages qu'il souhaite utiliser pour ses transcriptions, en fonction des buts poursuivis par son analyse.

Aussi, avant de transcrire nos données, nous avons réfléchi aux normes et aux codages

que nous souhaitions utiliser dans le cadre de notre objectif, à savoir analyser le langage d'un point de vue clinique orthophonique.

Dans cette perspective, la sélection des normes de transcription et de codage devait répondre à trois objectifs que nous avons tenté de concilier :

- les transcriptions ainsi codées devaient être analysables de façon optimale par EVAL ;
- elles devaient contenir suffisamment d'informations sur l'interaction pour permettre à l'orthophoniste de pouvoir analyser le langage du participant par la lecture de la transcription sans avoir besoin de réécouter l'enregistrement audio¹⁷ ;
- et les conventions de codage ne devaient pas être trop complexes afin d'être facilement utilisables par des cliniciens.

Afin de poursuivre ces trois objectifs, nous avons établi notre propre liste de normes et de codages en conservant certaines normes de codage prévues par le système de codage simplifié pour AphasiaBank, en enlevant d'autres, et en ajoutant d'autres normes provenant du système de normes plus détaillé du système CHAT. L'important était d'établir une liste de normes et de codages que nous allions utiliser de façon uniforme pour l'ensemble des corpus que nous allions retranscrire.

4.1.1.1. Sélection des normes générales de transcription et des codages

Le manuel EVAL propose d'effectuer une transcription très simplifiée car de nombreux éléments ne sont pas nécessaires au bon fonctionnement de l'analyse EVAL. Il est ainsi suggéré de ne pas transcrire les propos de l'examineur et de transcrire uniquement ce que dit le participant en réponse directe aux incitations du protocole, en omettant tous les échanges sociaux, les mots de remplissage tels que « *euh, hum* » et les fragments de mots.

Étant donné que notre objectif n'était pas uniquement d'obtenir une analyse EVAL mais également de permettre à l'orthophoniste d'effectuer une analyse clinique sur le discours retranscrit, nous avons fait le choix de transcrire l'interaction de façon plus détaillée. Ainsi, nous avons décidé de transcrire les propos de l'examineur, les échanges sociaux, ainsi que tous les mots et fragments de mots produits par le participant. En effet, pour analyser le discours du patient, l'orthophoniste a besoin de connaître le contexte global de l'interaction et de connaître l'ensemble des productions émises par le participant. Afin de ne pas trop complexifier la tâche de transcription, nous avons cependant décidé de ne pas segmenter les propos de l'examineur en plusieurs énoncés et de ne pas coder les reformulations, les répétitions ou les erreurs du discours de l'examineur.

17 Ceci a nécessité de conserver suffisamment de codages pour préciser divers éléments non verbaux ou paraverbaux (exemple : codages des marqueurs de fin d'énoncés explicitant la présence de phrases interrogatives ou exclamatives, d'interruptions du locuteur, ou encore la présence de pauses ou de gestes dans le discours, etc.).

Nous avons choisi de coder l'ensemble de nos corpus selon les normes de codages suivantes :

Codage	Définition du codage
.	Marqueur de fin d'énoncé affirmatif
?	Marqueur de fin d'énoncé interrogatif
!	Marqueur de fin d'énoncé exclamatif
+/.	Marqueur de fin d'énoncé quand le locuteur a été interrompu par un autre locuteur
+//.	Marqueur de fin d'énoncé quand le locuteur s'est auto-interrompu
+"/.	Codage du discours indirect : à la fin d'un énoncé pour annoncer qu'une citation suivra sur la ligne suivante
+"	Codage du discours indirect : au début de l'énoncé où commence la citation
+".	Codage du discours indirect : à la fin de l'énoncé à l'endroit où la citation se termine
&	Codage permettant d'exclure certains mots (comme les mots de remplissage, exemple : &euh, &hum), fragments de mots (amorces de mots, conduites d'approches, ...) ou de coder du non verbal (exemple : &gestedemanger). Les fragments de mots sont retranscrits en alphabet phonétique international (qui figure en annexe p. 88) et le symbole @u est attaché au fragment pour indiquer au lecteur que l'alphabet phonétique est utilisé à la place de l'alphabet standard (ex : &βø@u). Il est possible de préciser la raison de l'utilisation du symbole & par l'ajout du symbole &= suivi de l'explication (par exemple : &pfff &=soupir). Les communicateurs ¹⁸ ne sont pas considérés comme des mots de remplissage, ils ne sont donc pas exclus de l'analyse EVAL (exemple : oh, ben,...).
xxx	Mots ou groupes de mots inintelligibles
www	Événements non transcriposables ou qu'on ne souhaite pas retranscrire Une ligne supplémentaire (%exp) peut ensuite être ajoutée afin d'expliquer l'événement qui s'est produit, comme dans l'exemple suivant : *INV: www. %exp: énoncé à l'intention d'une tierce personne.
[/]	Codage des répétitions Mettre les éléments répétés entre <> à moins qu'un seul mot ait été répété, CLAN supposera alors que seul le mot avant les crochets a été répété. Si un mot est répété plusieurs fois mettre toutes les répétitions successives entre <>. *PAR: &euh <oui oui oui oui> [/] oui ben <j'ai> [/] j'ai nagé principalement fait de la natation.
[//]	Codage des reformulations Mettre les éléments reformulés entre <> à moins qu'un seul mot ait été changé, CLAN supposera alors que seul le mot avant les crochets a été reformulé. Exemple : *PAR: &euh et puis y a des situations de conflit <qui sont> [//] qui peuvent ne pas se résoudre.
(.)	Pause entre des mots Nous avons choisi d'utiliser ce codage à partir de 5 secondes de pause pour disposer d'un critère objectif.
()	Codage des abréviations Dans le manuel EVAL, il est conseillé de transcrire le mot cible directement si le participant abrège un mot, mais nous faisons le choix de transcrire les phonèmes manquants entre parenthèses pour rester fidèle au discours du participant tout en permettant au programme MOR de reconnaître les mots abrégés. Exemple : kiné(sithérapie)

18 Le terme de communicateur est utilisé dans les fichiers lexiques de la grammaire MOR pour désigner les mots ayant une fonction phatique (visant à entrer en relation ou à maintenir le contact avec l'interlocuteur).

[+ exc]	<p>Codage permettant d'exclure un énoncé de l'analyse EVAL</p> <p>Utilisé pour les énoncés qui ne doivent pas être inclus dans l'analyse EVAL, en particulier dans les tâches de discours « dirigé » (description d'images, narration d'histoire, discours procédural), par exemple :</p> <ul style="list-style-type: none"> - commentaires à propos de la tâche (« <i>c'est dur</i> », « <i>je n'y arrive pas</i> ») - apartés (« <i>j'ai besoin de mes lunettes</i> ») - réponses aux incitations ne faisant pas partie des instructions du protocole AphasiaBank qui fournissent au participant des informations dont ne disposent pas les autres participants.
----------------	--

Tableau n°13 : Sélection des normes générales de codage

4.1.1.2. Sélection des codages d'erreurs

L'utilisation de l'ensemble du système de codage d'erreurs CHAT n'est pas nécessaire pour EVAL. En effet, EVAL mesure le pourcentage d'erreurs au niveau du mot et d'erreurs au niveau de l'énoncé mais n'analyse pas le type d'erreur de façon plus détaillée. Ainsi seuls deux codes d'erreurs, [*] pour les erreurs au niveau du mot, et [+ *] pour les erreurs au niveau de l'énoncé, sont nécessaires. Cependant le transcripateur peut, s'il le souhaite, utiliser des codes supplémentaires afin de détailler davantage le type d'erreurs produites par le participant.

L'analyse du discours en orthophonie nécessitant de connaître avec précision le type d'erreurs produites, nous avons sélectionné des codages d'erreurs supplémentaires pour affiner l'analyse. Cependant le système de codage d'erreurs complet est trop compliqué et son utilisation rendrait la tâche de transcription trop longue. Ainsi, afin de ne pas trop complexifier la tâche de transcription, nous n'avons pas utilisé l'ensemble des codages proposés par le système CHAT mais avons sélectionné ceux qui nous paraissaient les plus pertinents.

Nous avons ainsi sélectionné les codages d'erreurs suivants :

Codages d'erreurs au niveau du mot

Codage	Définition du codage
[* p]	Erreurs phonologiques
[* s]	Erreurs sémantiques
[* s:per]	Persévérations au niveau du mot
[* n]	Néologismes
[* d]	Disfluences à l'intérieur d'un mot
[* m]	Erreurs morphologiques

Tableau n°14 : Sélection des codages d'erreurs au niveau du mot

Concernant les erreurs au niveau du mot, nous avons sélectionné cinq des sept catégories principales d'erreurs (erreurs phonologiques, erreurs sémantiques, néologismes, disfluences et erreurs morphologiques). Afin de ne pas trop complexifier les normes de codages, nous avons décidé de regrouper les deux autres catégories d'erreurs principales, "erreurs d'accord morphologique" ([* m:a]) et "outils lexicaux formels" ([* f]) (qui concernent des erreurs dans le choix des déterminants), sous le codage des erreurs morphologiques [* m]. En effet, nous

estimons que les erreurs d'accord morphologique et les erreurs concernant l'utilisation des déterminants constituent des erreurs morphologiques et peuvent donc, pour simplifier la tâche de transcription, être regroupées sous le même code d'erreur. Dans le même souci de simplification, nous avons choisi de ne pas reprendre toutes les sous-catégories d'erreurs permettant de détailler chacune des catégories principales. Nous avons seulement ajouté une sous-catégorie des erreurs sémantiques afin de pouvoir coder les persévérations [* s:per]. Nous avons également choisi de coder comme persévérations les amorces de mot qui constituent une persévération sur un mot précédemment employé, bien que le participant n'emploie pas le mot dans son intégralité.

Le document "Error coding" donne une définition très précise de ce qui peut être considéré comme une erreur phonologique (cf annexe p. 31-32). Nous avons choisi de ne pas respecter ces critères car ils ne sont pas appliqués en pratique clinique orthophonique française. Nous avons choisi de coder une erreur phonologique lorsque l'erreur affecte le niveau de deuxième articulation du langage, soit lorsque la forme phonologique du mot est modifiée et que le mot est substitué par un mot n'appartenant pas au lexique de la langue (paraphasies phonémiques), soit lorsque le mot est substitué par un mot du lexique partageant des phonèmes communs avec le mot cible (paraphasies verbales formelles et morphologiques).

Si l'erreur n'est pas un vrai mot, nous avons choisi de la retranscrire aussi précisément que possible en utilisant l'alphabet phonétique international.¹⁹ Afin d'indiquer au lecteur que l'erreur est écrite en alphabet phonétique, et non en alphabet standard, nous attachons le symbole @u à l'erreur (exemple : bɔzjɛ@u).

Si le mot cible pour les erreurs phonologiques, sémantiques et les néologismes est connu, il doit être retranscrit entre crochets avec deux points suivis d'un espace, afin que le programme MOR puisse identifier la catégorie grammaticale appropriée, comme dans l'exemple ci-dessous :

PAR: j'avais pas bɔzjɛ@u [: besoin] [p] de [/] de +//·

Codages d'erreurs au niveau de l'énoncé

Codage	Définition du codage
[+ * gram]	<p>Erreurs grammaticales</p> <p>Inclut l'agrammatisme et le paragrammatisme mais véhicule avec succès un message en rapport avec le contexte :</p> <ul style="list-style-type: none"> - Énoncés dans lesquels des éléments grammaticaux nécessaires (sujets, verbes, auxiliaires, prépositions) sont manquants ou mal utilisés, à l'exception des réponses appropriées d'un seul mot ou d'autres communicateurs d'un seul mot (« oui », « hum ») - Énoncés contenant des erreurs dans l'ordre des mots, de syntaxe ou de morphologie grammaticale - Ne pas coder les énoncés qui ne sont pas grammaticalement corrects du fait de l'utilisation de formules verbales régionales comme erreurs grammaticales

19 Nous avons utilisé l'alphabet phonétique disponible à l'adresse suivante : <http://w3.uohprod.univ-tlse2.fr/UOH-PHONETIQUE-FLE/DOCS/DOC01.pdf>. Ce document figure en annexe p. 88.

[+ * jar]	Jargon Généralement fluent et correct au niveau prosodique mais discours en grande partie vide de sens (contenant des paraphrasies, des néologismes ou des éléments inintelligibles) ; ressemble à la syntaxe française
[+ * es]	Discours vide Discours syntaxiquement correct mais véhiculant peu ou pas de sens général, résulte souvent de la substitution de mots génériques (« chose », « truc ») à des termes plus spécifiques
[+ * per]	Persévérations au niveau de l'énoncé Répétition d'un énoncé quand cela n'est plus approprié
[+ * cir]	Circonlocutions Tourner autour des concepts ou des mots

Tableau n°15 : Sélection des codages d'erreurs au niveau de l'énoncé

Les énoncés incomplets, où le locuteur s'est auto-interrompu, ne doivent pas être codés comme des erreurs au niveau de l'énoncé, même si des éléments grammaticaux sont manquants, car nous ne pouvons pas connaître l'énoncé cible.

Enfin, bien qu'il ne soit précisé nulle part explicitement dans le protocole AphasiaBank si les codages d'erreurs doivent être utilisés également dans les transcriptions des sujets contrôles, nous avons fait le choix d'utiliser les mêmes codages d'erreurs à la fois dans les transcriptions des sujets contrôles et dans les transcriptions des sujets aphasiques.

Les étapes de transcription et de préparation des transcriptions à l'analyse EVAL se sont faites en deux temps. Nous avons d'abord transcrit et préparé les transcriptions des sujets contrôles, puis, dans un second temps, celles des sujets aphasiques.

4.1.2. Transcription et codage des données

Les données recueillies sur support audio doivent être transcrites dans le programme de traitement de texte CLAN selon les normes de transcription et les codages CHAT que nous avons pré-sélectionnés.

La première étape de la transcription consiste à créer un nouveau fichier de transcription et d'y préciser les en-têtes obligatoires et additionnelles permettant d'identifier le corpus et le participant. Pour chacune de nos transcriptions, nous avons copié les en-têtes à partir du fichier « *Transcript Template* »²⁰, nous les avons insérées dans nos fichiers et les avons modifiées pour les adapter à nos corpus. Nous avons donc dû modifier :

- l'en-tête « Languages » pour y préciser que la langue utilisée était le français ;
- l'en-tête « Participants » pour y indiquer l'identité du participant ;
- les en-têtes « ID » pour préciser les caractéristiques du participant (langue utilisée, nom de

²⁰ Ce document est un exemple des en-têtes figurant au début d'une transcription. Il est disponible sur le site AphasiaBank.

l'investigateur, âge et sexe du participant, type d'aphasie ou « Control » si le participant était un sujet contrôle) ;

– l'en-tête « Date » pour y préciser la date de passation des épreuves de discours ;

– nous avons ajouté l'en-tête « Media » pour indiquer le nom du fichier audio correspondant à la transcription ;

– enfin, nous avons ajouté les en-têtes notées @G afin de délimiter les différentes tâches de discours dans la transcription :

En-tête	Description de la tâche de discours
@G : Speech	Tâche consistant à parler de sa rencontre avec des personnes en difficulté de communication (pour les sujets contrôles) ou de son langage (pour les patients aphasiques)
@G : Stroke	Tâche consistant à raconter l'histoire d'une maladie ou d'une blessure (pour les sujets contrôles) ou de son AVC (pour les sujets aphasiques), et de leur récupération
@G : Important_Event	Tâche consistant à raconter un événement important
@G : Window	Tâche consistant à raconter une histoire à partir d'images
@G: Umbrella	Tâche consistant à raconter une histoire à partir d'images
@G : Cat	Tâche consistant à raconter une histoire à partir d'une image
@G: Cinderella	Tâche consistant à raconter une histoire connue, sans support imagé sous les yeux
@G: Sandwich	Tâche de discours procédural consistant à décrire comment réaliser un sandwich jambon beurre

Tableau n°16 : Description des en-têtes Gem et des tâches de discours correspondantes

Lorsque les participants n'ont pas voulu ou pas pu raconter l'histoire de Cendrillon, nous avons laissé l'en-tête « @G : Cinderella » dans la transcription mais nous avons exclu tous les énoncés qui ne correspondaient pas à la tâche de narration demandée afin qu'ils ne soient pas considérés comme des corpus à comparer à la base de données.

Lorsque les participants ont raconté un autre conte de fée à la place de l'histoire de Cendrillon, nous avons modifié l'intitulé de l'en-tête pour y préciser le nom du conte de fée raconté par le participant (ex : @G : Le_Petit_Chaperon_Rouge). Ainsi, le discours produit par les participants pourra être analysé par EVAL mais ne sera pas comparé aux autres corpus de la base de données relatifs à la narration de l'histoire de Cendrillon.

Une fois les en-têtes précisées, nous avons pu passer à la transcription proprement dite. Deux exemples de nos transcriptions figurent en annexe p. 89-93 (sujet contrôle) et p. 94-99 (sujet aphasique). Pour effectuer la transcription, il a été nécessaire d'écouter de nombreuses fois le fichier audio et de l'écouter parfois au ralenti pour mieux discerner les propos du participant.²¹ Nous avons d'abord transcrit l'ensemble de l'interview et avons procédé au codage de la transcription dans un second temps.

Lors de la première phase de transcription, avant la phase de codage, de nombreuses

21 Pour ralentir la lecture des fichiers audio, nous avons utilisé certains logiciels tels qu'Audacity ou VLC.

questions ont progressivement émergé.

En effet, malgré les recommandations du manuel EVAL et la réflexion préalable que nous avons menée sur la sélection des codages et des normes de transcription, nos transcriptions manquaient d'uniformité, nous avons donc établi des principes plus précis de transcription, concernant notamment la segmentation en énoncés, l'exclusion des énoncés, les morphèmes à solidariser et le codage des répétitions.

• **Segmentation en énoncés**

La décision de segmenter en énoncés est importante car elle influence plusieurs mesures calculées par EVAL (*notamment le nombre total d'énoncés, la longueur moyenne des énoncés et le nombre de verbes par énoncé*). Or cette décision est souvent difficile à prendre, et ce malgré les recommandations du manuel EVAL. Ce dernier suggère de s'appuyer préférentiellement sur la syntaxe et l'intonation, puis sur les pauses et les indices sémantiques, et de préférer les énoncés courts aux énoncés longs.

L'indice le plus important permettant de délimiter un énoncé doit donc être la syntaxe : une phrase syntaxiquement bien formée doit être considérée comme un énoncé (bien qu'un énoncé puisse être considéré comme un énoncé même s'il n'est pas grammaticalement correct). On pourrait donc penser qu'il suffit de segmenter en énoncés dès qu'une construction classique type sujet-verbe a été formée, mais, en discours oral continu, la présence de nombreuses digressions, incises et apartés, peut rendre difficile la perception de la cohérence syntaxique. En effet, la cohérence syntaxique peut parfois s'établir à travers une succession d'énoncés distincts, comme dans l'exemple suivant, où la phrase « *j'avais été chaussée d'une espadrille très fine* » fonctionne comme une incise venant s'insérer au sein de la phrase syntaxiquement bien formée « *je pose mon pied sur une planche avec un clou qui dépassait* ».

***PAR:** et en reculant &euh pour prendre ma photo je pose mon pied j'avais été chaussée d'une petite espadrille très fine sur une planche avec un clou qui dépassait.

Dans ce deuxième exemple, plusieurs propositions sont insérées au sein d'un même énoncé :

***PAR:** et mais [/] mais Cendrillon a été enfermée <par la> [/] par la belle+mère qui ne veut surtout pas qu'elle apparaisse parce_qu' elle est quand_même nettement plus avenante que ses propres filles.

Face à de nombreux cas complexes rendant le choix de la segmentation en énoncés difficile à prendre, nous avons tenté d'établir des règles supplémentaires afin d'uniformiser notre façon de transcrire. Nous avons ainsi procédé aux choix suivants :

Segmenter :

- toutes les propositions indépendantes (*séparées par des conjonctions telle que donc, alors, mais, en fait, et, ...*),

Ne pas segmenter :

- les phrases complexes comprenant plusieurs propositions reliées entre elles (*par des conjonctions telles que parce_que, quand, si, qui, que, avec, pour, où, ...*);

- les énumérations et lorsqu'un même sujet est employé pour plusieurs verbes sans être repris (*exemple : « le papa prend l'échelle la met dans l'arbre »*);

- lorsqu'un commentaire modalisateur (« *je pense* », « *il me semble* »,...) ou une expression phatique (« *tu vois* ») étaient insérés au sein d'un énoncé, ces éléments fonctionnant comme des incises :

***PAR:** mais c'est pas la même relation que avec une fille je crois.

***PAR:** et sa maman lui dit +"/.

***PAR:** +" et oui tu_vois t'es bien mouillé +"

Exclusion des énoncés

La décision d'exclure certains énoncés de l'analyse EVAL est également une décision importante car elle influence aussi les calculs effectués par le programme EVAL. Malgré les recommandations du manuel EVAL et du document « Error coding », cette décision a parfois été difficile à prendre. Afin d'uniformiser nos transcriptions, nous avons choisi d'établir des règles plus précises :

Dans les tâches de discours libre (@G : *Speech, Stroke, Important_Event*)

- Ne rien exclure

Dans les tâches de discours dirigé (@G : *Window, Umbrella, Cat, Cinderella, Sandwich*).

- Ne conserver que les énoncés concernant la tâche à accomplir ;

- Exclure tous les commentaires personnels, les réponses du participant aux consignes, les échanges sociaux ;

- Exclure tous les énoncés produits dans des conditions différentes de celles exigées par le protocole (par exemple lorsque le participant commence à raconter l'histoire de Cendrillon alors qu'il a encore le livre sous les yeux). En effet, dans ce cas, le participant dispose d'informations dont ne disposent pas les autres participants et cela pourrait biaiser la comparaison entre les corpus.

A titre d'exemple, les énoncés suivants ont été exclus :

***PAR:** l'histoire a [/] a trait à ce que je vois j'imagine ? [+ exc])

***PAR:** il ressemble à mon chat. [+ exc]

***PAR:** c'est une histoire qui m'est souvent arrivée ça en classe &rires. [+ exc]

• Morphèmes à solidariser

Le programme MOR effectue un étiquetage des catégories morphologiques des différents morphèmes de façon automatique. Or, dans certains cas, la catégorie morphologique doit s'appliquer non pas à un seul morphème mais à plusieurs morphèmes, qui constituent une unité morphologique à part entière. C'est par exemple le cas des morphèmes suivants :

- mots composés (lave+linge, petit+déjeuner,...);
- locutions verbales (être_en_train_de,...);
- locutions conjonctives (parce_que, ou_bien,...);
- locutions adverbiales (par_contre, en_fait,...);
- locutions prépositionnelles (en_face_de, en_dehors_de,...).

Les morphèmes ainsi solidarisés sont considérés comme une unité solidaire par le programme MOR. Ceci a des conséquences sur les analyses quantitatives effectuées par EVAL sur les corpus.

Par exemple, au lieu de compter deux prépositions « en » et « de » et un substantif « face » dans le segment « en face de », le fait de considérer « en face de » comme une unité morphémique solidaire revient à comptabiliser l'occurrence d'une seule préposition, et non de deux prépositions et d'un substantif.

Ainsi, pour que le programme MOR parvienne à identifier ces morphèmes comme une entité morphologique à part entière, il est nécessaire de les transcrire et de les entrer dans les fichiers lexiques de façon solidaire. Les morphèmes peuvent être solidarisés de plusieurs manières, par l'utilisation de tirets bas, de plus, d'apostrophes et de tirets du 6.

Cependant, l'utilisation des apostrophes et des tirets de 6 n'a pas systématiquement pour effet de signifier au programme MOR que les morphèmes doivent être considérés comme une entité solidaire. En effet, il existe dans le dossier lexique de la grammaire MOR française des fichiers qui permettent au programme MOR d'identifier des formes lexicales transcrites avec des apostrophes ou des tirets du 6 de façon solidarisée comme des unités morphologiques indépendantes :

Le fichier lex « 0affix-v »	<ul style="list-style-type: none">- Contient une liste de pronoms considérés comme des affixes (-t-elle, t-il, - moi, -lui,...)- Ils doivent être transcrits de façon solidarisée à la forme verbale avec le tiret du 6 pour être identifiés par MOR comme des pronoms
Le fichier lex « 0preclict.cut »	<ul style="list-style-type: none">- Contient une liste de morphèmes considérés comme des préclitiques (c', m', j',...).- Ils doivent être transcrits de façon solidarisée avec l'apostrophe pour être correctement identifiés par MOR.

Tableau n°17 : Fichiers lexiques de morphèmes distincts bien que transcrits de façon solidarisée

Le manuel EVAL ne donne pas d'information sur les choix à effectuer en matière de solidarisation des morphèmes. Afin d'uniformiser nos transcriptions, nous avons donc dû établir des règles nous permettant de déterminer quels morphèmes devaient être solidarisés et de quelle façon ils devaient l'être. Pour cela, nous nous sommes basées sur les informations contenues dans le manuel CLAN, le manuel CHAT et sur les fichiers lexiques de la grammaire MOR française qui contenaient des morphèmes solidarisés (MacWhinney, 2000a, 2000b).

Nous avons choisi de transcrire les morphèmes solidarisés de la façon suivante :

Tirets bas	Plus
Adverbes de négation (<i>pas_du_tout, plus_rien, pas_du_tout...</i>)	Adjectif + adjectif (<i>sous+marin, sud+est, demi+sel, tragi+comique,...</i>)
Adverbes de lieu (<i>en_dehors, en_bas...</i>)	Adjectif + prep+nom (<i>tape+à+l'oeil, mal+en+point, mis+en+place...</i>)
Adverbes composés (<i>bien_sûr, de_plus_en_plus, à_nouveau...</i>)	Nom + det + nom (<i>haut+le+coeur,...</i>)
Noms anglais (<i>w_c, rock_n_roll,...</i>)	Nom + nom (<i>petit+déjeuner</i>)
Communicateurs (<i>ça_aussi, ben_oui,...</i>)	Nom + prep + nom (<i>sac+à+dos, boîte+aux+lettres, chasse+d+eau...</i>)
Conjonctions (<i>y_compris, c'est_à_dire,...</i>)	Verbe + nom (<i>entre+tuer, sous+louer, sous+estimer,...</i>)
Déterminants (<i>quelques_uns,...</i>)	
Pronoms (<i>tout_le_monde,...</i>)	
Préposition + article (<i>bien_que, à_moins,...</i>)	
Sigles, acronymes (<i>A_V_C, S_N_C_F,...</i>)	

Tableau n°18: Conventions adoptées pour solidariser les morphèmes 1/2.

Apostrophes	Tirets du 6
Entités morphémiques solidaires s'écrivant avec une apostrophe en français standard (<i>d'ailleurs, d'accord, quelqu'un...</i>)	Inversion sujet-verbe (<i>donnez-moi, que se passe-t-il, ...</i>)
Pré-clitiques (<i>c', m', j', qu'...</i>)	
Dans certains cas, nous avons dû désolidariser certaines expressions en insérant un espace après l'apostrophe afin que MOR n'étiquette pas l'ensemble des expressions comme une entité solidaire. Par exemple, « à côté d'un » a été transcrit « à_côté_d' un », ou « parce qu'il » a été transcrit « parce_qu' il ».	

Tableau n°19: Conventions adoptées pour solidariser les morphèmes 2/2.

Au fur et à mesure de nos transcriptions, nous avons constitué une liste des morphèmes que nous avons choisi de solidariser, celle-ci figure en annexe p. 100-104.

• Codage des répétitions

Nous avons également effectué des choix concernant le codage des répétitions. Nous avons décidé de coder toutes les répétitions, à partir du moment où un même mot était répété au sein d'un même énoncé, que ce soit immédiatement ou à distance de la première occurrence du mot répété, comme dans l'exemple suivant :

***PAR:** okay et ben alors en_fait <y a> [/] d'après ce que j'en comprends y a un chat qui est coincé sur un arbre.

Nous avons également choisi de coder toutes les répétitions, qu'elles soient ou non intentionnelles. Celles visant à marquer un effet de style d'insistance ont ainsi été codées comme répétitions, comme dans les exemples suivants :

<p>*PAR: et elle me dit +"/.</p> <p>*PAR: +" mais non <vas y vas y vas y> [/] vas y +".</p> <p>*PAR: +" faut absolument que tu assistes à ça +".</p> <p>*PAR: je vais à la clinique.</p> <p>*PAR: et puis là <on attend on attend> [/] on attend.</p>

Une fois l'ensemble des données retranscrites, nous avons utilisé les codages précédemment décrits pour annoter nos transcriptions.

4.1.3. Vérification des transcriptions

Une fois les données transcrites et codées, nous avons lancé le programme CHECK afin de vérifier que nos transcriptions respectaient bien les normes de transcription du système CHAT. Nous avons ensuite corrigé chacune des erreurs signalées jusqu'à l'obtention du message « *Success ! No errors Found !* ».

4.1.4. Alignement des transcriptions avec les fichiers audio

Nous avons ensuite procédé à l'alignement de chacune de nos transcriptions avec le fichier audio correspondant. Pour cela, nous avons certaines fois eu besoin d'utiliser le logiciel Audacity. En effet, ce logiciel nous a permis d'extraire les pistes audio des entretiens que nous n'avions qu'en format vidéo ou encore d'isoler les parties concernant le protocole de recueil de discours AphasiaBank des autres épreuves que nous avons faites passer aux participants aphasiques.

4.1.5. Le lancement du programme d'étiquetage morphologique

Avant de pouvoir lancer le programme EVAL sur nos transcriptions, nous avons dû les préparer à l'analyse EVAL en procédant au lancement du programme d'étiquetage morphologique. En effet, EVAL ne peut fonctionner que si les données langagières ont été analysées par le programme MOR et que les catégories morphologiques ont été spécifiées.

Cette phase préparatoire à l'analyse EVAL a donc comporté plusieurs étapes : la vérification de la reconnaissance des mots par le programme MOR, la création d'un nouveau fichier lexique afin d'y insérer les mots non reconnus par MOR et l'exécution du programme MOR.

4.1.5.1. Reconnaissance des mots par MOR

La première étape de préparation à l'analyse EVAL a été l'étape de vérification de la reconnaissance des mots par le programme MOR. En effet, afin que le programme d'étiquetage morphologique puisse déterminer la classe morphologique des mots présents dans les corpus, tous les mots doivent être reconnus par le programme. Pour cela, ils doivent figurer dans les fichiers lexiques de la grammaire MOR.

Nous avons donc lancé, pour chacune de nos transcriptions, la commande permettant de vérifier si tous les mots de la transcription figurent dans les fichiers lexiques (cf annexe p. 19). Nous avons ainsi obtenu pour chaque transcription la liste des mots non reconnus par MOR. Certains mots n'étaient pas reconnus par MOR car ils étaient mal transcrits, d'autres ne figuraient pas dans les fichiers lexiques de la grammaire MOR française.

4.1.5.2. Création d'un nouveau fichier lexique français

Nous avons créé un nouveau fichier lexique qui figure en annexe p. 105, auquel nous avons ajouté tous les mots non reconnus en spécifiant manuellement leur catégorie morphologique, sous la forme suivante :

```
constant {[scat adj]}
```

Dans le cas des mots composés, nous avons spécifié la catégorie morphologique du mot mais également les catégories morphologiques des morphèmes les composant, comme dans les exemples suivants :

```
femme+à+tout+faire {[scat n][gen fem][comp n+prep+adv+v]}  
lampe+de+chevet {[scat n][gen fem][comp n+prep+n]}  
tartes+tatin {[scat n][gen fem][tn pl][comp n+n]}
```

Nous avons effectué le choix de la catégorie morphologique des mots en nous appuyant sur le site du Centre National de Ressources Textuelles et Lexicales créé par le Centre National de la Recherche Scientifique (CNRS) [Site CNRTL].

Certains morphèmes ont été entrés dans le lexique de façon solidarisée, selon les conventions que nous avons précédemment décrites. Les noms et les adjectifs ont été ajoutés au fichier lexique au singulier et les verbes à la forme infinitive. En effet, le programme MOR identifie automatiquement les formes fléchies à partir des règles qui sont spécifiées dans les fichiers grammaire, il n'est donc pas nécessaire d'entrer dans les fichiers lexiques les formes au pluriel ou conjuguées. Cependant, dans certains cas, il a été nécessaire d'ajouter des entrées lexiques au pluriel. En effet, quand la mise au pluriel n'obéit pas aux règles classiques de mise au pluriel du français, le programme MOR ne

parvient pas à identifier automatiquement la forme au pluriel à partir d'une entrée lexicale au singulier. Nous avons rencontré ce problème avec des mots composés dont la mise au pluriel ne se faisait pas uniquement sur le deuxième mot. Ainsi, nous avons dû entrer les formes suivantes au pluriel dans le fichier lexique :

```
belles+soeurs {[scat n][gen f][num pl][comp n+n]}
tartes+tatin {[scat n][gen fem][tn pl][comp n+n]}
sages+femmes {[scat n][gen f][num pl][comp n+n]}
petits+déjeuners {[scat n][num pl][comp n+n]}
```

Au sein de notre nouveau fichier lexique, nous avons classé les mots par catégorie grammaticale, et, au sein de chaque catégorie grammaticale, les mots ont été classés par ordre alphabétique. Nous avons fait ce choix afin de faciliter le travail ultérieur des concepteurs de la base de données pour intégrer les mots de notre fichier à la grammaire MOR française générale.

Nous avons ensuite placé ce fichier dans le dossier lexique de l'application CLAN afin que le programme MOR s'y réfère automatiquement et puisse spécifier la catégorie morphologique de chacun des mots que nous y avons entrés.

La création de ce fichier lexique a permis au programme MOR de reconnaître l'ensemble des mots présents dans nos transcriptions. Nous avons ainsi pu procéder à l'étiquetage morphologique de nos transcriptions, indispensable pour que l'évaluation quantitative et qualitative puisse ensuite être effectuée par le programme EVAL (exemple : calcul du pourcentage d'adverbes).

4.1.5.3. Lancement du programme MOR

Nous avons lancé le programme d'étiquetage morphologique sur chacune de nos transcriptions (cf annexe p. 21). Une ligne supplémentaire intitulée %mor a ainsi été générée sous chaque énoncé dans toutes nos transcriptions en indiquant la catégorie morphologique de chacun des mots en fonction de leur contexte d'utilisation (cf annexe p.21).

4.1.6. Contribution à la base de données AphasiaBank

Afin que le programme EVAL puisse comparer les mesures langagières d'un sujet à un groupe de comparaison, il a fallu procéder à la mise en ligne de nos corpus sur la base de données AphasiaBank.

Nous avons donc transmis nos corpus préparés à l'analyse EVAL, les fichiers audio correspondants, notre fichier lexique, nos tableaux de recueil de données démographiques, notre

tableau de résultats et le protocole AphasiaBank adapté à la langue française aux concepteurs de la base de données.

La mise en ligne de nos données sur le site Aphasiabank a fait l'objet d'une déclaration à la Commission Nationale de l'Informatique et des Libertés (CNIL).

Le site AphasiaBank regroupe des corpus de langues différentes. La base de données générale est donc segmentée en sous-groupes contenant des corpus de la même langue. Ces sous-groupes fonctionnent de manière distincte puisque que les corpus qu'ils contiennent sont comparés uniquement entre eux et non avec les corpus des autres langues du reste de la base de données. Nos corpus ont ainsi été ajoutés à la partie française de la base de données.²² La mise en ligne de nos données, qui correspondent à des transcriptions de mêmes tâches de discours, va ainsi permettre d'effectuer des comparaisons entre nos différents corpus. Les mesures EVAL d'un participant pourront ainsi être situées par rapport à celles des autres participants francophones.

4.1.7. Le lancement du programme EVAL

Nous avons ensuite lancé le programme d'analyse EVAL sur nos transcriptions. Nous avons ainsi obtenu un tableau de résultats basés sur les calculs du programme EVAL.

4.2. Résultats de l'opérationnalisation des données pour l'utilisation d'EVAL

L'opérationnalisation des données pour l'utilisation d'EVAL nécessite plusieurs étapes que nous avons décrites précédemment : la transcription et le codage des données selon les normes CHAT, la vérification des transcriptions, et le lancement du programme MOR. Nous allons donc présenter les résultats obtenus à chacune de ces différentes étapes.

4.2.1. Transcription et codage des données selon les normes CHAT

4.2.1.1. Durée du travail consacré à la transcription et au codage des données

Nous avons comptabilisé la durée du travail consacré à la transcription et au codage pour chacune de nos transcriptions et l'avons mise en rapport avec la durée effective des enregistrements que nous avons retranscrits dans les tableaux suivants. Cette durée ne comprend pas les phases ultérieures de vérification, d'alignement audio et de lancement du programme MOR.

Le premier tableau concerne les transcriptions des sujets contrôles, le second les transcriptions des sujets aphasiques.

²² Lors des comparaisons, EVAL ne tiendra compte que des corpus contenant des tâches de discours issues du protocoleAphasiaBank, et non des corpus hors protocole déjà déposés.

Transcription	Durée du travail consacré à la transcription et au codage	Durée de l'enregistrement retranscrit*
colin01a.cha	3 heures 30 minutes	11 minutes 21 secondes
colin02a.cha	6 heures 25 minutes	18 minutes 15 secondes
colin03a.cha	3 heures 55 minutes	17 minutes 20 secondes
colin04a.cha	3 heures 45 minutes	15 minutes 15 secondes
colin05a.cha	2 heures 50 minutes	7 minutes 43 secondes
colin06a.cha	3 heures	13 minutes 7 secondes
colin07a.cha	5 heures 30 minutes	17 minutes 5 secondes
lemeur01a.cha	5 heures 10 minutes	27 minutes 45 secondes
lemeur02a.cha	4 heures	14 minutes 38 secondes
lemeur03a.cha	2 heures 35 minutes	11 minutes 33 secondes
lemeur04a.cha	2 heures 45 minutes	19 minutes 28 secondes
lemeur05a.cha	3 heures 20 minutes	17 minutes
lemeur06a.cha	1 heure 45 minutes	15 minutes 56 secondes
lemeur07a.cha	3 heures 50 minutes	26 minutes 59 secondes

Tableau n°20 : Durée du travail consacré à la transcription et au codage des données et de l'enregistrement des sujets contrôlés

* La durée de l'enregistrement comprend l'ensemble de l'entretien, y compris les propos de l'examineur

Transcription	Durée du travail consacré à la transcription et au codage	Durée de l'enregistrement retranscrit*
colin09a.cha	4 heures	12 minutes 4 secondes
colin10a.cha	6 heures 15 minutes	14 minutes 9 secondes
colin11a.cha	6 heures	15 minutes 49 secondes
colin13a.cha	3 heures 45 minutes	14 minutes 16 secondes
colin14a.cha	5 heures 50 minutes	18 minutes 24 secondes
colin16a.cha	9 heures 35 minutes	15 minutes 54 secondes
lemeur09a.cha	2 heures 30 minutes	21 minutes 57 secondes
lemeur10a.cha	8 heures 5 minutes	22 minutes 18 secondes
lemeur11a.cha	2 heures	14 minutes 20 secondes
colin12a.cha	4 heures	9 minutes 15 secondes
colin15a.cha	3 heures 5 minutes	29 minutes 20 secondes

Tableau n°21 : Durée du travail consacré à la transcription et au codage des données et de l'enregistrement des sujets aphasiques

* La durée de l'enregistrement comprend l'ensemble de l'entretien, y compris les propos de l'examineur

Dans l'ensemble, 52 heures et 20 minutes ont été consacrées aux transcriptions des sujets contrôlés et 55 heures et 5 minutes aux transcriptions des sujets aphasiques. Ainsi, le temps total consacré à l'étape de transcription et de codage des données a été de 107 heures et 25 minutes.

Bien que les durées moyennes d'enregistrement étaient à peu près équivalentes chez les

sujets contrôles et les sujets aphasiques, la durée moyenne de transcription et de codage était nettement plus importante chez les sujets aphasiques que chez les sujets contrôles.

	Sujets contrôles		Sujets aphasiques	
	Durée de transcription et de codage	Durée de l'enregistrement	Durée de transcription et de codage	Durée de l'enregistrement
Moyenne	3 h 44 min 17 sec	16 min 40 sec	5 h 27 min 16 sec	17 min 4 sec
Minimum	1 h 45 min	15 min 56 sec	2 h	9 min 15 sec
Maximum	6 h 25 min	27 min 45 sec	9 h 35 min	29 min 20 sec
Écart-type	1 h 50 min 26 sec	5 min 30 sec	2 h 22 min 18 sec	5 min 37 sec

Tableau n°22: Comparaison entre les durées de transcription et de codage et d'enregistrement entre sujets contrôles et sujets aphasiques

4.2.1.2. Utilisation des codages et des normes générales de transcription

Nous avons utilisé l'ensemble des codages que nous avons sélectionnés avant de débiter nos transcriptions, soit 16 types de codages différents.²³

4.2.1.3. Utilisation des codages d'erreurs

Le nombre et le type d'erreurs codées au niveau du mot sur l'ensemble des corpus ont été calculés avec la commande CLAN suivante :

```
freq +s\[**\] +t*PAR *.cha
```

Le nombre et le type d'erreurs codées au niveau des énoncés dans l'ensemble des corpus ont été obtenus avec la commande CLAN suivante :

```
freq +s"<+ *>" +tPAR *.cha
```

Les erreurs qui sont incluses au sein d'une reformulation ne sont pas comptabilisées par ces calculs comme des erreurs.

²³ Ces normes sont détaillées dans un tableau p. 96.

Dans l'ensemble, nous avons codé 6 erreurs au niveau du mot dans les corpus des sujets contrôles et 177 dans les corpus des sujets aphasiques. Ces erreurs se répartissent de la façon suivante :

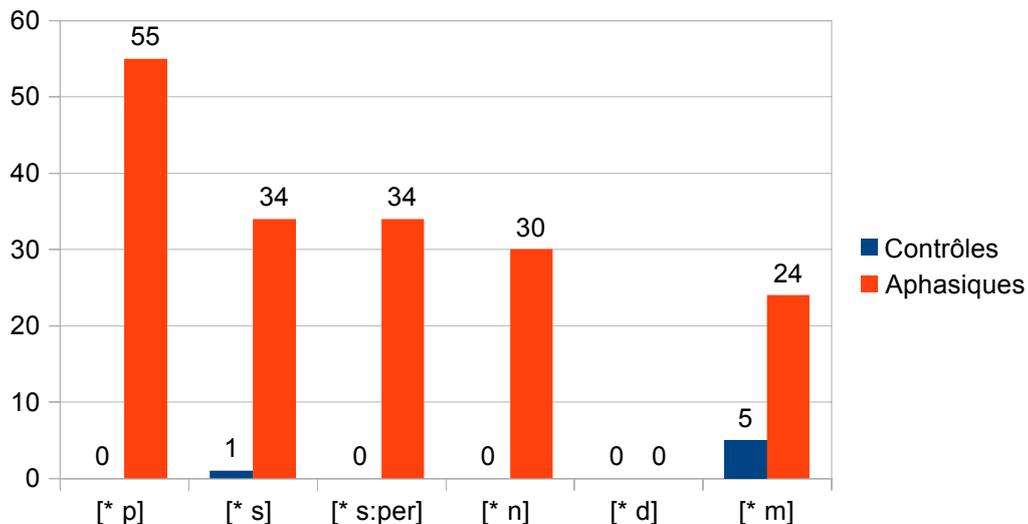


Diagramme n°1 : Répartition des erreurs au niveau du mot dans les corpus des sujets contrôles et aphasiques

[* p] : Erreurs phonologiques

[* s] : Erreurs sémantiques

[* s:per] : Persévérations au niveau du mot

[* n] : Néologismes

[* d] : Disfluences à l'intérieur d'un mot

[* m] : Erreurs morphologiques

Dans l'ensemble, nous avons codé 4 erreurs au niveau de l'énoncé dans les corpus des sujets contrôles et 133 erreurs dans les corpus des sujets aphasiques. Ces erreurs se répartissent de la façon suivante :

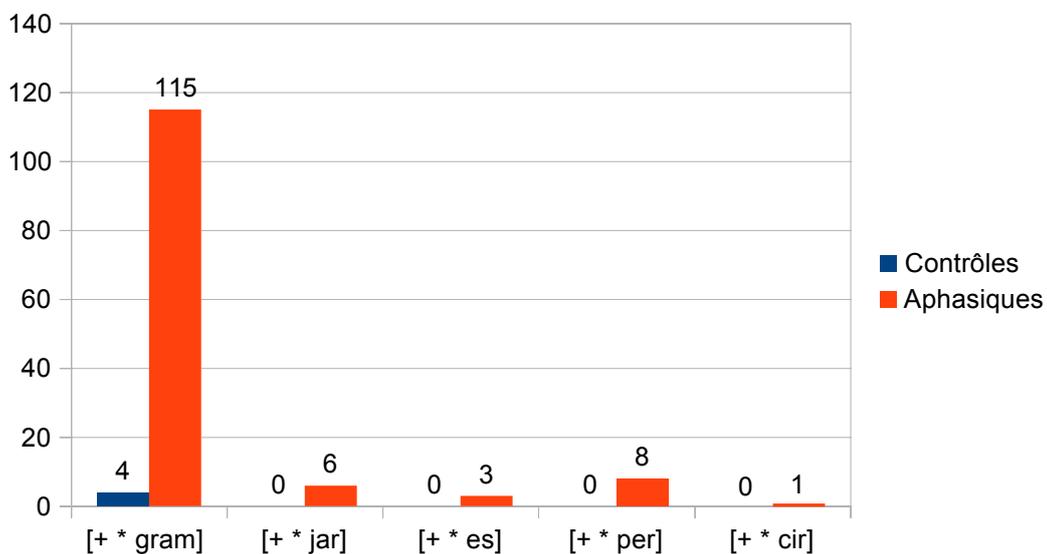


Diagramme n°2 : Répartition des erreurs au niveau de l'énoncé dans les corpus des sujets contrôles et aphasiques

[+ * gram] : Erreurs grammaticales

[+ * jar] : Jargon

[+ * es] : Discours vide

[+ * per] : Persévérations au niveau de l'énoncé

[+ * cir] : Circonlocutions

La répartition détaillée de chacune des erreurs dans les différentes transcriptions figure en annexe p. 110-111.

La comparaison entre le nombre d'erreurs codées pour les participants contrôles et les participants aphasiques montre la répartition suivante :

	Nombre total d'erreurs au niveau du mot	Nombre total d'erreurs au niveau de l'énoncé	Nombre total d'erreurs
Sujets contrôles	6	4	10
Sujets aphasiques	177	133	310

Tableau n°23 : Comparaison du nombre d'erreurs entre les sujets contrôles et les sujets aphasiques

Concernant les erreurs au niveau du mot, en moyenne, nous avons donc codé 0,4 erreur dans une transcription d'un sujet contrôle et 16,09 erreurs dans une transcription d'un sujet aphasique.

Concernant les erreurs au niveau de l'énoncé, en moyenne, nous avons donc codé 0,3 erreur dans une transcription d'un sujet contrôle et 12,09 erreurs dans une transcription d'un sujet aphasique.

4.2.2. Vérification des transcriptions par le programme CHECK

La durée de vérification des transcriptions comprend le lancement du programme CHECK et la correction de chacune des erreurs détectées par le programme.

La vérification d'une transcription durait entre 10 et 20 minutes selon les cas. Dans l'ensemble, au fur et à mesure des transcriptions et de notre maîtrise du logiciel, l'étape de vérification et de correction des erreurs a pris de moins en moins de temps. Nous avons finalement réussi à résoudre l'ensemble des erreurs détectées par le programme CHECK.

Les erreurs relevées par le programme CHECK se rapportent au respect des normes de transcription et de codage du système CHAT. Ces erreurs concernaient principalement la détection d'espaces manquants ou en trop, l'absence ou la redondance de signes de ponctuation en fin d'énoncé, les fautes de frappe ou le refus de certaines combinaisons de symboles (comme la présence de tirets bas après les apostrophes). Un certain nombre d'erreurs concernait le non respect des conventions de codage des erreurs, par exemple la nécessité que les signes <> soient suivis de crochets, que le mot cible de l'erreur soit retranscrit immédiatement après l'erreur ou encore que les codages des erreurs au niveau de l'énoncé soient situés après le signe de ponctuation clôturant un énoncé.

4.2.3. Alignement aux fichiers audio

L'alignement des transcriptions aux enregistrements audio a duré environ 30 minutes par transcription. La durée de l'alignement était généralement supérieure à la durée effective de l'enregistrement audio car la phase d'alignement a souvent conduit à effectuer des corrections de la transcription et certains retours en arrière dans l'écoute du fichier audio ont été nécessaires.

L'étape d'alignement audio est utile car elle permet de vérifier la transcription effectuée en réécoutant l'ensemble du fichier audio et ainsi d'effectuer des corrections permettant d'améliorer la précision de la transcription. Cette étape est également importante dans le cadre de la constitution de la base de données car elle permet aux cliniciens et aux chercheurs membres d'AphasiaBank d'avoir accès à la transcription et à l'audio correspondant. Cependant, l'alignement audio prend du temps et n'est pas indispensable à l'analyse EVAL, bien qu'il permette d'obtenir la durée précise de chaque Gem.

4.2.4. L'étape d'étiquetage morphologique

Comme nous l'avons précédemment décrit, l'étape d'étiquetage morphologique a nécessité plusieurs étapes : l'étape de vérification de la reconnaissance des mots par MOR, la création d'un nouveau fichier lexique et le lancement effectif du programme MOR.

4.2.4.1. Reconnaissance des mots par MOR et création d'un nouveau fichier lexique

La durée de la vérification de la reconnaissance des mots par MOR comprend le lancement de la commande permettant d'identifier les mots non reconnus dans la transcription (cf annexe p. 19), la correction des mots mal transcrits et l'ajout des mots non reconnus dans le fichier lexique en spécifiant leurs catégories morphologiques. Le temps le plus important de cette étape a été consacré à la recherche des catégories morphologiques appropriées à chaque entrée lexique.

En moyenne, le temps consacré à cette étape était de 1 heure. Ce temps a cependant diminué au fur et à mesure des transcriptions.

Nous avons indiqué dans le tableau suivant le nombre de mots non reconnus par MOR. Cela comprend les mots mal transcrits et les mots qui ne faisaient pas partie des fichiers lexiques. Nous avons précisé dans la colonne suivante le nombre de mots qui n'ont pas été reconnus par MOR car ils ne figuraient pas dans les fichiers lexiques et qui ont donc été ajoutés au nouveau fichier lexique.

	Nombre total de mots non reconnus par MOR	Nombre total de mots qui ont dû être ajoutés au nouveau fichier lexique
Transcriptions contrôles	315	179
Transcriptions aphasiques	113	44
TOTAL	428	223

Tableau n°24 : Nombre total de mots non reconnus par MOR et nombre total de mots qui ont dû être ajoutés au nouveau fichier lexique²⁴

Ainsi, sur l'ensemble des mots non reconnus par MOR, une partie seulement ne figurait pas dans les fichiers lexiques et a dû y être ajoutée. Les autres mots n'ont pas été reconnus pour les raisons suivantes essentiellement :

- fautes de frappe ou erreurs d'orthographe ;
- transcription sous une forme différente que celle de l'entrée figurant dans le lexique (par exemple la transcription de la forme « ok » au lieu la forme « okay », ou de « demi_soeur » au lieu de « demi+sœur »,...) ;
- erreurs de segmentation (mots solidarisés qui n'auraient pas dû l'être ou inversement ou mots segmentés de la mauvaise façon) ;
- non mots ou conduites d'approches non exclues ;
- présence d'un espace après une apostrophe dans des formes lexicales préclitiques (c', m',...);
- non précision du mot cible entre parenthèses dans les abréviations ;
- non notation du fait que certains mots sont des sigles (exempleA_V_C).

²⁴ La répartition détaillée du nombre de mots non reconnus par MOR et de mots qui ont dû être ajoutés au nouveau fichier lexique par transcription figure en annexe p. 112.

4.2.4.2. La constitution du nouveau fichier lexique

Au total, nous avons dû ajouter 223 mots à notre fichier lexique. Ces mots appartenait à 10 catégories morphologiques différentes, se répartissant de la façon suivante :

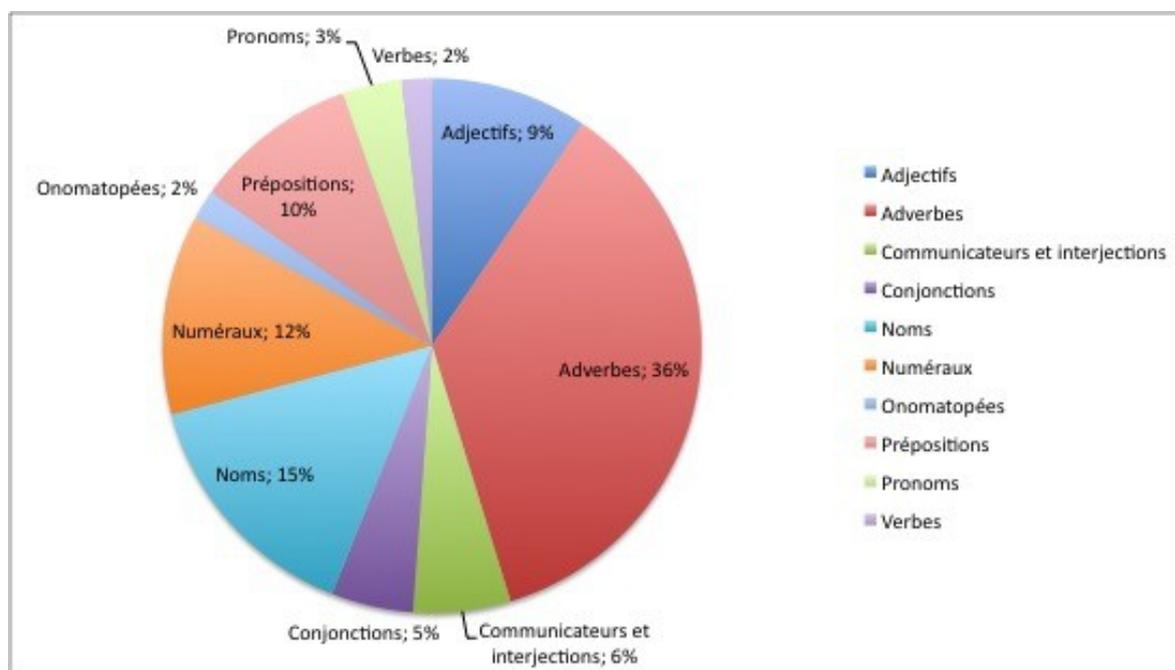


Diagramme n°3 : Pourcentage des différentes catégories morphologiques des mots entrés dans le fichier lexique

Nous avons transmis notre fichier lexique aux concepteurs de la base de données. Ils ont intégré la plupart des mots de notre fichier aux fichiers lexiques de la grammaire française, à l'exception de certaines combinaisons de morphèmes que nous avons choisi de solidariser. Les concepteurs ont en effet estimé que certaines entités devaient être traitées en entités distinctes plutôt qu'en entités solidaires.

4.2.4.3. Le lancement du programme d'étiquetage morphologique

L'ensemble des mots présents dans nos corpus étant désormais reconnu par le programme MOR, nous avons donc pu lancer le programme d'étiquetage morphologique et générer la ligne %mor sur l'ensemble de nos transcriptions.

Toutes les étapes préalables de préparation des corpus et d'adaptation du logiciel CLAN permettant de lancer le programme EVAL ont donc été faites.

4.2.5. La mise en ligne de nos corpus sur la base de données

Nos transcriptions et les fichiers audio correspondants ont ainsi été mis en ligne le 7 mai 2016, dans le strict respect de l'anonymat des participants. Ils sont accessibles aux cliniciens et

aux chercheurs membres d'AphasiaBank par un identifiant et un mot de passe.

4.2.6. Le lancement du programme EVAL

Nous avons pu lancer le programme EVAL sur nos transcriptions mais celui-ci n'est pas encore opérationnel pour la langue française.

En effet, les résultats obtenus par l'analyse EVAL ne sont pas tous exploitables car certains calculs ne sont pas encore totalement opérationnels. De plus, le logiciel CLAN doit être reconfiguré pour permettre au programme EVAL de prendre en compte les autres corpus de la base de données française afin qu'EVAL puisse effectuer des comparaisons entre les différentes transcriptions.

Les concepteurs de la base de données nous ont informé que le programme EVAL n'était en effet pas encore opérationnel pour la langue française et que les informaticiens du projet AphasiaBank allaient prochainement travailler sur cette reconfiguration du programme.²⁵²⁶

25 Mail de Brian MacWhinney (un des concepteurs du projet AphasiaBank) en date du 8 mai 2016 : « Il est possible de reconfigurer EVAL pour le français. Cependant, pour ce faire, L. S. va devoir changer plusieurs choses dans les programmes et il est actuellement très occupé à régler un (autre) problème (...) Lorsque ce travail sera fait, j'espère dans quelques semaines, nous pourrons revenir à cette question. » *(traduit de l'anglais au français)*

26 Mail de Brian MacWhinney en date du 12 mai 2016 : « Les prochaines étapes pour la création d'EVAL pour le français impliquent toutes des détails techniques concernant des modifications du programme. Vous n'avez rien à faire de particulier de votre côté. » *(traduit de l'anglais au français)*

PARTIE 4 : DISCUSSION DES RÉSULTATS

1. INTERPRÉTATION DES RÉSULTATS ET VALIDATION OU INVALIDATION DES HYPOTHÈSES

1.1. Hypothèse 1 : Le protocole AphasiaBank est adaptable à la langue française

Nous avons postulé que l'ensemble du protocole de recueil de données AphasiaBank était adaptable à la langue française. Ce protocole contient les documents permettant le recueil de données démographiques, le protocole de recueil de données discursives et enfin le protocole de recueil de données langagières issues des épreuves complémentaires à faire passer aux participants aphasiques. Chaque élément fait l'objet d'une sous-hypothèse que nous allons à présent discuter.

1.1.1 Sous-hypothèse 1 : Le protocole de recueil de données démographiques AphasiaBank est adaptable à la langue française

1.1.1.1. Synthèse des résultats

L'adaptation du formulaire a nécessité quelques changements du fait des particularités culturelles propres à chaque pays.

L'adaptation a donc pu être réalisée avec des améliorations ou des précisions qui étaient nécessaires, comme pour la question concernant les années d'étude.

Le tableau de recueil de données démographiques des participants contrôles a été adapté sans modification, ce qui nous a permis de rester fidèle à la version originale.

Le tableau de recueil de données démographiques des participants aphasiques n'a pas été adapté tel quel puisque nous avons ajouté trois colonnes et supprimé une colonne concernant la sélection du type d'aphasie selon le système de classification de Luria. Nous avons également traduit le terme de « semantic dementia » par « variante sémantique de l'aphasie primaire progressive ».

1.1.1.2. Discussion, intérêts et limites

En ne modifiant pas beaucoup d'éléments, nous avons ainsi pu suivre notre objectif qui était de rester au plus proche de la version originale. Cela présente l'intérêt d'uniformiser au maximum le recueil de données.

Cependant, la modification de certaines formulations nous a aussi permis de nous situer au plus près de la réalité clinique française.

Nous avons adapté le questionnaire dans l'optique d'obtenir des données les plus précises possibles afin que le recueil reflète la réalité exacte. Ce qui nous a semblé confus dans la version originale a donc été précisé. Ce travail d'adaptation a été le moyen de mettre en évidence certaines améliorations à apporter au protocole, qui sont essentielles selon nous. Il pourrait ainsi être intéressant de soumettre ces observations et remarques aux concepteurs.

Nous nous sommes ainsi éloignées du protocole original pour certaines questions mais cela constituait selon nous un aspect inhérent au travail d'adaptation aux particularités culturelles françaises.

1.1.1.3. Conclusion

L'ensemble du protocole de recueil de données a pu être adapté, **la sous-hypothèse 1 est donc validée.**

1.1.2 Sous-hypothèse 2 : Le protocole de recueil de données discursives AphasiaBank est adaptable à la langue française

1.1.2.1. Synthèse des résultats

L'adaptation du protocole de discours d'AphasiaBank destiné aux participants contrôles et aphasiques a pu être effectuée en gardant tous les éléments présents dans la version anglo-Saxonne.

Cependant, en choisissant par exemple de traduire « *can you tell me a story about something important that happened to you in your life?* » par « *Pourriez-vous me raconter un événement marquant de votre vie ?* » nous n'avons pas respecté le script original en ne reprenant pas le terme « important ».

Nous avons également remplacé certains termes comme le titre original de la section II « Picture Description » par « Discours narratif à partir d'images ». Il nous a en effet semblé important de préciser l'intitulé de l'épreuve car la tâche demandée aux participants est bien celle d'un récit à partir d'une image.

Concernant le matériel proposé, nous avons essayé de choisir le livre de Cendrillon qui soit le plus similaire possible à celui proposé dans la version originale.

1.1.2.2. Discussion, intérêts et limites

Concernant les **instructions** du protocole d'épreuves de discours, lors de la traduction de certains termes, nous nous sommes parfois éloignées de la version originale. Certains éléments n'avaient plus de sens une fois traduits de façon littérale en français. Nos choix concernant la traduction de certains termes ont donc été motivés par la volonté de proposer un protocole qui ait du sens et qui soit adapté au français.

Dans l'optique de rester fidèle au protocole d'origine, peut-être aurait-il fallu reprendre les mêmes formulations. Nous nous situons cependant dans le cadre d'une première traduction et des améliorations futures devront probablement être apportées, par le biais d'une évaluation de la pertinence de la traduction par exemple.

Sur le choix du sandwich jambon beurre, nous ne nous sommes pas éloignées du protocole original car il était bien précisé dans la version originale que les utilisateurs du protocole originaires d'autres pays pouvaient choisir un autre type de sandwich.

L'adaptation du protocole a d'abord présenté quelques difficultés liées au fait que certains **documents** n'étaient pas disponibles sur le site AphasiaBank. C'est le cas de la carte oui/non, des photos du sandwich au beurre de cacahuète et la référence du livre de Cendrillon. Nous avons donc dû demander ces informations directement aux concepteurs du protocole.

On pourrait également se demander si le **choix concernant le matériel** nécessaire à la passation du protocole était adapté. Le livre choisi pour la narration de l'histoire de Cendrillon n'est pas totalement similaire à celui proposé dans la version originale. Nous aurions pu choisir de proposer le livre utilisé dans la version américaine. Cependant ce livre n'étant pas facilement accessible en France, il aurait été moins facile pour les cliniciens ou les chercheurs qui souhaiteraient, à terme, utiliser le protocole de se le procurer. Trouver un livre totalement similaire en français n'était pas possible. Nous avons donc essayé de choisir un livre qui s'en rapprochait le plus, même si ceci constitue une limite à l'adaptation stricte du protocole original.

La carte OUI/NON, utilisée pour fournir une aide aux participants aphasiques, que nous avons créée nous-mêmes a été validée par les concepteurs du protocole.

Pour l'aide de l'épreuve du discours procédural, nous avons du choisir des photos correspondant au sandwich jambon beurre. Nous avons choisi d'acheter des photos libres de droits afin que n'importe quel clinicien puisse se les procurer. Cependant, les photos du protocole original n'étant pas disponibles, nous n'avons pas pu vérifier l'adéquation de celles que nous avons proposées avec celles-ci.

Ainsi, si tous les éléments présents dans les versions originales destinées aux participants contrôles et aphasiques ont pu être repris lors de l'adaptation du protocole, certaines améliorations peuvent encore être apportées au protocole que nous avons adapté. Nous avons dû réaliser quelques modifications qui, selon nous, sont nécessaires lorsque l'on effectue un travail d'adaptation. En effet, il nous semble que l'intérêt du travail d'adaptation réside également dans la capacité à se détacher du protocole original pour savoir s'adapter à la réalité clinique du pays où il est proposé.

1.1.2.3 Conclusion

L'ensemble du protocole de recueil de données discursives a pu être adapté, **la sous-hypothèse 2 est donc validée.**

1.1.3.Sous-hypothèse 3 : Le protocole de recueil de données issues des épreuves complémentaires AphasiaBank est adaptable à la langue française

1.1.3.1 Synthèse des résultats

Concernant l'adaptation des épreuves complémentaires, notre objectif était, dans la mesure du possible, de rester au plus proche des tests initialement proposés.

Nous n'avons cependant pas traduit les tests du protocole original. Il nous a en effet semblé que la simple traduction d'un test standardisé et validé dans une autre langue n'aurait pas de validité en français et présenterait de ce fait, peu d'intérêt.

De plus, il était nécessaire d'adapter les tests à la réalité clinique française. Dans une perspective d'utilisation future du protocole par les cliniciens français, nos choix ont été motivés par le fait qu'il fallait proposer, tant que cela était possible, des épreuves déjà connues et utilisées par les orthophonistes français, dans un souci de facilité d'utilisation, d'accessibilité et de connaissance des tests proposés.

Nous allons maintenant discuter de l'adéquation de l'adaptation proposée pour chaque test complémentaire du protocole AphasiaBank.

1.1.3.2.Discussion, intérêts et limites

• **Adaptation de l'*AphasiaBank Repetition Test*** (Holland et al.,2007)

Nous aurions pu choisir de traduire l'*AphasiaBank Repetition Test* mais cela ne nous a pas paru opportun. En effet, la traduction d'un test nécessite une réflexion aboutie, pour laquelle nous ne disposons pas de suffisamment de temps. Il pourrait cependant être intéressant d'adapter ce test au français car celui-ci présente de nombreuses spécificités que nous n'avons pas retrouvées dans des tests francophones. Ceci pourrait faire l'objet d'un autre mémoire.

Concernant l'épreuve de répétition de phrases, pour rester proche du test original, il aurait été intéressant de trouver une épreuve de span de mots et de phrases. Dans un premier temps, nous n'avons pas trouvé d'épreuve de ce type. Plus tard, après que les passations du protocole aient été réalisées, nous avons trouvé une épreuve semblable dans le *BIA* intitulée « Empan de Danneman » (Gatignol et al., 2012).

Pour l'adaptation de cette épreuve de répétition, nous nous sommes donc beaucoup éloignées du test original, en proposant deux épreuves différentes alors qu'une seule est proposée dans le test original. Ainsi, le type d'épreuve, les items proposés et la façon dont s'organise l'épreuve que nous avons proposée, diffèrent complètement de l'épreuve proposée dans le test

original. La limite principale réside dans le fait que nous n'obtenons pas du tout les mêmes scores et profils que dans les épreuves originales qui calculent un empan de mots, l'épreuve de répétition de mots et de phrases de l'*HDAE* évaluant simplement la mémoire auditivo-verbale sans calculer d'empan.

Notre adaptation présente cependant l'intérêt de proposer des épreuves déjà utilisées par les cliniciens, validées et standardisées en français. On pourrait néanmoins discuter de la pertinence de cette adaptation concernant l'adéquation des épreuves proposées à celle du protocole original, duquel nous avons dû beaucoup nous éloigner et dont les résultats obtenus sont trop éloignés de ce qui est recherché dans le test original.

- **Adaptation du *Verb Naming Test*** - tiré de *Northwestern Assessment of Verbs and Sentences-Revised*, Field Test Version (Thompson, 2011, cité sur le site AphasiaBank)

Le test équivalent que nous avons choisi est le *DVL 38*.

Concernant le nombre d'items, celui-ci est plus élevé dans le *DVL 38*. Ceci peut constituer une limite, car cela rend la durée de passation du test plus longue et risque d'engendrer une fatigue supplémentaire chez le participant. Nous aurions ainsi pu choisir de ne proposer que vingt-deux verbes à dénommer comme dans le *VNT*, cependant il aurait été compliqué de choisir les items appropriés et nous n'aurions pas pu calculer d'écart-types à l'issue de la passation du test.

Une épreuve de dénomination de verbes contenant le même nombre d'items que dans le test original existe, issue du *BIA*, mais nous nous en sommes aperçues qu'après la réalisation des passations. Il pourrait être opportun d'étudier la possibilité de la proposer à la place du *DVL 38*.

Une autre limite concerne les paramètres utilisés pour calculer les résultats du participant. Le *DVL 38* propose des sous-scores relatifs à la fréquence lexicale des verbes alors que le *VNT* classe les verbes selon leur structure argumentale. Cependant, la prise en compte de ce type de paramètres n'existe pas en français, et nous ne pouvons pas simplement traduire ce test, ceci relevant d'un travail approfondi de sélection des verbes appropriés.

Pour l'adaptation de ce test, la limite principale concerne donc les sous-scores obtenus qui ne correspondent pas à ce qui est proposé dans le test original.

L'intérêt de ce test réside cependant dans le fait qu'il est connu et très utilisé en clinique orthophonique française.

- **Adaptation du *Boston Naming Test*** - seconde édition, version courte (Kaplan et al., 2001, cité sur le site AphasiaBank)

Pour adapter le *Boston Naming Test* (BNT) nous avons choisi l'épreuve de dénomination du *MT86*. Nous aurions pu choisir de proposer la version longue française qui est une adaptation de la version anglaise. Cependant le nombre d'items proposés nous a paru trop élevé.

Les limites concernant l'adaptation de ce test concernent le nombre d'items présentés dans le *MT86* qui diffère de celui proposé dans le test original.

Elles concernent également la cotation qui est effectuée avec un simple score pour le *MT86* alors qu'elle est effectuée avec un écart-type dans le test original.

Cependant, l'utilisation de l'épreuve du *MT86* dans une version raccourcie, présente l'intérêt de ne proposer que des objets, comme c'est le cas dans le test original, ce qui nous permet de nous situer au plus proche du test original.

- **Adaptation de l'*Aphasia Quotient*** - tiré de la *Western Aphasia Battery-Revised* (Kertesz, 2007, cité par MacWhinney et al., 2011)

Il n'existe pas d'équivalent de la *WAB* en français. Les choix que nous avons faits concernant l'adaptation de cette épreuve comportent donc certaines limites car nous n'étions pas en mesure de proposer un équivalent français. Nous devons ainsi, pour l'adaptation de cette épreuve, nous référer à la pratique clinique en France, en nous basant sur une échelle de gravité utilisée couramment et sur le diagnostic clinique du thérapeute. Ceci nous a ainsi conduit à nous éloigner de notre objectif principal, qui était de rester au plus proche du protocole original.

- **Adaptation du *Complex Ideational Material*** - Version courte – tiré du *Boston Diagnostic Aphasia Examination* (Goodglass et al., 2001, cité sur le site *AphasiaBank*)

Nous sommes restées au plus proche du protocole original, en utilisant le même test que celui proposé dans la version anglaise, et non celui proposé dans la version française qui comporte plus d'items. Pour respecter la cohérence du test conçu en français, nous aurions pu proposer l'épreuve de logique et de raisonnement dans son intégralité. Cependant la cotation se faisant à l'aide d'un score, le choix de ne pas conserver la totalité de l'épreuve française présente l'intérêt d'être strictement identique à l'épreuve proposée dans le test original.

- **Adaptation du *Sentence Comprehension*** - Adapté de *Philadelphia Comprehension Battery* (Saffran et al, 1988, cité sur le site *AphasiaBank*)

Nous avons choisi de ne présenter que les vingt premiers items de l'épreuve du *MT86* afin de respecter le nombre d'items du test anglo-saxon ainsi que l'ordre de présentation des items des phrases du *MT86*. Nous avons donc décidé de ne pas faire passer la version française dans sa totalité. Nous avons en effet pensé que cela aurait été trop fatiguant pour le patient, au vu de toutes les épreuves qu'il avait déjà réalisées. Cela nous a également permis de rester fidèle à la version d'origine en proposant le même nombre d'items. Cependant, la progression et la cohérence d'un test est étudiée dans sa globalité et le fait de ne proposer que vingt items sur les trente-huit peut être sujet à discussion quant à la pertinence de proposer ce test sous cette forme.

De plus, nous proposons douze phrases actives dans l'épreuve du *MT86*. Nous aurions pu choisir de ne garder que deux phrases actives comme c'est le cas dans le test original. Cela aurait pu permettre de diminuer le nombre d'items, afin de ne pas générer une fatigue supplémentaire pour le patient. Cependant si nous avons fait ce choix, il nous aurait été difficile de sélectionner les deux phrases actives à garder parmi toutes celles du *MT86*, et nous n'aurions alors plus respecté l'ordre de présentation des items conçu par les concepteurs du *MT86*.

1.1.3.3. Conclusion

Lors de l'adaptation des tests complémentaires, il s'est révélé difficile de prendre en compte à la fois l'exigence de fidélité au protocole et celle de proposer des tests connus des cliniciens français. Notre choix de départ était cependant de rester le plus proche possible du protocole original. Nous ne disposons donc pas d'étalonnage permettant de situer le patient pour chaque épreuve, ce qui peut constituer une limite à l'utilisation des tests tels qu'ils sont présentés pour le moment, dans une perspective d'utilisation en clinique orthophonique. Il est cependant possible de comparer les résultats des patients entre eux pour chaque épreuve.

D'autre part, nous avons choisi de ne pas traduire certains tests, de raccourcir certaines épreuves ou encore d'en remplacer certaines par des épreuves quelque peu éloignées de celles présentes dans la version originale. Nous avons cependant pu proposer un équivalent pour chaque épreuve. **La sous-hypothèse 3 est donc validée.**

L'ensemble du protocole AphasiaBank a pu être adapté à la langue française. L'hypothèse 1 est donc validée.
--

1.2. Hypothèse 2 : Le protocole AphasiaBank est utilisable avec des sujets contrôles et des sujets aphasiques francophones

Une fois les éléments du protocole adaptés, nous nous sommes demandées si chaque élément du protocole était concrètement utilisable auprès de participants contrôles et de participants aphasiques. Chaque partie du protocole fait donc l'objet d'une sous-hypothèse concernant sa possibilité d'utilisation, que nous allons à présent discuter.

1.2.1. Sous-hypothèse 1 : Le protocole de recueil de données démographiques AphasiaBank est utilisable avec des sujets contrôles et des sujets aphasiques francophones

1.2.1.1. Synthèse des résultats

Le remplissage du formulaire par les participants contrôles a nécessité des reformulations ou des précisions orales suite aux questions qui ont été posées par les participants.

Dans le formulaire concernant les participants aphasiques, nous avons dû rechercher des informations supplémentaires en dehors des informations demandées dans le questionnaire puisque le remplissage du tableau de recueil de données démographiques nécessitait la mention de ces informations. Nous avons fait le choix de ne pas demander la totalité de ces informations dans le questionnaire mais de les rechercher par nous-mêmes dans le dossier médical et orthophonique du patient.

1.2.1.2. Discussion, intérêts et limites

Si les **participants contrôles** ont pu répondre à toutes les questions du formulaire, celui-ci pourrait tout de même faire l'objet de précisions et d'améliorations au niveau des questions. En effet, la taille de l'échantillon de la population contrôle à laquelle nous avons soumis le questionnaire était relativement peu élevée, ce qui nous a permis de prendre du temps pour vérifier l'exactitude des informations contenues dans le formulaire et de répondre aux questions des participants. Cependant, il serait important de modifier le questionnaire afin que celui-ci soit plus clair pour les participants, ce qui constituerait un gain de temps pour l'examineur qui n'aurait pas besoin de vérifier ces informations pour chaque participant.

D'autres questions ne nous ont pas paru assez précises quant à l'information recherchée. Par exemple, la question « *Ressentez-vous des signes pouvant indiquer une humeur dépressive ?* » ne nous a pas paru assez adaptée pour évaluer la dépression, la réponse du participant ne reflétant pas toujours la réalité concernant la présence d'un état dépressif. On peut alors se poser la question de la fiabilité des informations contenues dans ce questionnaire. La passation des tests complémentaires proposés aux participants contrôles présente donc un intérêt. Si nous avons choisi de ne pas les faire passer, nous aurions pu les utiliser lorsque nous n'étions pas sûres de la fiabilité de la réponse d'un participant. Ces tests ont donc tout leur intérêt au sein du protocole de recueil de données destiné aux participants contrôles, et pourraient également être intégrés au protocole de recueil de données français.

Concernant le formulaire de recueil de données à destination des **participants aphasiques**, la recherche des informations supplémentaires nous a demandé du temps et de la disponibilité. Pour certains participants nous n'avons d'ailleurs pas pu retrouver toutes les informations

demandées. En effet, les dossiers médicaux et orthophoniques sur lesquels nous nous sommes appuyées pour recueillir les informations n'étaient parfois pas assez détaillés. Cependant il ne nous semble pas pertinent de rajouter des questions supplémentaires au formulaire afin de ne pas surcharger le patient.

D'autre part, dans le cas des participants aphasiques, il est demandé dans le tableau de recueil de données démographiques de préciser le type d'aphasie du sujet. Si le type d'aphasie n'est pas connu, l'examineur peut simplement entrer l'information « Aphasia » à la place du type d'aphasie. Cependant la mention du type d'aphasie est importante pour permettre au programme EVAL d'effectuer des comparaisons adéquates entre les différents corpus aphasiques.

Or la détermination du type d'aphasie n'est pas toujours évidente. Les cliniciens américains disposent de certains outils pour les aider dans cette tâche. En effet, la *Western Aphasia Battery-Revised (WAB)* (Kertész, 2007, cité par MacWhinney et al., 2011) permet à partir de différentes épreuves de déterminer le type d'aphasie. Il n'existe pas d'outil équivalent en France où la détermination du type d'aphasie se fait essentiellement de façon clinique. Or il est, dans certains cas, relativement difficile de catégoriser le trouble d'un patient en référence à la taxonomie classique. Dans le cadre de notre étude, certains orthophonistes nous ont fait part de leur difficulté à déterminer précisément le type d'aphasie de leur patient.

1.2.1.3. Conclusion

Le protocole de recueil de données démographiques AphasiaBank est utilisable avec des sujets contrôles et des sujets aphasiques francophones, **la sous-hypothèse 1 est donc validée.**

1.2.2. Sous-hypothèse 2 : Le protocole de recueil de données discursives AphasiaBank est utilisable avec des sujets contrôles et des sujets aphasiques francophones

1.2.2.1 Synthèse des résultats

Concernant le temps de passation du protocole, celui-ci a duré en moyenne 16 minutes et 40 secondes pour les participants contrôles et 17 minutes et 04 secondes pour les participants aphasiques. Les passations ont pu se dérouler en une seule fois pour la totalité des participants.

Si tous les participants contrôles ont acceptés d'être filmés, seuls huit participants aphasiques ont accepté. Les conditions de passation n'ont pas toujours été optimales du fait de la présence de personnes tierces ou du lieu de passation choisi.

Concernant les consignes, certaines ont dû être précisées oralement et certaines épreuves ont posé beaucoup de difficultés aux participants aphasiques, comme les épreuves de discours spontané notamment, auxquelles certains participants aphasiques n'ont pas été en mesure de répondre.

1.2.2.2. Discussion, intérêts et limites

• **Durée de la passation**

Le protocole de recueil de données discursives présente l'intérêt de pouvoir être réalisé en une seule fois, comme cela a été le cas pour tous les participants. La durée moyenne de passation des épreuves auprès des participants contrôles nous semble raisonnable puisqu'elle est de 16 minutes et 40 secondes. Cependant, le temps de passation peut être très variable d'un participant à l'autre. En effet, le temps de passation va de 7 minutes et 43 secondes pour la plus courte et jusqu'à 27 minutes et 45 secondes pour la plus longue. Cette grande variabilité peut s'expliquer par la longueur des réponses du participant, qui peut choisir ou non de détailler ses réponses aux questions.

Pour les participants aphasiques, la durée moyenne d'une passation ne diffère pas beaucoup de la durée moyenne des participants contrôles, puisqu'elle est de 17 minutes et 04 secondes. On pourrait se demander pourquoi cette durée diffère si peu alors même que les participants aphasiques présentent des troubles que l'on ne retrouve pas chez les participants contrôles.

Ceci pourrait s'expliquer par plusieurs raisons : les participants contrôles ont davantage détaillé leurs réponses aux différentes épreuves, leur récit étant davantage fourni en détails et précisions. Les participants aphasiques ont quant à eux moins détaillé leurs réponses, cependant certains participants ont réalisé beaucoup de pauses dans leur discours, ce qui a allongé le temps de passation. D'autres ont éprouvé des difficultés à comprendre la consigne, que nous avons dû réexpliquer plusieurs fois. Si leur récit était, pour la plupart, moins détaillé, la durée de la passation s'explique par la nature des troubles présentés par les participants aphasiques qui ont davantage eu besoin d'être sollicités pour effectuer les différentes épreuves de discours.

Comme dans le cas des participants contrôles, il existe une grande variabilité dans la durée de passation du protocole avec les participants aphasiques, puisqu'elle va de 9 minutes et 15 secondes à 29 minutes et 20 secondes .

Il est possible que le type et la sévérité des troubles présentés par le participant aphasique puissent être un facteur de variabilité de la durée de la tâche.

• **Les conditions de passation**

L'enregistrement au format vidéo n'a pas posé de problème pour les participants contrôles. En revanche, trois participants aphasiques ont refusé d'être filmés. L'enregistrement au format vidéo présente l'intérêt d'apporter des indications supplémentaires de nature qualitative au clinicien qui souhaite analyser le discours du participant aphasique. Toutefois, le protocole de recueil de données discursives est tout à fait utilisable avec un enregistrement audio.

Les différents événements survenus lors de la passation, ont montré que l'environnement dans lequel se déroule la passation est primordial. Ainsi, la façon dont se sont déroulées certaines

passations présente certaines limites, telles que l'intervention d'un tiers ou la présence de la radio en bruit de fond, qu'il faudra veiller à éviter au maximum. En effet, si la passation n'est pas réalisée dans un endroit calme et silencieux, cela peut d'une part perturber le participant et, d'autre part, altérer la qualité de l'enregistrement et rendre ainsi plus difficile le travail de transcription des données.

• **L'utilisation du protocole de recueil de données discursives**

▪ **Passation auprès des participants contrôles**

Les participants contrôles ont pu effectuer toutes les épreuves sans difficulté. Quelques précisions ont été demandées pendant la passation, ce qui pourrait s'expliquer par la formulation de certaines consignes qui ne sont peut-être pas assez claires. Dans une perspective d'amélioration du protocole, il sera donc peut-être nécessaire de modifier certaines consignes ou de changer certains termes qui ne sont peut-être pas adaptés, comme le choix du terme « événement marquant » qui a pu perturber plusieurs participants. Peut-être pourrions nous être plus précis et demander aux participants de nous raconter un souvenir de voyage. Ceci impliquerait moins le participant au niveau émotionnel et pourrait faciliter le récit.

Les participants contrôles connaissant pour la plupart l'histoire de Cendrillon, il semble que le choix de ce conte soit approprié et corresponde aux habitudes culturelles françaises.

▪ **Passation auprès des participants aphasiques**

❖ **Ordre des épreuves**

La passation du protocole auprès des participants aphasiques a été réalisable bien qu'elle ait nécessité beaucoup d'adaptations pour certains participants. En effet nous avons dû modifier l'ordre de présentation des épreuves pour deux patients, pour lesquels nous avons jugé que l'ordre prévu initialement générerait une trop grande fatigue. Il est donc possible d'utiliser le protocole en modifiant l'ordre des épreuves, ceci étant précisé dans le protocole original.

❖ **Réalisation des épreuves**

Pour certains participants aphasiques, la réalisation de certaines épreuves n'a pas été possible, notamment les épreuves de discours spontané. Ceci s'explique peut-être par le fait qu'il s'agit de questions très larges, sans support sur lequel s'appuyer. Le fait de commencer par des questions à propos de leur accident et de leur troubles n'est peut-être pas approprié et l'on pourrait se demander dans quelle mesure ceux-ci ne sont pas déstabilisés par la nature de ces questions.

Le questionnaire d'aide, qui permet de fournir une aide supplémentaire au participant si celui-ci est en difficulté, nous a été très utile puisque nous avons dû l'utiliser neuf fois. Ce questionnaire présente ainsi un grand intérêt au sein du protocole, en proposant, après plusieurs niveaux d'incitations, de terminer l'administration de la tâche de façon plus simple si le patient ne peut pas

répondre à la question posée.

❖ **Commentaires verbaux hors script**

Il n'a pas non plus toujours été possible de respecter le script prévu initialement. En effet, suivant la sévérité et le type de troubles présentés par le participant aphasique, nous avons dû faire beaucoup de commentaires hors protocole afin de canaliser le participant ou bien le relancer. Le script du protocole destiné aux participants aphasiques présente donc des limites concernant les incitations à fournir et le respect strict de ce protocole semble difficilement réalisable avec certains participants aphasiques. Nous nous sommes par ailleurs autorisées à préciser le sens de lecture des images pour l'épreuve de narration à partir d'images. Nous nous sommes donc éloignées du script sur ce point et avons fourni une indication supplémentaire au participant, ce qui peut constituer un biais dans la réalisation de l'épreuve. Cependant, sans cette indication, la réalisation de l'épreuve par le participant n'aurait pas été possible.

Ainsi, si nous avons pu utiliser le protocole de recueil de discours avec les participants aphasiques, et obtenir un échantillon des différents types de discours pour chaque participant, nous avons dû nous en éloigner à plusieurs reprises. Le script du protocole semble donc constituer une base sur laquelle s'appuyer mais il ne sera jamais possible, avec les participants aphasiques, de le respecter à la lettre. Chaque examinateur donne des indications différentes qui peuvent orienter d'une certaine manière la réponse du participant. On peut ainsi s'interroger sur la comparabilité des résultats présents dans la base de données AphasiaBank. En effet, si nous avons essayé de fournir un maximum d'incitations non verbales, les incitations verbales ou les commentaires fournis par chaque examinateur sont impossibles à éviter, spécialement lorsqu'on évalue le discours.

Les épreuves que nous proposons incitent à un échange, le discours se construit aussi en interaction et les participants ont souvent sollicité notre intervention en nous posant des questions. Il est ainsi plus difficile de contrôler les paramètres extérieurs que dans d'autres épreuves de langage, comme une épreuve de dénomination par exemple.

❖ **Matériel utilisé**

Les participants aphasiques ont parfois été perturbés par les images présentes dans le livre de Cendrillon car celles-ci sont très fournies en détails qui n'ont pas de rapport direct avec l'histoire. Quelques participants sont restés focalisés sur ces détails au détriment des images globales qui illustraient l'histoire. Le choix du livre pourrait donc être repensé en prenant en compte ce paramètre.

1.2.2.3. Conclusion

Bien que quelques améliorations pourraient être apportées, le protocole de recueil de données discursives AphasiaBank est utilisable avec des participants contrôles et des participants aphasiques francophones. **La sous-hypothèse 2 est donc validée.**

1.2.3. Sous-hypothèse 3 : Le protocole de recueil de données des épreuves complémentaires AphasiaBank est utilisable avec des sujets contrôles et des sujets aphasiques francophones

1.2.3.1 Synthèse des résultats

La durée moyenne de passation des épreuves complémentaires est de 32 minutes et 53 secondes. Elles ont pu être réalisées en entier pour chacun des participants aphasiques.

1.2.3.2. Discussion, intérêts et limites

• **Durée de passation des épreuves complémentaires**

La durée de passation des épreuves complémentaires, par rapport à celle des épreuves de discours, est beaucoup plus longue. Les passations de deux participants ont dû se dérouler en deux fois, du fait de la fatigue générée par la passation des différents tests. On pourrait alors se demander si la durée de passation des épreuves complémentaires, à la suite des épreuves de discours, n'est pas trop longue. De plus, la fatigue consécutive à la longueur de ces épreuves pourrait avoir un impact sur les capacités des participants. Dans une perspective d'amélioration du protocole, la réduction du temps de passation des épreuves complémentaires pourra être un élément à prendre en compte.

• **Intérêt des données recueillies**

Ces épreuves complémentaires présentent l'intérêt de recueillir des résultats chiffrés précis pour chaque patient, ce qui permet ensuite de pouvoir comparer les patients entre eux. En outre, ces données précises pourront être utiles pour les chercheurs qui pourront les exploiter dans le cadre de leurs recherches.

• **Durée totale des épreuves langagières**

Concernant l'ensemble du protocole de recueil de données langagières, la durée totale de passation est en moyenne de 49 minutes et 57 secondes, la durée la plus longue étant d'1 heure et 13 minutes. Il s'agit d'une durée relativement longue. Si l'ensemble du protocole est utilisable présenté ainsi, à des fins de recherche, cette durée devra peut-être être réduite dans une

perspective d'utilisation future en clinique orthophonique.

1.2.3.3. Conclusion

L'ensemble des épreuves complémentaires a pu être utilisé avec les participants aphasiques, la sous-hypothèse 3 est donc validée.

L'ensemble du protocole AphasiaBank a pu être utilisé avec les participants contrôles et aphasiques francophones, **l'hypothèse 2 est donc validée.**

1.3. Hypothèse 3 : EVAL est opérationnel avec des corpus issus de sujets contrôles et de sujets aphasiques francophones

Nous avons postulé qu'EVAL était opérationnel avec des corpus issus de sujets contrôles et de sujets aphasiques francophones.

L'utilisation d'EVAL nécessite plusieurs étapes préalables destinées à préparer les corpus de données langagières francophones. Ces étapes comprennent la transcription et le codage des données selon les normes CHAT, la vérification des transcriptions par le programme CHECK, le lancement du programme d'étiquetage morphologique, la mise en ligne des données sur le site AphasiaBank et le lancement effectif d'EVAL. Chacune de ces étapes fait l'objet d'une sous-hypothèse que nous allons à présent discuter.

1.3.1. Sous-hypothèse 1 : Les corpus francophones peuvent être transcrits et codés selon les normes CHAT

1.3.1.1. Synthèse des résultats

• **Durée du travail consacré à la transcription et au codage**

La durée de transcription et de codage des données est en moyenne plus élevée chez les sujets aphasiques que chez les sujets contrôles (3 heures 44 minutes 17 secondes chez les sujets contrôles, 5 heures 27 minutes 16 secondes chez les sujets aphasiques). Pourtant, la durée moyenne des enregistrements retranscrits est à peu près équivalente pour les deux populations (16 minutes 40 secondes chez les sujets contrôles, 17 minutes 04 secondes chez les sujets aphasiques).

Chez les sujets contrôles, la transcription qui a demandé un temps de transcription et de codage le moins important (1 heure 45 minutes) correspondait à un enregistrement de 15 minutes 56 secondes. Celle qui a demandé le plus de temps (6 heures 35 minutes) correspondait à un enregistrement de 18 minutes et 15 secondes.

Chez les sujets aphasiques, la transcription qui a demandé un temps de transcription et de

codage le moins important (2 heures) correspondait à un enregistrement de 14 minutes 20 secondes. Celle qui a demandé le plus de temps (9 heures 35 minutes) correspondait à un enregistrement de 15 minutes et 54 secondes.

• **Utilisation des codages et des normes de transcription**

Nous avons pré-sélectionné 16 codages liés aux normes générales de transcription, 6 codages d'erreurs au niveau du mot et 5 codages d'erreurs au niveau de l'énoncé.

Nous avons utilisé l'ensemble des normes de codage et des codages d'erreurs au niveau du mot que nous avons pré-sélectionnés. Concernant les codages d'erreurs au niveau de l'énoncé, nous n'avons utilisé que quatre codages sur les cinq que nous avons pré-sélectionnés. Nous n'avons jamais utilisé le codage « disfluence » ([* d]) car il ne nous a pas semblé approprié aux corpus que nous avons eu à transcrire.

Au total, nous avons codé 6 erreurs au niveau du mot chez les sujets contrôles et 4 au niveau de l'énoncé. Parmi les erreurs au niveau du mot, nous avons comptabilisé 5 erreurs morphologiques et 1 erreur sémantique. Toutes les erreurs au niveau de l'énoncé étaient des erreurs grammaticales.

Nous avons codé 177 erreurs au niveau du mot et 133 au niveau de l'énoncé chez les sujets aphasiques.

1.3.1.2. Discussion, intérêts et limites

• **Durée du travail consacré à la transcription et au codage**

D'après ces résultats, il semble que la durée de transcription et de codage ne soit pas uniquement tributaire de la durée de l'enregistrement à transcrire. En effet, à durée d'enregistrement à peu près équivalente, le temps consacré à la transcription et au codage peut être très variable. Parmi les paramètres pouvant expliquer l'augmentation du temps de transcription et de codage, nous pouvons citer le degré de connaissance des normes de transcription et de codage du transcripateur, ainsi que la complexité du discours à retranscrire et l'importance du nombre d'erreurs à coder. Ces deux derniers paramètres pourraient ainsi expliquer le fait que le temps de transcription et de codage du discours d'un sujet aphasique soit nettement plus élevé que celui d'un sujet contrôle. En effet, chez les sujets aphasiques, le discours est souvent plus difficile à transcrire du fait de certaines déviations affectant l'intelligibilité du discours. De plus, le plus grand nombre d'erreurs à coder chez les sujets aphasiques nécessite davantage de travail de la part du transcripateur.

La durée moyenne de transcription et de codage pour un corpus aphasique est de 5 heures 27 minutes 16 secondes. Ce temps rend bien évidemment l'utilisation d'ÉVAL difficilement réalisable pour les cliniciens. Ce travail s'est inscrit dans une démarche d'adaptation du

programme à la langue française. Pour cela, la transcription d'un nombre suffisant de corpus afin d'initier la création d'une base de données en langue française était nécessaire, ce qui explique le temps important consacré à la transcription et au codage des données.

- **Utilisation des codages et des normes de transcription**

- **Normes de transcription**

Concernant l'exclusion des énoncés, nous avons choisi de ne rien exclure dans les tâches de discours libre. En effet, ces tâches de discours se présentent sous forme de discussion libre, bien que celle-ci soit guidée par les incitations de l'examineur. Or, dans ce type d'interaction, les commentaires, les apartés et les énoncés modalisateurs, sont très fréquents et ils reflètent le discours produit par le participant dans ce cadre de production du discours. Si tous les commentaires, apartés et énoncés modalisateurs, étaient exclus de l'analyse EVAL lors de ces tâches de discours, l'analyse EVAL ne refléterait pas l'intégralité du discours du participant. De plus, dans ce cadre d'interaction, il est particulièrement difficile de juger ce qui relève ou non du protocole et de la tâche à accomplir, et donc de décider des énoncés à exclure ou à inclure dans l'analyse.

Conformément aux recommandations du manuel et du document « Error coding », nous avons choisi d'être plus strictes concernant l'exclusion des énoncés dans les tâches de discours dirigé. En effet, dans ce type de tâche, le discours du participant est plus contraint que dans les tâches de discours libre, il est donc plus facile de discerner ce qui relève ou non du protocole et donc d'exclure ce qui ne relève pas de la tâche précise à accomplir.

Les choix que nous avons effectués concernant les normes de transcription (segmentation en énoncés, exclusion des énoncés,...) ont permis d'améliorer l'uniformité de nos transcriptions mais n'ont pas permis de résoudre tous les cas de figure. Une part de subjectivité persiste et, dans certains cas, la décision finale doit relever du choix du transcripteur.

- **Réflexion sur les codages d'erreurs**

Dans l'ensemble, nous avons codé peu d'erreurs chez les sujets contrôles, en moyenne moins d'une erreur par transcription. Il est cependant intéressant de constater que des erreurs sont effectivement présentes dans les transcriptions des sujets contrôles. Ce constat est en faveur de l'intérêt de coder les erreurs également chez les sujets contrôles afin de pouvoir effectuer une comparaison plus fine avec les erreurs constatées dans les productions de sujets aphasiques. En effet, chez ces derniers, il se peut que certaines erreurs relèvent également de processus normaux relatifs à l'élaboration du langage continu. Le fait de coder toutes les erreurs de la même façon chez tous les sujets permet ensuite de comparer la quantité des erreurs produites par chacun des groupes et de déterminer quel niveau d'erreurs permet de distinguer les erreurs relevant de la

normalité et de la pathologie.

L'analyse du type d'erreurs produites révèle parfois une importante prédominance d'un type d'erreur, notamment des erreurs phonologiques ou des erreurs sémantiques, chez certains sujets, alors que chez d'autres sujets les types d'erreurs sont plus variés. Ces résultats sont en faveur de l'intérêt de préciser davantage les codages d'erreurs, afin de disposer, au niveau clinique, d'une analyse plus fine du type d'erreurs produites.

- **Réflexions sur la quantité de codages utilisés**

L'ensemble des codages que nous avons effectivement utilisés nous ont paru nécessaires, dans le cadre de notre objectif, pour retranscrire nos données avec suffisamment de précision. Ce choix présente cependant certaines limites pouvant être envisagées depuis deux points de vue contradictoires.

D'un certain point de vue, le nombre de codages que nous avons choisi peut sembler trop restreint par rapport à la grande quantité de codages disponibles dans le système CHAT. Le choix de n'utiliser qu'un nombre limité de codages entraîne nécessairement un codage moins précis des données langagières et peut empêcher d'effectuer par la suite certains calculs proposés par le logiciel CLAN nécessitant des codages spécifiques.

D'un autre point de vue, le nombre de codages que nous avons choisi d'utiliser peut paraître trop étendu par rapport à l'objectif de départ que nous nous étions fixées, à savoir préparer nos données à l'utilisation du programme EVAL. Nous avons en effet choisi de coder avec davantage de précision que ce qui était préconisé dans le manuel EVAL. De plus, l'utilisation d'un nombre plus important de codages que les codages nécessaires à l'analyse EVAL peut être considérée comme un travail non pertinent et trop chronophage par rapport à l'objectif poursuivi. Nous avons fait ce choix car notre objectif n'était pas uniquement de préparer l'utilisation d'EVAL, mais d'envisager son utilisation ultérieure en clinique orthophonique.

Nous avons effectué ce choix *a priori*, avant d'entreprendre le travail de transcription. À l'issue de ce travail, nous sommes conscientes que quel que soit le choix effectué en matière de transcription et de codage, ces deux limites pourront toujours nous être opposées. Ce choix est propre au transcripateur et reste adaptable en fonction des objectifs d'analyse poursuivis par chaque transcripateur.

- **Manque d'uniformisation et de précision des codages d'erreurs**

Certaines limites peuvent également être relevées au sujet du codage des erreurs. Une de ces limites concerne le manque d'uniformisation des outils AphasiaBank. Les codages d'erreur sont mentionnés dans plusieurs documents disponibles sur le site AphasiaBank et ils ne sont pas systématiquement concordants. Ainsi, le codage des erreurs au niveau de l'énoncé n'est pas le même dans le document « Error coding » [+ gram] et dans le manuel EVAL [+ *]. De la même

façon, d'après le manuel EVAL, les néologismes sont transcrits en alphabet standard suivi de @n et d'après le document « Error coding » ils doivent être transcrits en alphabet phonétique suivi de @u.

Une autre de ces limites concerne les codages d'erreurs proposés par le système CHAT. Certains codages manquent de précision quant aux erreurs auxquelles ils doivent s'appliquer. En effet, pendant la phase de transcription et de codage des données, nous avons parfois eu des difficultés à choisir le codage d'erreur approprié, notamment en ce que concerne les codages d'erreur morphologique [* m] et d'erreur grammaticale [+ * gram]. Dans de nombreux cas, nous avons beaucoup hésité dans le choix du codage approprié, comme dans l'exemple suivant :

*PAR: (...) c'était je pense la personne (...) <u>auquel</u> j'étais le_plus proche (...). [+ * gram]
--

- **Intérêts de la transcription, du codage et de l'étiquetage morphologique**

Malgré ces limites, la transcription, le codage et l'étiquetage morphologique des données présentent un certain nombre d'intérêts. Les données langagières transcrites et codées au format CHAT permettent de stocker des échantillons de discours oral à long terme et d'effectuer de multiples analyses langagières.

Le stockage des données transcrites, ainsi que des fichiers audio correspondants, permet aux cliniciens et aux chercheurs de conserver leurs données et de les partager. Les données langagières sont ainsi constamment à disposition dans un format qui les rend prêtes à être analysées.

La transcription, le codage et l'étiquetage morphologique permettent en outre d'effectuer de nombreuses analyses langagières précises qu'il serait difficile d'effectuer manuellement. En effet, le logiciel CLAN offre, en plus du programme d'analyse EVAL, de nombreuses options permettant de calculer diverses mesures d'analyse linguistique qui sont détaillées dans le manuel CLAN (MacWhinney, 2000). Ainsi, à tout moment, les cliniciens et les chercheurs peuvent revenir sur leurs corpus stockés pour y effectuer des analyses supplémentaires et compléter leurs analyses précédentes.

Au point de vue clinique, ce système de transcription, de codage et d'étiquetage morphologique peut permettre aux orthophonistes de stocker des données langagières orales de leurs patients et de les conserver pour les analyser, au besoin, plus finement, les joindre au bilan, ou encore les comparer à des données langagières produites par un même patient à un temps différent et ainsi de suivre l'évolution de son discours oral de façon précise au cours du temps.

1.3.1.3 .Conclusion

Les données langagières recueillies auprès de sujets contrôles et aphasiques francophones lors de l'administration du protocole AphasiaBank ont pu être transcrites et codées au format CHAT sur le logiciel CLAN. **La sous-hypothèse 1 est donc validée.**

1.3.2.Sous-hypothèse 2 : Les corpus francophones peuvent être vérifiés par le programme CHECK et les erreurs détectées peuvent être corrigées

1.3.2.1. Synthèse des résultats

La vérification de chacune des transcriptions dure en moyenne entre 10 et 20 minutes. Nous avons vérifié et corrigé chacune de nos transcriptions jusqu'à l'obtention du message de réussite.

1.3.2.2. Discussion, intérêts et limites

L'utilisation du programme CHECK est indispensable car, malgré la vigilance du transcripateur, le programme détecte toujours quelques erreurs qui peuvent ainsi être corrigées.

Cependant, nous avons constaté que le programme CHECK ne détectait pas l'ensemble des erreurs de la transcription et qu'il ne relevait pas certaines erreurs pouvant affecter la fiabilité des calculs de CLAN et d'EVAL. Ainsi, la présence de majuscule en début d'énoncé n'est pas signalée, ce qui peut conduire à des erreurs d'étiquetage morphologique, le programme MOR considérant tout mot avec majuscule comme un nom propre. De même, le programme CHECK ne détecte pas certaines erreurs dans la façon de transcrire les codages d'erreurs ce qui peut conduire le logiciel CLAN à comptabiliser le même type d'erreurs dans deux catégories distinctes et donc biaiser les calculs d'erreurs. Ainsi, en cas de mauvaise transcription du codage d'erreur, le logiciel CLAN comptabilisera par exemple de façon distincte les erreurs codées par le même code avec ou sans espace, comme dans l'exemple suivant où les persévérations sont comptées de façon distincte du fait de la présence d'un espace avant l'accolade :

```
From file <colin11a.cha> Speaker:
```

```
*PAR:
```

```
44 [+ *gram]
```

```
2 [+ *per ]
```

```
3 [+ *per]
```

1.3.2.3. Conclusion

Malgré ces limites, l'ensemble de nos transcriptions ont été vérifiées par le programme CHECK et l'ensemble des erreurs détectées ont pu être corrigées. **La sous-hypothèse 2 est donc validée.**

1.3.3. Sous-hypothèse 3 : Le programme d'étiquetage morphologique peut être lancé sur les corpus francophones

1.3.3.1. Synthèse des résultats

Le lancement du programme d'étiquetage morphologique prend en moyenne 1 heure. Ce temps, qui s'explique par la longueur des transcriptions que nous avons à retranscrire, a cependant diminué au fur et à mesure des transcriptions du fait de notre maîtrise des corrections à apporter aux mots mal transcrits et de la baisse du nombre de mots à ajouter au fichier lexique. En effet, les mots ajoutés dans les transcriptions précédentes étaient reconnus par le programme dans les transcriptions suivantes.

La plupart des mots non reconnus par MOR l'ont été parce qu'il ont été mal retranscrits et non parce qu'ils ne figuraient pas dans les fichiers lexiques de la grammaire française.

L'ajout des mots non reconnus dans les fichiers lexiques s'est révélé indispensable afin de permettre au programme d'étiquetage morphologique de fonctionner correctement et donc de permettre à CLAN et EVAL d'effectuer des calculs corrects.

De plus, la création d'un nouveau fichier lexique qui est ensuite transmis aux concepteurs d'AphasiaBank permet de contribuer à l'enrichissement de la grammaire française générale. Ainsi, au fur et à mesure des contributions de chaque transcripateur, les transcripateurs suivants auront moins de mots à ajouter aux fichiers lexiques et ce travail prendra ainsi moins de temps.

L'analyse de la composition du fichier lexique que nous avons créé révèle une prédominance d'adverbes. Ceci peut s'expliquer par le fait que nous avons fait le choix de solidariser un nombre important de locutions adverbiales, choix qui n'avait pas été celui des précédents transcripateurs ayant constitué et contribué à l'enrichissement des fichiers lexiques de la grammaire MOR française générale.

1.3.3.2. Discussion, intérêts et limites

Nous avons fait le choix de solidariser certains morphèmes afin que le programme d'étiquetage morphologique les considère comme des entités solidaires. Un grand nombre des combinaisons que nous avons choisi de solidariser ne figurait pas dans les fichiers lexiques de la grammaire française, ce qui nous a conduit à les ajouter dans le fichier lexique que nous avons créé. La majorité des mots que nous avons ajoutés dans le fichier sont des morphèmes que nous avons choisi de solidariser alors qu'ils s'écrivent en unités distinctes dans l'orthographe française conventionnelle. Au vu de l'analyse de ces résultats, nous pensons que le choix de solidariser de si nombreuses combinaisons de morphèmes n'était finalement peut-être pas le plus approprié car il a nécessité un travail considérable lié à l'ajout de nouvelles entrées lexiques. De plus, les concepteurs de la base de données nous ont indiqué qu'il était finalement préférable de restreindre le nombre de combinaisons de morphèmes solidarisés à un nombre minimum.

De plus, nous avons choisi de retranscrire les formes lexicales élidées, comme « *qu'*, *lorsqu'*, *puisqu'*, *parce_qu'* » de la même manière qu'en orthographe française conventionnelle. Ce choix a entraîné de nombreux mots non reconnus par le programme MOR et la nécessité d'ajouter ces formes au fichier lexique. De plus, les concepteurs de la base de données nous ont indiqué que ces formes peuvent poser des problèmes lors de la génération de la ligne d'étiquetage morphologique. Afin de contourner ces difficultés, il est préférable de transcrire ces formes élidées de la même façon que les abréviations, c'est-à-dire de la façon suivante : « *qu(e)*, *lorsqu(e)*, *puisqu(e)*, *parce_qu(e)* ».

Une autre limite concerne la désambiguïsation. Lors de l'étape d'étiquetage morphologique, le programme MOR liste toutes les catégories grammaticales possibles de chaque mot. Le programme POST est ensuite lancé automatiquement afin de procéder à une « désambiguïsation », en déterminant la catégorie appropriée en fonction du contexte d'utilisation du mot. Ce programme de désambiguïsation ne semble pas totalement opérationnel pour la langue française. En effet, lorsque nous avons fait des essais en générant une ligne %mor à partir de l'utilisation d'un même mot dans plusieurs contextes phrastiques différents, nous nous sommes rendus compte que la fonction de désambiguïsation du programme MOR n'était pas tout à fait opérationnelle. Si la plupart des morphèmes sont correctement étiquetés par le programme d'étiquetage morphologique, certains morphèmes sont néanmoins étiquetés de façon incorrecte.

1.3.3.3. Conclusion

L'ensemble des mots présents dans nos transcriptions a pu être reconnu par le programme MOR et la ligne d'étiquetage morphologique a pu être générée sur l'ensemble de nos corpus. **La sous-hypothèse 3 est donc validée.**

1.3.4. Sous-hypothèse 4 : Les corpus francophones peuvent être mis en ligne sur la base de données française d'AphasiaBank

1.3.4.1. Synthèse des résultats

Nos transcriptions et les fichiers audio correspondants ont été mis en ligne le 7 mai 2016. La base de données française est donc désormais constituée de 25 transcriptions de données issues de la passation du protocole AphasiaBank, 14 corpus de sujets contrôles et 11 de sujets aphasiques.

1.3.4.2. Discussion, intérêts et limites

La création d'une base de données en langue française rend désormais possible des comparaisons entre différentes transcriptions de mêmes tâches de discours. Les corpus et les

fichiers audio étant désormais accessibles en ligne sur le site AphasiaBank, les cliniciens et les chercheurs membres d'AphasiaBank peuvent effectuer de nombreuses analyses linguistiques et effectuer des analyses comparées entre les différents corpus grâce aux différentes fonctionnalités proposées par le logiciel CLAN. De plus, nos transcriptions sont prêtes à être analysées par le programme EVAL lorsque celui-ci aura été adapté à la langue française. La constitution de la base de données française rendra alors possible le calcul des écarts-types des différentes mesures calculées par EVAL.

Certaines de nos passations ont pu être recueillies au format vidéo. Nous avons cependant fait le choix de ne déposer sur la base de données que les fichiers audio, dans le souci de préserver l'anonymat des participants ayant participé à notre étude. Nous sommes cependant conscientes que les enregistrements audio présentent certaines limites par rapport aux enregistrements vidéo. En effet, une transcription et un fichier audio ne permettent pas de rendre compte de l'interaction de façon aussi détaillée et précise que ne l'aurait permis un enregistrement vidéo. En choisissant de ne conserver que les fichiers audio, nous n'avons conservé aucune trace, si ce n'est quelques annotations dans nos transcriptions, concernant les aspects non verbaux de l'interaction qui sont fondamentaux dans le cadre des interactions conversationnelles. Nous avons ainsi exclu une source d'information très importante pour l'interprétation complète de l'interaction. Cependant, dans le cadre de notre objectif de création d'une base de données visant l'utilisation d'EVAL, l'important était d'avoir des corpus transcrits d'après lesquels il était possible de générer des analyses quantitatives et qualitatives du point de vue formel.

Dans la perspective d'initier la création d'une base de données, le nombre de corpus que nous avons intégré à la base de données nous a semblé suffisant. Cependant, dans la perspective de l'utilisation d'EVAL et de la comparaison entre les différents corpus, la question de la représentativité de la base de données se pose. En effet, afin d'améliorer cette représentativité, il est nécessaire que la base de données soit enrichie par la suite d'autres corpus francophones, à la fois issus de sujets contrôles et de sujets aphasiques, par d'autres membres contributeurs. La mise à disposition d'une version française du protocole AphasiaBank sur le site va rendre possible ces futures contributions.

1.3.4.3. Conclusion

L'ensemble de nos transcriptions est désormais en ligne sur la base de données et accessible aux cliniciens et aux chercheurs membres d'AphasiaBank. **La sous-hypothèse 4 est donc validée.**

1.3.5. Sous-hypothèse 5 : Le programme EVAL peut être lancé de façon opérationnelle sur les corpus francophones

1.3.5.1. Synthèse des résultats

Le programme EVAL a pu être lancé sur les transcriptions francophones mais les résultats obtenus ne sont pas exploitables.

1.3.5.2. Discussion, intérêts et limites

Le programme EVAL n'est pas encore opérationnel dans l'analyse des données langagières francophones. En effet, certaines mesures ne sont pas encore adaptées à l'analyse de données francophones, notamment en ce qui concerne les comparaisons avec les corpus de la base de données. Le programme doit encore être reconfiguré par les informaticiens du projet AphasiaBank pour pouvoir fonctionner de façon optimale sur des corpus en langue française.

1.3.5.3. Conclusion

Le programme EVAL n'est donc pas actuellement opérationnel sur des corpus francophones.

La sous-hypothèse 5 est donc invalidée.

L'ensemble des étapes nécessaires à la préparation des corpus à l'analyse EVAL a pu être effectué, mais le programme EVAL n'est pas encore configuré pour fonctionner de façon optimale sur des corpus en langue française. De ce fait, **l'hypothèse 3 est invalidée.**

2. LES LIMITES

Notre travail présente certaines limites sur lesquelles nous allons à présent nous interroger. Ces limites concernent la population sélectionnée dans le cadre de notre étude et certaines limites inhérentes au projet AphasiaBank lui-même.

2.1. Concernant la population

La sélection de la population contrôle, comme de la population aphasique, comporte de nombreuses limites. Notre objectif était simplement d'initier la création d'une base de données. Ainsi, à ce jour, les populations contrôles et aphasiques ne sont pas assez représentatives pour pouvoir obtenir des résultats suffisamment fiables.

2.2. Concernant le projet AphasiaBank

Nous allons à présent discuter des limites inhérentes au projet AphasiaBank lui-même. Celles-ci concernent les critères d'inclusion du protocole, les tâches de transcription et de codage des données, ainsi que l'étiquetage morphologique.

2.2.1. Limites des critères d'inclusion

Nous avons relevé certaines limites concernant les critères d'inclusion du projet AphasiaBank. Dans la liste des critères d'inclusion des sujets contrôles et aphasiques, proposée sur le site AphasiaBank, il n'est pas précisé que les participants ne doivent pas avoir d'antécédent de trouble(s) de la parole, du langage ou de la communication (autre que l'aphasie pour les participants aphasiques). Nous pensons que ce critère devrait être explicité. En effet, il nous paraît important d'obtenir des données langagières reflétant le discours « normal » chez les sujets contrôles, sans processus pathologique sous-jacent, et des données reflétant les processus pathologiques dus uniquement à l'aphasie chez les participants aphasiques et pas à d'autres troubles langagiers.

De plus, concernant les critères d'inclusion de la population aphasique, il n'est pas précisé que les sujets doivent avoir une vue et une audition suffisante pour effectuer la passation, ne pas avoir de troubles dépressifs au moment de la passation, et parler couramment la langue, alors que ces critères sont explicités dans la liste des critères d'inclusion des participants contrôles. Il nous semble important de les expliciter également pour la population aphasique afin d'obtenir des données langagières ne reflétant pas d'autres perturbations que celles liées à l'aphasie. De plus, cela permettrait de minimiser les différences entre les populations contrôles et aphasiques, et pourrait améliorer la fiabilité des comparaisons.

2.2.2. Limites de la transcription et du codage des données

Nous nous sommes interrogées sur les limites inhérentes à la tâche de transcription et de codage. Celles-ci sont principalement liées aux particularités mêmes du discours oral, à la part de subjectivité et aux écueils impliqués par la tâche de transcription, ainsi qu'aux difficultés relatives à certains codages.

• **Particularités du discours oral**

Les difficultés inhérentes à la tâche de transcription s'expliquent en partie par les particularités du discours oral que nous avons précédemment décrites dans la partie théorique. En effet, le discours oral n'obéit pas aux règles conventionnelles de la langue écrite. Certains éléments morpho-syntaxiques ne sont pas explicités du fait de la présence d'éléments paraverbaux ou

contextuels. De plus, le mode d'élaboration continu du discours oral rend ce dernier particulièrement complexe du fait de la fréquence de répétitions, reformulations et d'incises venant rompre l'écoulement du message. L'ensemble de ces particularités rend le discours oral particulièrement difficile à retranscrire sous une forme écrite.

- **Part de subjectivité inhérente à la transcription**

Tout au long de la tâche de transcription, le transcripteur est amené à prendre de nombreuses décisions, concernant par exemple la segmentation du discours en énoncés ou le codage des erreurs. La tâche de transcription est ainsi nécessairement tributaire de la subjectivité du transcripteur. Cette subjectivité est importante car elle fait appel au sens d'analyse et à l'expérience clinique du transcripteur. En effet, considérer une production comme une erreur ou non est déjà une analyse, une interprétation qualitative faite par le clinicien, et cela va influencer les calculs d'ÉVAL. Le jugement du clinicien est donc sollicité dès la phase de transcription et le jugement clinique est fondamental durant cette étape. Cependant, cette part de subjectivité, qui implique une part d'interprétation, peut induire des différences importantes dans la façon de transcrire et de coder les données. Nous nous sommes ainsi rendues compte que selon le transcripteur un même enregistrement audio pouvait être transcrit et codé de façon différente et produire ainsi des résultats différents dans les calculs effectués par le logiciel CLAN. De plus, nous nous sommes aperçues qu'en réécoutant l'enregistrement audio, nous avons tendance à revenir sur nos choix de transcription et de codage antérieurs. Ainsi, un même transcripteur pourrait retranscrire de façon différente un même passage selon le moment de l'écoute. Ceci remet en question la reproductibilité des transcriptions.

- **Écueils de la transcription : sous-normalisation et sur-normalisation**

Dans l'introduction du manuel CHAT, MacWhinney évoque les deux principaux dangers de la transcription : la sous-normalisation et la sur-normalisation (MacWhinney, 2000a). Ses propos concernent la transcription du langage enfantin mais des difficultés similaires se posent dans la transcription du langage déviant de type aphasique. La sous-normalisation renvoie au fait de ne pas faire correspondre une forme erronée à une forme standard alors qu'il y avait en fait une correspondance. Dans le cas du langage aphasique, cela pourrait concerner la non reconnaissance du mot cible d'une paraphasie et le codage d'un néologisme au lieu d'une paraphasie avec mention du mot cible. La sur-normalisation consiste à l'inverse à appliquer une correspondance entre une forme déviante et une forme standard alors qu'en fait il n'y a pas de véritable correspondance (MacWhinney, 2000a). Dans le cas du langage aphasique, cela reviendrait par exemple à interpréter une production déviante comme une tentative de produire un mot cible et de l'indiquer dans la transcription alors que ce mot cible ne correspond en fait pas à ce que voulait produire le participant. Cette difficulté est d'autant plus importante dans le cadre de la

transcription du discours oral continu, car, contrairement à d'autres tâches langagières, comme les tâches de dénomination, l'investigateur ne connaît pas la cible et doit inférer ce que le participant veut produire.

Ces difficultés de transcription peuvent aussi se produire chaque fois qu'un mot oral peut correspondre à plusieurs formes écrites (par exemple le mot /mãʒe/ pourrait être retranscrit par les formes lexicales « *manger* », « *mangé* », « *mangeait* » ou encore « *mangez* »). Le transcripteur est ainsi amené à effectuer de nombreux choix pouvant influencer l'analyse des données langagières retranscrites. Ces difficultés peuvent cependant être contournées par l'utilisation de l'alphabet phonétique (MacWhinney, 2000a).

• **Manques concernant les codages**

Nous avons également relevé certains manques concernant les codages d'erreurs existants. Ainsi il n'existe pas de codage spécifique permettant de coder les conduites d'approches. Les fragments de mots, qui incluent les conduites d'approches, sont exclus de l'analyse par l'utilisation du symbole & afin de permettre au programme MOR de reconnaître l'ensemble des mots. De ce fait, les conduites d'approches ne sont pas comptabilisées comme des erreurs et n'apparaissent pas dans les calculs informatisés. Or l'utilisation des conduites d'approches est une perturbation langagière fréquemment observée chez les sujets aphasiques et ce comportement langagier est important dans la caractérisation du profil discursif du patient aphasique. À des fins cliniques, le calcul des conduites d'approches présenterait un intérêt significatif au service de l'analyse orthophonique.

De plus, les codages d'erreurs actuels ne permettent pas de distinguer les erreurs phonétiques (affectant le niveau de la troisième articulation du langage) et les erreurs phonologiques (affectant le niveau de la deuxième articulation du langage). Or cette distinction est importante pour l'évaluation orthophonique car elle permet de définir le niveau d'atteinte langagière du patient. Face à cette limite, nous avons choisi de ne coder dans nos transcriptions que les erreurs relevant selon nous d'une atteinte phonologique. Nous n'avons donc pas codé les erreurs stables et récurrentes qui nous semblaient relever d'une atteinte phonétique. Ainsi, la présence de déviations phonétiques ne figure pas dans nos résultats concernant le nombre et le type d'erreurs relevées dans chacune des transcriptions. L'analyse des erreurs a donc été faite de façon incomplète.

En outre, le codage des erreurs grammaticales nous semble manquer de précision. En effet, ce codage est utilisé indifféremment dans le cas d'un agrammatisme et dans le cas d'une dyssyntaxie. Il ne permet pas de distinguer ces deux types de perturbations qui relèvent pourtant de processus pathologiques différents.

Nous avons également relevé un manque concernant le calcul des erreurs ayant fait l'objet

d'une reformulation, comme dans l'exemple suivant :

***PAR:** et [/] et là il est <sous le vol [: averse] [* s]> [/] sous l'averse.

Nous avons précédemment calculé le nombre et le type d'erreurs qui ont été codées dans nos transcriptions à l'aide d'une commande CLAN.²⁷ En comparant le résultat ainsi obtenu et un résultat obtenu manuellement, nous nous sommes rendues compte que les erreurs incluses dans des reformulations n'apparaissaient pas dans le calcul d'erreur effectué par CLAN.

Le programme EVAL étant basé sur l'utilisation des commandes CLAN, nous pensons que ces erreurs n'apparaîtront pas non plus dans le calcul du nombre d'erreurs effectué par le programme EVAL. Or, le nombre d'erreurs incluses dans les reformulations correspond aux erreurs qui ont fait l'objet d'une tentative (réussie ou non) d'auto-correction par le participant. Cette information renseigne donc sur les capacités de jugement du sujet sur son propre discours et ses capacités d'auto-correction, une information qui serait pertinente dans le cadre d'une utilisation d'EVAL en clinique orthophonique.

Enfin, nous avons constaté que le discours de certains participants aphasiques ne contenait que peu d'erreurs formelles que nous aurions pu coder par les codages d'erreurs précédemment décrits et que ce dernier était pourtant très perturbé, au niveau du sens, de la cohérence ou de l'adéquation du propos à la situation de communication. Il nous a alors semblé intéressant, pour compléter l'analyse, d'envisager la possibilité de compléter les codages existants par des codages permettant de coder des éléments liés aux aspects plus pragmatiques du discours, ou de compléter l'analyse EVAL par d'autres outils en clinique orthophonique. Le logiciel CLAN est tout à fait adaptable et il est possible de créer des nouveaux codages afin d'analyser de nouveaux aspects.

3. PERSPECTIVES

3.1. Perspectives en recherche fondamentale

3.1.1. Perspectives à court terme

L'utilisation d'EVAL sur des corpus de langue française n'est actuellement pas encore possible. La prochaine étape, afin de permettre son utilisation, concerne donc la reconfiguration du programme par les informaticiens du projet AphasiaBank. L'adaptation du programme à la langue française va être réalisée très prochainement par les concepteurs d'AphasiaBank. Lors de cette étape d'adaptation, il est possible que certaines mesures calculées par EVAL sur les corpus en langue anglaise ne soient pas adaptées à la langue française. Ainsi, l'adaptation d'EVAL à la

²⁷ Commande pour obtenir le détail des erreurs au niveau du mot : freq +s\[**\] +t*PAR *.cha
Commande pour obtenir le détail des erreurs au niveau de l'énoncé : freq +s"<+ *>" +tPAR *.cha

langue française pourra peut-être nécessiter le retrait et l'ajout de certaines mesures linguistiques.

L'amélioration du désambiguisateur et du programme CHECK devrait également être rapidement entreprise afin de permettre une analyse plus précise des corpus par CLAN et EVAL. La précision du désambiguisateur dépend de la quantité et de la qualité de l'étiquetage manuel effectué dans le dossier /train de la grammaire française. Le désambiguisateur a pour le moment été entraîné sur un petit nombre de corpus en langue française. Afin d'améliorer cette fonction, il faudrait donc procéder à l'étiquetage manuel d'un plus grand nombre de corpus. Cela pourrait constituer l'objet d'une étude ultérieure en linguistique.

3.1.2. Perspectives à moyen et long terme

A moyen et long terme, afin d'améliorer la significativité des comparaisons effectuées par EVAL, il est important que la base de données française continue d'être enrichie de nouveaux corpus issus à la fois de sujets aphasiques et non aphasiques. La mise à disposition du protocole AphasiaBank en version française rend désormais ce travail possible.

Dans le domaine de la recherche, l'accroissement du nombre de corpus déposés en langue française pourrait permettre de mener des recherches approfondies sur la neuropsycholinguistique du discours oral dans le domaine de l'aphasiologie. L'archivage et la mutualisation des données représentent un enjeu majeur pour la recherche clinique. L'utilisation d'EVAL pourrait permettre d'affiner la connaissance des profils discursifs aphasiques en mettant en évidence la singularité des profils de patients atteints du même type d'aphasie et des dissociations entre les différents types de discours.

Ainsi la base de données pourra être utilisée à des fins de recherche mais également dans une perspective de formation à destination des chercheurs et des cliniciens. En effet, la mise à disposition de corpus de participants aphasiques peut également être utilisée dans la formation des chercheurs et des orthophonistes à l'analyse du discours aphasique.

Concernant notre adaptation du protocole AphasiaBank, il pourrait être utile d'évaluer la pertinence de notre traduction et de nos choix d'adaptation. Il pourrait être intéressant de proposer par exemple un questionnaire à faire remplir aux participants contrôles sur le test qu'ils ont passé et d'adapter par la suite le protocole en fonction des réponses obtenues.

Concernant l'adaptation des épreuves équivalentes du protocole AphasiaBank, il pourrait être utile à terme d'adapter l'*AphasiaBank Repetition Test* (Holland et al.,2007) à la langue française afin d'uniformiser le plus possible le protocole dans les différentes langues.

Nous avons également mis en évidence les difficultés rencontrées dans la tâche de transcription. Afin d'uniformiser les transcriptions et d'améliorer leur reproductibilité, il serait intéressant de préciser davantage certaines normes de transcription, concernant les points

précédemment évoqués (segmentation en énoncés, ...).

3.2. Perspectives concernant l'utilisation du protocole AphasiaBank et d'EVAL en clinique orthophonique

3.2.1. Perspectives à court terme

Le nombre actuel de corpus sur la base de données française peut déjà permettre l'utilisation d'EVAL, lorsque le programme aura été adapté à la langue française, par un orthophoniste afin d'obtenir, pour un patient, un profil discursif basé sur des données objectives et comparé à d'autres sujets contrôles ou aphasiques. Cela peut permettre à l'orthophoniste de stocker les données langagières du patient et d'effectuer des analyses lui permettant de cerner avec finesse les déficits linguistiques au niveau formel. Cette analyse permise par EVAL pourra ainsi servir l'analyse qualitative plus globale effectuée par l'orthophoniste.

L'orthophoniste pourra utiliser EVAL de différentes manières :

- lancer l'analyse EVAL sur des épreuves issues du protocole AphasiaBank, pour effectuer une comparaison entre un sujet et la base de données ;
- lancer l'analyse EVAL sur des épreuves issues du protocole AphasiaBank, pour effectuer une comparaison du discours d'un même patient à des temps différents ;
- lancer l'analyse EVAL sur des échantillons de discours hors protocole, pour effectuer une comparaison du discours d'un même patient à des temps différents.

EVAL permet ainsi d'effectuer des comparaisons entre différents sujets, ou chez un même sujet à des temps différents, en conservant le même référentiel d'analyse, ce qui permet d'évaluer avec objectivité le discours d'un patient.

3.2.2. Perspectives à moyen et long terme

EVAL est actuellement utilisé en clinique orthophonique aux Etats-Unis.²⁸ Mais l'utilisation d'EVAL en clinique en France nécessite cependant d'être encore évaluée à différents points de vue, notamment en termes d'utilisabilité²⁹ et d'acceptabilité³⁰ d'une part, et d'utilité d'autre part en contexte clinique français.

28 Communication personnelle avec Margie Forbes (mail du 20 mai 2016), membre fondateur d'AphasiaBank, et Jessica Richardson (mail du 21 mai 2016), clinicienne et chercheuse basée aux Etats-Unis.

29 La notion d'utilisabilité renvoie au « *degré selon lequel un produit peut être utilisé, par des utilisateurs identifiés, pour atteindre des buts définis avec efficacité, efficacité et satisfaction, dans un contexte d'utilisation spécifié* ». (Norme ISO 9241, 1998, cité par Février, 2011, p.15).

30 La notion d'acceptabilité renvoie à la valeur de la représentation mentale que se construisent les individus concernant un objet technologique donné (notamment à propos de son utilité et de son utilisabilité, mais pas seulement) (Tricot et al., 2003, cité par Février, 2011, p.15).

• **Evaluation de l'utilisabilité et de l'acceptabilité d'EVAl en clinique orthophonique**

L'utilisabilité d'EVAl en clinique devra faire l'objet d'une étude spécifique, prenant en compte l'ergonomie et la facilité d'utilisation du programme, mais également la faisabilité en termes de temps. L'utilisation d'EVAl par un clinicien devra être adaptée à des fins cliniques. L'orthophoniste n'aura par exemple pas nécessairement besoin d'utiliser l'ensemble du protocole AphasiaBank et l'ensemble des codages s'il n'envisage pas d'alimenter la base de données.

Ainsi un clinicien engagé dans un protocole de recherche pourra utiliser un nombre plus important de codages et l'ensemble du protocole AphasiaBank. Un clinicien désirant évaluer le discours d'un patient pourra choisir de retranscrire une seule tâche de discours en utilisant un nombre plus restreint de codages et choisir de le comparer ou non à la base de données.

Un protocole plus court, adapté à la clinique orthophonique pourrait ainsi être proposé, afin de permettre une utilisation pertinente d'EVAl dans un cadre clinique, comme c'est le cas aux Etats-Unis où les cliniciens n'utilisent généralement qu'une tâche du protocole (la tâche Window).

Concernant la transcription, l'utilisation d'un logiciel de reconnaissance vocale pourrait être intéressante afin d'effectuer une première version de la transcription qui serait ensuite reprise par l'orthophoniste. Cela pourrait permettre un gain de temps considérable.³¹

Le clinicien pourra choisir les codages qui lui semblent les plus appropriés et ainsi enlever certains codages lui paraissant superflus et en ajouter d'autres qu'il considère importants pour son analyse.

Pour faciliter l'utilisation d'EVAl, il pourrait également être intéressant d'envisager la création d'un tutoriel vidéo en français ou des ateliers de formation à la transcription.

L'évaluation de l'acceptabilité d'EVAl en pratique clinique orthophonique pourrait par exemple être évaluée en demandant à des orthophonistes d'utiliser EVAl et en collectant leurs retours et leurs suggestions d'amélioration.

• **Evaluation de l'utilité d'EVAl en clinique orthophonique**

- **Comparaison entre les résultats obtenus avec EVAl et ceux obtenus avec un bilan orthophonique classique**

Il serait également nécessaire d'évaluer objectivement l'utilité d'EVAl en clinique orthophonique, ce que nous n'avons pas eu le temps de faire dans cette étape exploratoire. Nous envisagions au départ de réaliser une étude de cas afin de comparer les résultats obtenus à un bilan orthophonique classique et les résultats supplémentaires obtenus par l'utilisation d'EVAl. Nous n'avons cependant pas pu mener ce travail à terme du fait des problèmes rencontrés dans l'adaptation du programme EVAl à la langue française. Ce travail reste donc à mener dans le

31 L'utilisation d'un logiciel de reconnaissance vocale en pratique clinique orthophonique a fait l'objet d'un mémoire d'orthophonie en juin 2015 (Balland et Courtade Jouanicq, 2015).

cadre de l'évaluation de la pertinence d'EVAL en clinique orthophonique. L'intérêt d'un tel outil pour mesurer l'efficacité d'une rééducation pourrait également être étudié à travers la comparaison de deux profils discursifs fournis par EVAL à deux temps différents, en début et en milieu ou en fin de thérapie par exemple.

- **Sélection des codages pertinents en clinique orthophonique**

L'étude de la pertinence d'EVAL en clinique orthophonique pourrait ainsi conduire à la sélection d'un nombre restreint de codages et de mesures EVAL pertinents. La sélection d'un nombre limité de codages et de mesures pourrait ainsi rendre l'utilisation d'EVAL plus pertinente en s'adaptant aux contraintes temporelles de passation d'un bilan orthophonique. De même, parmi les trente mesures calculées par EVAL, il est peut-être envisageable de sélectionner les plus pertinentes pour l'évaluation orthophonique. Les mesures permettant de distinguer les aphasies fluentes et non fluentes pourraient par exemple mettre en évidence l'intérêt des mesures EVAL pour aider à une classification basée sur des critères cliniques. La sélection des mesures pertinentes devra néanmoins être affinée par des études ultérieures utilisant des critères de sélection objectifs.

- **Précision du type d'erreurs produites par le patient**

Dans la perspective d'une utilisation clinique d'EVAL, il sera probablement nécessaire d'adjoindre aux mesures EVAL la commande CLAN permettant d'obtenir la liste précise du type d'erreurs produites par le patient car EVAL répertorie les erreurs de façon indifférenciée.

Il pourrait également être intéressant d'effectuer des corrélations entre les scores obtenus à l'échelle de gravité de l'aphasie de l'*HDAE* et le type d'aphasie du patient au nombre et au type d'erreurs codées dans les transcriptions.

- **L'intérêt d'évaluer le discours avec EVAL**

Un questionnaire sur l'intérêt potentiel d'EVAL au sein du bilan orthophonique global pourrait également être mené.

D'une part, l'évaluation du discours n'est qu'une partie du bilan orthophonique global. Ce bilan se fonde également sur d'autres tests ciblés qui fournissent des informations complémentaires sur le profil langagier du patient. Ainsi l'étude du discours ne doit pas être décorrélée de l'ensemble du bilan orthophonique.

D'autre part, l'analyse du discours de façon strictement formelle de façon isolée n'a pas véritablement de sens. Le discours est toujours adressé, il est construit au sein d'une interaction et s'inscrit dans un contexte plus global. EVAL doit donc être utilisé dans la perspective de servir une évaluation globale qui prendra en compte les capacités pragmatiques et conversationnelles de

sujet. L'utilisation d'EVAL en clinique orthophonique devra donc être complétée par l'utilisation d'autres outils.

Ainsi, EVAL peut constituer un outil au service de la clinique mais ne peut évidemment pas se substituer à l'analyse clinique orthophonique qui intègre des aspects bien plus larges, qui permettent de donner sens aux mesures fournies par le programme.

CONCLUSION

L'objectif de notre mémoire était de proposer une adaptation du projet AphasiaBank à la langue française. Dans cette perspective, nous avons entrepris d'adapter le protocole de recueil de données AphasiaBank et de l'administrer à des sujets contrôles et aphasiques francophones. Les données ainsi recueillies ont ensuite été opérationnalisées afin de lancer le programme d'analyse EVAL sur nos corpus.

Nous avons ainsi pu constituer une adaptation française du protocole de données AphasiaBank qui est désormais accessible aux cliniciens et aux chercheurs membres d'AphasiaBank.

La démarche que nous avons découverte et dans laquelle nous nous sommes inscrites, auprès des différents chercheurs impliqués dans le projet AphasiaBank avec lesquels nous avons échangé ces derniers mois, nous a montré l'intérêt de mener des recherches collaboratives à l'échelle internationale entre chercheurs et cliniciens en formation et expérimentés.

Toutefois, nous n'avons pas pu mener notre travail à son terme puisque l'utilisation effective d'EVAL sur nos corpus n'a pas permis d'obtenir des analyses exploitables. En effet, le programme EVAL n'est donc pas encore opérationnel sur des données en langue française.

L'adaptation et l'utilisation d'EVAL à la langue française doivent donc encore se poursuivre afin de permettre son utilisation optimale par les cliniciens et les chercheurs.

À l'issue de ce travail, nous pensons néanmoins pouvoir affirmer que l'utilisation d'EVAL présente des perspectives intéressantes quant à l'évaluation clinique orthophonique du discours oral aphasique. En effet, la base de données constitue, en arrière plan, une référence de discours continu produit par des patients aphasiques et des locuteurs ordinaires, référence à laquelle les performances d'un patient particulier peuvent être relativisées, et archivées si l'on souhaite garder des cotations objectives reflétant son parcours de récupération en discours continu. L'utilisation d'un tel outil devra néanmoins faire l'objet d'études ultérieures afin d'optimiser le protocole de recueil de données et le programme EVAL aux besoins spécifiques de la clinique.

A travers notre démarche d'adaptation du projet AphasiaBank à la langue française, nous espérons avoir contribué à la diffusion des différents outils proposés par ce projet aux cliniciens et chercheurs francophones. Ceux-ci peuvent également être utiles pour la formation et la recherche en orthophonie. Nous espérons qu'ils pourront à terme constituer un outil supplémentaire à disposition du clinicien afin d'améliorer l'évaluation du discours oral chez la personne aphasique, et ainsi contribuer à l'établissement d'une prise en charge la plus adaptée possible.

BIBLIOGRAPHIE

Ouvrages, chapitres d'ouvrages et articles

Blanche-Benveniste C, Bilger M. *Français parlé - oral spontané. Quelques réflexions*. Rev Fr Ling Appl. 1999 ;IV(2):21-30.

Brin-Henry F, Courrier C, Lederlé E, Masy V. *Dictionnaire d'Orthophonie*. Isebergues : Ortho Edition; 2004:303.

Cendrillon. Paris : Hachette Livre. 2014.

Charaudeau P, Maingueneau D. *Dictionnaire d'analyse du discours*. Paris : Editions du Seuil; 2002:661.

Chomel-Guillaume S, Leloup G, Bernard I. *Les aphasies, évaluation et rééducation*. Issy-les-Moulineaux: Elsevier; 2010: 268.

Cook G, Marquis A, Achim A, Dessureault E, Dan NgyuenK, Ooleepeka Neill D, Royle P. *AndroidBAT : Une « boîte à outils » pour la collecte de données vocales, oculaires et tactiles chez les patients aphasiques bilingues ou plurilingues*. Rééducation orthophonique. Mar 2013;253:98-120.

Crevier-Buchman L. *La prise en charge du Handicap en rééducation orthophonique*. Rééducation orthophonique. Déc 2005;224:4-6.

EVAL Manual. (page consultée le 08/04/16). [en ligne].
<http://talkbank.org/APhasiaBank/manuals/eval.pdf>

Fromkin AV, Ratner NB. *Speech Production*. In: Gleason JB, Ratner NB. editors. *Psycholinguistics*. 2nd ed. New York: Harcourt Brace;1998. p.309-346.

Gil R. *Les formes cliniques des aphasies corticales*. Rééducation orthophonique. Juin 1999;198:29-40.

Grimes N. *Walt Disney's Cinderella*. New York: Random House; 1993.

Hellwig B. (page consultée le 22/05/16). *ELAN – Linguistic Annotator, version 4.9.4*, [en ligne].
<http://www.mpi.nl/corpus/manuals/manual-elan.pdf>

Levelt WJM. *Language Use in normal speakers and its disorders*. In: Blanken G, Dittman J, Grimm H, Marshall JC, Wallesch CW, editors. *Linguistic disorders and pathologies : an International handbook*. New York: Gruyter;1993. p.1-15.

Levelt WJM. *Producing spoken language : A blueprint of the speaker*. In: Hagoort P, Brown CM , editors. *The neurocognition of language*. Oxford: Oxford University Press;1999. p. 83-114.

MacWhinney B. *The CHILDES Project: Tools for Analyzing Talk 3rd Edition Part 1: The CHAT Transcription Format*. Mahwah, NJ: Lawrence Erlbaum Associates (ed). 2000a.[en ligne] [mise à jour le 14/04/16 et consulté le 14/04/16]. Disponibilité sur internet :
<<http://childes.talkbank.org/manuals/chat.pdf>>

MacWhinney B. *The CHILDES Project: Tools for Analyzing Talk 3rd Edition Part 2: The CLAN Programs*. Mahwah NJ : Lawrence Erlbaum Associates (ed). 2000b. [en ligne] [mise à jour le 06/04/16 et consulté le 17/04/16]. Disponibilité sur internet : ><http://childes.talkbank.org/manuals/clan.pdf>>

MacWhinney B, Fromm D, Forbes M, Holland A. *AphasiaBank : Methods for studying discourse*. *Aphasiology*. Nov 2011;25(11):1286-1307.

Nespoulous JL. *Référence et modalisation : De la dissociation de leur gestion psycholinguistique et neuropsycholinguistique dans le discours aphasique*. In : Neveu F, Muni Toke V, Durand J, Klingler T, Mondada L, Prévost S, editors. CMLF 2010: Congrès Mondial de Linguistique Française; 2010; Paris.

Riegel M, Pellat JC, Rioul R. *Grammaire méthodique du français*. 5ème ed. Paris: Presses Universitaires de France. 2014:1107.

Sahraoui H, Nespoulous JL. *Across-task variability in agrammatic performance*. *Aphasiology*. 2012;26(6):785-810.

UNADREO. (page consultée le 21/05/16). *Présentation et utilisation de l'OADO*. Avril 2013. [en ligne] <http://www.unadreo.org/assets/medias/fichiers/2014-09-18-14-19-33_9622344.pdf>

Viader F, Lambert J, de la Saynette V, Eustache F, Morin P, Morin I, Lechevalier B. *Aphasie*. *Neurologie*. 2010:17-035-A-96.

Wright HH. *Discourse in aphasia : An introduction to current research and future directions*. *Aphasiology*. Nov 2011;25(11):1283-1285.

Sites internet

Aphasia Software Finder. (page consultée le 25/05/16). [en ligne]. <http://www.aphasiasoftwarefinder.org/>.

AphasiaBank. (dernière date de consultation : 22/05/16). [en ligne]. <http://talkbank.org/APhasiaBank/>

Centre National de Ressources Textuelles et Lexicales. (dernière date de consultation : 16/04/16). [en ligne]. <http://www.cnrtl.fr/definition/>

Fotolia. (page consultée le 31/07/15). [en ligne]. <https://fr.fotolia.com/>

Ortho & Co. (page consultée le 25/05/16) Les nouvelles technologies au service de l'orthophonie, [en ligne]. <http://www.ortho-n-co.fr/>

Praat : doing phonetics by computer. (page consultée le 22/05/16). [en ligne]. <http://www.fon.hum.uva.nl/praat/>

Sémanticlopédie. Dictionnaire de sémantique. (page consultée le 07/06/16) [en ligne]. <http://www.semantique-gdr.net/dico/index.php/Argument>

Agre-ink. Glossaire terminologique de Grammaire Linguistique. (page consultée le 07/06/15 [en ligne]. http://agreg-ink.net/index.php?title=Glossaire_terminologique_de_Grammaire_Linguistique

Mémoires et thèses

Balland O, Courtade Jouanicq A. *Etude de faisabilité d'un logiciel de reconnaissance vocale adapté à des tâches d'évocation lexicale*. [Mémoire présenté en vue de l'obtention du Certificat de Capacité d'Orthophoniste]. Toulouse : Université Paul Sabatier Toulouse III; 2015.

Février F. *Vers un modèle intégrateur « expérience-acceptation » : rôle des affects et de caractéristiques personnelles et contextuelles dans la détermination des intentions d'usage d'un environnement numérique de travail*. [Thèse de Doctorat d'Université, Psychologie cognitive]. Rennes : Université Rennes 2, Université Européenne de Bretagne;2011.
[en ligne] (page consultée le 06/06/16) <<https://tel.archives-ouvertes.fr/tel-00608335/document>>

Fleurier A. « *Et si on bavardait un peu...* » : *Etude de l'intérêt de l'évaluation du langage spontané lors de neurochirurgies éveillées et création d'un protocole*. [Mémoire présenté en vue de l'obtention du Certificat de Capacité d'Orthophoniste]. Toulouse : Université Paul Sabatier Toulouse III; 2013.

Sahraoui H. *Contribution à l'étude des stratégies compensatoires dans l'agrammatisme. Approche neuropsycholinguistique de la performance de six locuteurs agrammatiques en production orale : caractérisation quantitative et fonctionnelle des variabilités*. [Thèse de Doctorat d'Université, Sciences du langage]. Toulouse : Université de Toulouse II-Le Mirail;2009.

Tests

Ducarne de Ribaucourt B. *Test pour l'examen de l'aphasie*. Paris: Editions du Centre de Psychologie Appliquée; 1989.

Gatignol P, Jutteau S, Oudry M, Weill-Chounlamountry A. *Bilan informatisé d'Aphasie*. Isebergues: Ortho Edition; 2012.

Gatignol P, Marin Curtoud S, ERU 16. *BIMM - Batterie informatisée du manque du mot*. Paris: Editions du Centre de Psychologie Appliquée; 2007.

Hammelrath C. *Test de Dénomination de Verbes Lexicaux en images DVL 38*. Isebergues: Ortho Edition; 1999.

Holland A, MacWhinney B, Fromm D, Forbes M. *AphasiaBank Repetition Test*. 2007.

Mazaux JM, Orgogozo JM. *Boston diagnostic aphasia examination - HDAE - échelle française*. Paris: ESP; 1981.

Nespoulous JL, Lecours AR, Lafond D, Lemay A, Puel M, Joannette Y, et al. *Protocole Montréal-Toulouse d'examen linguistique de l'aphasie*. Isebergues: Ortho Edition; 1986.

Rousseaux M, Dei Cas P. *Test de Langage Elaboré pour adultes*. Isebergues: Ortho Edition; 2012.

Sainson C, Guyou H. *Grille d'Analyse Linguistique d'Interaction libre*. Isebergues: Ortho Edition; 2016.

Signoret JL. *Batterie d'efficience mnésique, BEM 144*. Paris: Elsevier; 1991.

BIBLIOGRAPHIE D'AUTEURS CITÉS PAR D'AUTRES AUTEURS

Ouvrages, chapitres d'ouvrages et articles

Austin JL. *Quand dire, c'est faire*. (trad. fr. par Lane G). (1ère éd. *How to do Things with Words*. Oxford ; 1962). Paris: Seuil; 1970.

Benveniste E. *Problèmes de linguistique générale*. Paris: Gallimard; 1966.

Charaudeau P. *Grammaire du sens et de l'expression*. Paris: Hachette; 1992.

Damasio AR. *Signs of aphasia*. In: Sarno MT, editor. *Acquired aphasia*. San Diego: Academic Press; 1991. p. 27-43.

Ducrot O. *Dire et ne pas dire. Principes de sémantique linguistique*. Paris: Hermann; 1972.

Fournié M. *Essai de psychologie*. Paris; 1887.

Gardiner AH. *Langage et acte de langage. Aux sources de la pragmatique*. (1ère éd. *The Theory of Speech and Language*. Oxford : Clarendon Press; 1932). Presses universitaires de Lille; 1989.

Garrett M. *The analysis of sentence production*. In: Bower G, editor. *Psychology of learning and motivation, Vol 9*. New York: Academic Press; 1975.

Garrett M. *Syntactic processes in sentence production*. In: Wales R, Walker E, editors. *New approaches to language mechanisms* Amsterdam: North Holland; 1976. p. 231-256.

Garrett M. *The organization of processing structure for language production*. In Caplan D, Lecours AR, Smith A, editors. *Biological perspectives on language*. Cambridge: MIT Press; 1984. p. 172-193.

Garrett M. *Processes in language production*. In: Newmeyer FJ, editor. *Linguistics : The Cambridge survey III. Language : Psychological and biological aspects*. Cambridge: Cambridge University Press; 1988. p. 69-96.

Goldman-Eisler F. *Psycholinguistics : Experiments in spontaneous speech*. New York: Academic Press; 1968.

Gordon D, Lakoff G. « Postulats de conversation », *Langages*, 30; 1973: 32-55.

Grice HP. « *Logique et conversation* ». (trad. fr.). *Communications*, 30 (1ère éd. « *Logic and conversation* » . In: COLE P, MORGAN JL, editors *Syntax and Semantics*, vol. III, *Speech Acts*. New York : Academic Press; 1975 :41-58); 1979: 57-72.

Guillaume G. *Principes de linguistique théorique*. Valin R, éditeur. Québec : Presses de l'université Laval/Paris, Klincksieck; 1973.

Lashley KS. *The problem of serial order in behaviour*. In: Jeffress LA, editor. *Cerebral mechanisms in behaviour*. New York: Wiley; 1951. p. 112-136.

Maclay H, Osgood CE. *Hesitation phenomena in spontaneous English speech*. *Word*, 15. 1959. p. 19-44.

Martinet A. *Éléments de linguistique générale*. Paris: Armand Colin; 1967.

Saffran EM, Berndt RS, Schwartz MF. *The quantitative analysis of agrammatic production: Procedure and data*. *Brain and Language*, 37. 1989. p. 440-79.

Saussure de F. *Cours de linguistique générale*, édition critique de T. de Mauro. (1ère ed. 1916). Paris : Payot; 1972.

Searle JR. *Les Actes de langage*. (trad. fr.). (1ère ed. *Speech Acts*. Cambridge : Cambridge University Press;1969). Paris: Hermann; 1972.

Ska B, Duong A, Joannette Y. *Discourse impairments*. In: Kent RD, editor. *The MIT encyclopedia of communication disorders*. Cambridge, MA: The MIT Press; 2004. p. 302-304.

Tricot A, Plégat-Soutjis F, Camps JF, Amiel A, Lutz G, Morcillo A. *Utilité, utilisabilité, acceptabilité : interpréter les relations entre trois dimensions de l'évaluation des EIAH*. In: Desmoulins C, Marquet P, Bouhineau D. *Environnements Informatiques pour l'Apprentissage Humain*. Apr 2003. Strasbourg,, France. ATIEF; INRP, pp.391-402. [en ligne] (page consultée le 06/06/16) <<https://hal.inria.fr/file/index/docid/1674/filename/n036-80.pdf>>

Tests

Beukelman D, Mirenda P. *Augmentative and alternative communication: management of severe communication disorders in children and adults*. Baltimore, MD: Paul H. Brookes Publishing Co; 1998.

Brink TL, Yesavage JA, Lum O, Heersema P, Adey MB, Rose TL. *Screening tests for geriatric depression*. *Clinical Gerontologist*, 1, 37-44.1982.

Darrigrand B, Mazaux JM. *L'Echelle de communication verbale de Bordeaux (E.C.V.B.)*. Isebergues: Ortho Edition; 2002.

Davis H, Silverman SR, editors. *Hearing and deafness*. 4th ed. New York :Holt, Rinehart & Winston; 1978.

De Partz MP, Bilocq V, De Wilde V, Seron X, Pillon A. *Lexis*. Paris: Solal; 2002.

Deloche G, Hannequin D. *DO80*. Paris : Editions du Centre de Psychologie Appliquée; 2007.

Folstein M, Folstein S, Fanjiang G. *Mini-mental State Examination*. Lutz, FL: Psychological Assessment Resources, Inc.; 2002.

Goodglass H, Kaplan E. *The Assessment of Aphasia and Related Disorders*. Philadelphia: Lea & Febiger; 1972.

Goodglass H, Kaplan E, Barresi B. *Boston Diagnostic Aphasia Examination*. 3rd ed. Philadelphia: Lippincott Williams & Wilkins; 2001.

Joannette Y, Ska B, Côté H. *Protocole Montréal d'Evaluation de la Communication (MEC)*. Isbergues: Ortho Edition; 2005.

Kaplan E, Goodglass H, Weintraub S. *Boston Naming Test*. 2nd ed. Austin, TX: Pro-Ed; 2001.

Kertesz A. *Western Aphasia Battery-Revised*. San Antonio, TX: Psychological Corporation; 2007.

Rousseaux M, Delacourt A, Wyrzykowski N, Lefeuve M. *TLC: Test Lillois de Communication*. Isbergues: Ortho Edition; 2001.

Saffran, EM, Schwartz MF, Linebarger M, Martin N, Bochetto P. *The Philadelphia Comprehension Battery for Aphasia*. 1988. *Unpublished manuscript*.

Thompson CK. *Northwestern assessment of verbs and sentences – revised*. Evanstons, IL: Northwestern University Press; 2011. *Manuscript in preparation*.

LISTE DES FIGURES, TABLEAUX, DIAGRAMMES

LISTE DES FIGURES

Figure n°1 : Modèle de Levelt (tiré de Levelt, 1999)	20
--	----

LISTE DES TABLEAUX

Tableau n°1 : Les aphasies à langage réduit (d'après Gil, 1999)	26
Tableau n°2 : Les aphasies à langage fluide (d'après Gil, 1999)	26
Tableau n°3 : Mesures EVAL – Mesures générales (d'après EVAL Manual, 2016).....	50
Tableau n°4 : Mesures EVAL – Calcul des différentes catégories de mots (d'après EVAL Manual, 2016).....	51
Tableau n°5 : Mesures EVAL - Nombre de répétitions et de reformulations (d'après EVAL Manual, 2016).....	52
Tableau n°6 : Adaptation des épreuves complémentaires du protocole AphasiaBank en français..	76
Tableau n°7 : Description de la population contrôle	82
Tableau n°8 : Description de la population aphasique – Données générales	84
Tableau n°9 : Description de la population aphasique – Données médicales	84
Tableau n°10 : Durée des passations auprès des sujets contrôles.....	87
Tableau n°11 : Durée moyenne de la passation de chaque tâche de discours du protocole chez les participants aphasiques	89
Tableau n°12 : Durée de la passation des épreuves de discours et des épreuves complémentaires auprès des sujets aphasiques.....	94
Tableau n°13 : Sélection des normes générales de codage	96
Tableau n°14 : Sélection des codages d'erreurs au niveau du mot	97
Tableau n°15 : Sélection des codages d'erreurs au niveau de l'énoncé	98
Tableau n°16 : Description des en-têtes Gem et des tâches de discours correspondantes.....	100
Tableau n°17 : Fichiers lexiques de morphèmes distincts bien que transcrits de façon solidarisée	103
Tableau n°18: Conventions adoptées pour solidariser les morphèmes 1/2.....	104
Tableau n°19 : Conventions adoptées pour solidariser les morphèmes 2/2.....	104
Tableau n°20 : Durée du travail consacré à la transcription et au codage des données et de l'enregistrement des sujets contrôles.....	109
Tableau n°21 : Durée du travail consacré à la transcription et au codage des données et de l'enregistrement des sujets aphasiques	109
Tableau n°22 : Comparaison entre les durées de transcription et de codage et d'enregistrement entre sujets contrôles et sujets aphasiques.....	110
Tableau n°23 : Comparaison du nombre d'erreurs entre les sujets contrôles et les sujets aphasiques.....	112
Tableau n°24 : Nombre total de mots non reconnus par MOR et nombre total de mots qui ont dû être ajoutés au nouveau fichier lexique	114

LISTE DES DIAGRAMMES

Diagramme n°1 : Répartition des erreurs au niveau du mot dans les corpus des sujets contrôles et aphasiques.....	111
Diagramme n°2 : Répartition des erreurs au niveau de l'énoncé dans les corpus des sujets contrôles et aphasiques	111
Diagramme n°3 : Pourcentage des différentes catégories morphologiques des mots entrés dans le fichier lexique.....	115

TABLES DES MATIERES

GLOSSAIRE

INTRODUCTION	1
PARTIE 1 : PARTIE THEORIQUE	3
1. Le discours	3
1.1. Le discours : approche linguistique	3
1.1.1. Le discours : généralités (d'après Charaudeau et Maingueneau, 2002)	3
1.1.2. Les différents modes d'organisation du discours : prolongement pour la clinique	6
1.1.3. Notion de compétence discursive et enjeu pour la clinique	7
1.1.4. Les règles du discours : apport de la pragmatique	8
1.1.5. Les particularités du discours oral	9
1.1.5.1. Discours oral et discours écrit	9
1.1.5.2. Les différents registres de langue	12
1.2. Le discours oral : approche psycholinguistique (d'après Fromkin et Ratner, 1998).....	12
1.2.1. Les phénomènes de l'oral : une fenêtre sur les processus psycholinguistiques sous-jacents	13
1.2.1.1. Généralités.....	13
1.2.1.2. Apports de l'étude des erreurs de performance : de la normalité à la pathologie	13
1.2.1.2.1. Les unités de production du discours oral	13
1.2.1.2.2. Les principes fondamentaux à l'œuvre dans les processus de production du discours oral.....	16
1.2.1.2.2.1. Le discours est organisé à l'avance	16
1.2.1.2.2.2. Le lexique est organisé sémantiquement et phonologiquement.....	16
1.2.1.2.2.3. Les mots morphologiquement complexes sont assemblés.....	17
1.2.1.2.2.4. Les erreurs affectent différemment les affixes, les mots fonction, les éléments mineurs et les éléments majeurs de la phrase	17
1.2.1.2.2.5. Les erreurs de discours reflètent la connaissance de la règle.....	18
1.2.2. Les principaux modèles de production du discours oral.....	18
1.2.2.1. Le modèle de Levelt.....	20
2. L'aphasie	22
2.1. Définition de l'aphasie	22
2.2. Etiologie des aphasies.....	23
2.3. Perturbations langagières orales observées.....	23
2.4. Les différents types d'aphasie	24
2.4.1. Les aphasies à langage réduit (d'après Gil, 1999).....	25
2.4.2. Les aphasies à langage fluide (d'après Gil, 1999).....	26
2.5. Handicap et communication dans l'aphasie : mieux évaluer le discours déviant.....	27
3. L'évaluation en clinique orthophonique	29
3.1. L'évaluation orthophonique de l'aphasie.....	29
3.1.1. Domaines de l'évaluation et outils existants	29
3.1.2. L'analyse et l'interprétation des résultats.....	33
3.2. L'évaluation orthophonique du discours continu oral en aphasiologie.....	33
3.2.1. Les paramètres d'évaluation du discours.....	34
3.2.2. Les outils d'évaluation orthophoniques du discours en aphasiologie.....	34
3.2.3. Synthèse des outils actuels.....	38

4. L'utilisation des outils informatiques dans l'étude du discours en aphasiologie	39
4.1. Outils informatisés actuels dédiés à l'évaluation orthophonique de l'aphasie en France...	39
4.2. Les bases de données et les projets de développement collaboratif dans l'étude du discours en aphasiologie	41
4.2.1. Les principes d'une base de données.....	41
4.2.2. La base de données internationale AphasiaBank : intérêt pour la recherche fondamentale.....	42
4.3. Présentation des outils utilisés pour le traitement automatique informatisé du langage oral au sein du projet AphasiaBank	45
4.3.1. Le système CHAT	45
4.3.2. Le logiciel CLAN	46
4.4. Analyse informatisée du discours en aphasiologie : fonctionnement du programme EVAL	46
4.4.1. EVAL : généralités	46
4.4.2. Préparation d'une transcription pour l'analyse EVAL	47
4.4.3. Les mesures linguistiques calculées par EVAL	50
4.4.4. Le lancement du programme EVAL	52
4.4.4.1. Comparaison d'un sujet à la base de données.....	52
4.4.4.2. Comparaison d'un même sujet à deux temps différents	52
PARTIE 2 : PROBLEMATIQUE ET HYPOTHESES	54
1. Problématique	54
1.1. Raisons du choix de cette problématique : Intérêts potentiels de l'analyse informatisée par EVAL du discours aphasique dans l'évaluation clinique orthophonique	54
1.1.1. L'importance de l'évaluation du discours en aphasiologie.....	54
1.1.2. Apport potentiel d'EVAL par rapport aux autres outils d'évaluation du discours	54
1.1.3. L'intérêt de l'utilisation d'EVAL dans l'évaluation du discours.....	55
1.2. Problématique résumée.....	57
2. Hypothèses	58
PARTIE 3 : PARTIE EXPERIMENTALE	60
1. Adaptation du protocole de recueil de données AphasiaBank	60
1.1. Méthodologie dédiée à l'adaptation du protocole de recueil de données AphasiaBank....	60
1.1.1. Adaptation des documents permettant le recueil des données démographiques.....	61
1.1.1.1. Description des formulaires originaux à faire remplir par les participants	61
1.1.1.2. Adaptation des formulaires à faire remplir par les participants.....	62
1.1.1.3. Description des tableaux de recueil de données démographiques originaux de la base de données AphasiaBank.....	62
1.1.1.4. Adaptation des tableaux de recueil de données démographiques de la base de données AphasiaBank	63
1.1.2. Adaptation du protocole de recueil de données langagières AphasiaBank	64
1.1.2.1. Adaptation des épreuves de discours.....	64
1.1.2.1.1. Description du protocole original destiné aux participants contrôles.....	64
1.1.2.1.2. Adaptation du protocole destiné aux participants contrôles	65
1.1.2.1.3. Description du protocole destiné aux participants aphasiques	66
1.1.2.1.4. Adaptation du protocole destiné aux participants aphasiques	66
1.1.2.2. Adaptation des épreuves complémentaires	67
1.1.2.2.1. Description des épreuves complémentaires originales.....	67
1.1.2.2.2. Adaptation des épreuves équivalentes : prise en compte du contexte clinique en France.....	79

1.1.2.2.3. Adaptation du tableau de résultats.....	73
1.2. Résultats concernant l'adaptation du protocole de recueil de données AphasiaBank	74
1.2.1 Résultats concernant l'adaptation des documents permettant le recueil de données démographiques	74
1.2.1.1 Résultats concernant l'adaptation du formulaire destiné aux participants	74
1.2.1.2. Résultats concernant l'adaptation des tableaux de recueil de données démographiques de la base de données AphasiaBank.....	75
1.2.2. Résultats concernant l'adaptation des épreuves de discours AphasiaBank.....	75
1.2.2.1. Résultats concernant l'adaptation des épreuves de discours destinées aux participants contrôles	75
1.2.2.2. Résultats concernant l'adaptation des épreuves de discours destinées aux participants aphasiques.....	76
1.2.2.3. Résultats concernant l'adaptation des épreuves complémentaires.....	76
1.2.2.4. Résultats concernant l'adaptation du tableau de recueil des résultats aux épreuves complémentaires	78
2. Population	79
2.1. Critères de sélection des sujets	79
2.1.1. Critères de sélection de la population contrôle	79
2.1.2. Critères de sélection de la population aphasique	80
2.2. Description de l'échantillon de population	81
2.2.1. Description de la population contrôle	81
2.2.2. Description de la population aphasique	82
3. Le recueil de données	85
3.1. Méthodologie dédiée au recueil de données	85
3.1.1. Création d'une lettre d'information et d'un formulaire de consentement éclairé	85
3.1.2. Le recueil de données démographiques	85
3.1.3. Le recueil de données langagières	85
3.2. Résultats concernant l'utilisation du protocole AphasiaBank avec des sujets contrôles et des sujets aphasiques francophones.....	86
3.2.1. Résultats concernant le recueil de données démographiques	86
3.2.1.1 Résultats concernant l'utilisation du formulaire à remplir par les participants contrôles	86
3.2.1.2 Résultats concernant l'utilisation du formulaire à remplir par les participants aphasiques	86
3.2.2. Résultats concernant le recueil de données langagières.....	87
3.2.2.1. Résultats concernant le recueil de données langagières auprès des participants contrôles	87
3.2.2.1.1. Durée de la passation	87
3.2.2.1.2. Conditions de passation.....	88
3.2.2.1.3. Utilisation du protocole	88
3.2.2.2. Résultats concernant le recueil de données langagières auprès des participants aphasiques	89
3.2.2.2.1. Résultats concernant le recueil de données discursives auprès des participants aphasiques.....	89
3.2.2.2.1.1. Durée de la passation	89
3.2.2.2.1.2. Concernant les conditions de passation	89
3.2.2.2.1.3. Utilisation du protocole de discours	90
3.2.2.2.2. Résultats concernant le recueil de données issues des épreuves complémentaires.....	93
3.2.2.2.2.1 Durée de la passation	93

3.2.2.2.2. Utilisation des épreuves complémentaires.....	93
3.2.2.2.3. Résultats concernant le recueil de l'ensemble des données langagières	93
4. L'opérationnalisation des données pour l'utilisation d'ÉVAL	94
4.1. Méthodologie de l'opérationnalisation des données pour l'utilisation d'ÉVAL	94
4.1.1. Sélection des conventions de transcription et des codages pertinents	94
4.1.1.1. Sélection des normes générales de transcription et des codages.....	95
4.1.1.2. Sélection des codages d'erreurs.....	97
4.1.2. Transcription et codage des données	99
4.1.3. Vérification des transcriptions	105
4.1.4. Alignement des transcriptions avec les fichiers audio	105
4.1.5. Le lancement du programme d'étiquetage morphologique.....	105
4.1.5.1. Reconnaissance des mots par MOR.....	106
4.1.5.2. Création d'un nouveau fichier lexique français.....	106
4.1.5.3. Lancement du programme MOR.....	107
4.1.6. Contribution à la base de données AphasiaBank	107
4.1.7. Le lancement du programme ÉVAL	108
4.2. Résultats de l'opérationnalisation des données pour l'utilisation d'ÉVAL.....	108
4.2.1. Transcription et codage des données selon les normes CHAT	108
4.2.1.1. Durée du travail consacré à la transcription et au codage des données.....	108
4.2.1.2. Utilisation des codages et des normes générales de transcription.....	110
4.2.1.3. Utilisation des codages d'erreurs.....	110
4.2.2. Vérification des transcriptions par le programme CHECK	112
4.2.3. Alignement aux fichiers audio	113
4.2.4. L'étape d'étiquetage morphologique	113
4.2.4.1. Reconnaissance des mots par MOR et création d'un nouveau fichier lexique.....	113
4.2.4.2. La constitution du nouveau fichier lexique.....	115
4.2.4.3. Le lancement du programme d'étiquetage morphologique	115
4.2.5. La mise en ligne de nos corpus sur la base de données	115
4.2.6. Le lancement du programme ÉVAL	116
PARTIE 4 : DISCUSSION DES RÉSULTATS	117
1. Interprétation des résultats et validation ou invalidation des hypothèses	117
1.1. Hypothèse 1 : Le protocole AphasiaBank est adaptable à la langue française	117
1.1.1. Sous-hypothèse 1 : Le protocole de recueil de données démographiques AphasiaBank est adaptable à la langue française	117
1.1.1.1. Synthèse des résultats.....	117
1.1.1.2. Discussion, intérêts et limites.....	117
1.1.1.3. Conclusion.....	118
1.1.2. Sous-hypothèse 2 : Le protocole de recueil de données discursives AphasiaBank est adaptable à la langue française	118
1.1.2.1. Synthèse des résultats.....	118
1.1.2.2. Discussion, intérêts et limites.....	118
1.1.2.3. Conclusion.....	120
1.1.3. Sous-hypothèse 3 : Le protocole de recueil de données issues des épreuves complémentaires AphasiaBank est adaptable à la langue française	120
1.1.3.1. Synthèse des résultats.....	120
1.1.3.2. Discussion, intérêts et limites.....	120
1.1.3.3. Conclusion.....	123
1.2. Hypothèse 2 : Le protocole AphasiaBank est utilisable avec des sujets contrôles et des sujets aphasiques francophones	123

1.2.1. Sous-hypothèse 1 : Le protocole de recueil de données démographiques AphasiaBank est utilisable avec des sujets contrôles et des sujets aphasiques francophones.....	124
1.2.1.1. Synthèse des résultats.....	124
1.2.1.2. Discussion, intérêts et limites.....	124
1.2.1.3. Conclusion.....	125
1.2.2. Sous-hypothèse 2 : Le protocole de recueil de données discursives AphasiaBank est utilisable avec des sujets contrôles et des sujets aphasiques francophones	125
1.2.2.1. Synthèse des résultats.....	125
1.2.2.2. Discussion, intérêts et limites.....	126
1.2.2.3. Conclusion.....	129
1.2.3. Sous-hypothèse 3 : Le protocole de recueil de données des épreuves complémentaires AphasiaBank est utilisable avec des sujets contrôles et des sujets aphasiques francophones	129
1.2.3.1 Synthèse des résultats.....	129
1.2.3.2. Discussion, intérêts et limites.....	129
1.2.3.3. Conclusion.....	130
1.3. Hypothèse 3 : EVAL est opérationnel avec des corpus issus de sujets contrôles et de sujets aphasiques francophones	130
1.3.1. Sous-hypothèse 1 Les corpus francophones peuvent être transcrits et codés selon les normes CHAT	130
1.3.1.1. Synthèse des résultats.....	130
1.3.1.2. Discussion, intérêts et limites.....	131
1.3.1.3. Conclusion.....	135
1.3.2. Sous-hypothèse 2 : Les corpus francophones peuvent être vérifiés par le programme CHECK et les erreurs détectées peuvent être corrigées.....	135
1.3.2.1. Synthèse des résultats.....	135
1.3.2.2. Discussion, intérêts et limites.....	135
1.3.2.3. Conclusion.....	135
1.3.3. Sous-hypothèse 3 : Le programme d'étiquetage morphologique peut être lancé sur les corpus francophones	136
1.3.3.1. Synthèse des résultats.....	136
1.3.3.2. Discussion, intérêts et limites.....	136
1.3.3.3. Conclusion.....	137
1.3.4. Sous-hypothèse 4 : Les corpus francophones peuvent être mis en ligne sur la base de données française d'AphasiaBank.....	137
1.3.4.1. Synthèse des résultats.....	137
1.3.4.2. Discussion, intérêts et limites.....	137
1.3.4.3. Conclusion.....	138
1.3.5. Sous-hypothèse 5 : Le programme EVAL peut être lancé de façon opérationnelle sur les corpus francophones.....	139
1.3.5.1. Synthèse des résultats.....	139
1.3.5.2. Discussion, intérêts et limites.....	139
1.3.5.3. Conclusion.....	139
2. Les limites.....	139
2.1. Concernant la population.....	139
2.2. Concernant le projet AphasiaBank	140
2.2.1. Limites des critères d'inclusion.....	140
2.2.2. Limites de la transcription et du codage des données.....	141
3. Perspectives.....	143
3.1.Perspectives en recherche fondamentale	143

3.1.1. Perspectives à court terme.....	143
3.1.2. Perspectives à moyen et long terme.....	144
3.2. Perspectives concernant l'utilisation du protocole AphasiaBank et d'EVAL en clinique orthophonique.....	145
3.2.1. Perspectives à court terme.....	145
3.2.2. Perspectives à moyen et long terme.....	145
CONCLUSION.....	149
BIBLIOGRAPHIE.....	150
BIBLIOGRAPHIE D'AUTEURS CITÉS PAR D'AUTRES AUTEURS.....	153
LISTE DES FIGURES, TABLEAUX, DIAGRAMMES.....	156
TABLES DES MATIERES.....	157
RESUME.....	173

Résumé

Le discours oral, central dans la communication, doit être évalué au sein du bilan orthophonique du patient aphasique. L'analyse fournie par les outils actuels, essentiellement qualitative, permet de dresser un profil qui pourrait encore s'affiner avec des analyses quantitatives complémentaires objectivées à l'aide d'EVAL, un outil d'analyse automatique du discours oral.

En vue de pouvoir utiliser EVAL, le protocole AphasiaBank a été élaboré par une équipe de chercheurs et cliniciens en anglais afin de recueillir des échantillons de discours oral et d'en proposer une analyse informatisée. EVAL permet ainsi de fournir des données quantitatives et qualitatives précises sur les aspects formels du discours oral du patient aphasique, en comparant les résultats obtenus, pour un patient en particulier, à une base de données de sujets contrôles ou aphasiques.

Dans un premier temps, nous avons testé la possibilité d'adapter et d'utiliser le protocole auprès de 14 sujets contrôles et 11 sujets aphasiques francophones, dont les corpus ont été déposés sur AphasiaBank. Nous avons ensuite testé la possibilité d'utiliser le programme EVAL sur les échantillons de discours ainsi recueillis.

Les résultats ont montré qu'il est possible d'adapter et d'utiliser le protocole AphasiaBank en langue française, celui-ci nécessitant quelques améliorations. Ils ont également montré que, s'il est possible de transcrire les échantillons de discours et de les préparer pour l'analyse EVAL, à ce jour le programme n'est pas encore opérationnel pour la langue française. A terme, il sera possible d'obtenir des résultats exploitables permettant de situer le profil discursif d'un participant par rapport à la base de données française.

Mots-clés : Aphasie, Evaluation, Analyse du discours, Outils informatisés, AphasiaBank, EVAL.

Abstract

Aphasic patient's speech needs to be assessed in therapy. The analysis provided by the current tools is essentially qualitative. It draws a profile that may still be refined with additional objective analyzes using EVAL, an automatic analysis tool dedicated to speech assessment.

In order to run EVAL, the AphasiaBank protocol has been developed by a team of researchers and clinicians in English, aiming at collecting control and aphasic speech production samples and propose a computerized analysis. EVAL provides more accurate quantitative and qualitative speech analyses, by comparing a particular patient's profile to a control or a selective aphasic data set.

Firstly, we made and tested an adaptation of the protocol to French. So we performed the protocol with 14 control and 11 aphasic French-speaking subjects. Their speech has been transcribed and put on the Aphasiabank database. Secondly we tested the possibility of using the EVAL program on French speech samples.

Our study showed that it is now possible to adapt and use the AphasiaBank protocol in French, although some improvements are steel needed. We also concluded that even if it is now possible to transcribe speech samples and prepare them for EVAL analysis, the program is not yet completely operational for the French language. Still, further work to make EVAL fully useful will contribute to a better assessment of speech in aphasic speakers for therapy.

Keywords : Aphasia, Assessment, Discourse analysis, Computerized analysis, AphasiaBank, EVAL.