



MEMOIRE présenté pour l'obtention du
CERTIFICAT DE CAPACITE D'ORTHOPHONISTE

Par

MAILLÉ Elsa
THUILLIER Constance

MESURE DE LA DENSITÉ DES IDÉES DANS LE
VIEILLISSEMENT COGNITIF NORMAL EN FONCTION
DU NIVEAU SOCIOCULTUREL ET DE L'ÂGE.

*Incidences sur l'analyse des productions orales de patients atteints
de la maladie d'Alzheimer.*

Maître de Mémoire

Dr. BARKAT-DEFRADAS Mélissa

**Chargée de Recherches UMR 5267 CNRS & Université de
Montpellier**

Membres du Jury

AUJOGUES Emmanuelle

GAYRAUD Frédérique

PEILLON Anne

Date de Soutenance

01 juillet 2010

1. Université Claude Bernard Lyon1

Président
Pr. COLLET Lionel

Vice-président CEVU
Pr. SIMON Daniel

Vice-président CA
Pr. ANNAT Guy

Vice-président CS
Pr. MORNEX Jean-François

Secrétaire Général
M. GAY Gilles

1.1. Secteur Santé :

U.F.R. de Médecine Lyon Est
Directeur **Pr. ETIENNE Jérôme**

U.F.R d'Odontologie
Directeur **Pr. BOURGEOIS Denis**

U.F.R de Médecine Lyon-Sud
Charles Mérieux
Directeur **Pr. GILLY François
Noël**

Institut des Sciences
Pharmaceutiques et Biologiques
Directeur **Pr. LOCHER François**

Institut des Sciences et Techniques de
Réadaptation
Directeur **Pr. MATILLON Yves**

Comité de Coordination des
Etudes Médicales (C.C.E.M.)
Pr. GILLY François Noël

Département de Formation et Centre
de Recherche en Biologie Humaine
Directeur **Pr. FARGE Pierre**

1.2. Secteur Sciences et Technologies :

U.F.R. de Sciences et Technologies
Directeur **Pr GIERES François**

IUFM
Directeur **M. BERNARD Régis**

U.F.R. de Sciences et Techniques
des Activités Physiques et
Sportives (S.T.A.P.S.)
Directeur **Pr. COLLIGNON
Claude**

Ecole Polytechnique Universitaire de
Lyon (EPUL)
Directeur **Pr. LIETO Joseph**

Institut des Sciences Financières
et d'Assurance (I.S.F.A.)
Directeur **Pr. AUGROS Jean-Claude**

Ecole Supérieure de Chimie Physique
Electronique de Lyon (CPE)
Directeur **M. PIGNAULT Gérard**

IUT LYON 1
Directeurs **M. COULET Christian et
Pr. LAMARTINE Roger**

2. **Institut Sciences et Techniques de Réadaptation FORMATION
ORTHOPHONIE**

Directeur ISTR
Pr. MATILLON Yves

Directeur de la formation
Pr. TRUY Eric

Directeur des études
BO Agnès

Directeur de la recherche
Dr. WITKO Agnès

Responsables de la formation clinique
THEROND Béatrice

GUILLON Fanny

Chargée du concours d'entrée
PEILLON Anne

Secrétariat de direction et de scolarité
BADIOU Stéphanie

CLERGET Corinne

REMERCIEMENTS

REMERCIEMENTS

Nous souhaitons tout d'abord remercier notre maître de mémoire, Mélissa Barkat-Defradas pour nous avoir orientées vers ce sujet passionnant qui nous a permis d'approfondir nos connaissances en linguistique. Merci pour la confiance qu'elle nous a accordée, son soutien, et l'aide apportée tout au long de notre mémoire.

Un immense merci à Hye Ran Lee pour ses apports théoriques précieux, sa grande générosité et le temps qu'elle nous a consacré malgré un emploi du temps surchargé. Merci également à Pierre pour son accueil chaleureux, le gîte et le couvert !

Nous tenons à remercier tout particulièrement Ariane Delemasure pour sa disponibilité et ses précieux conseils cliniques.

Merci beaucoup à Philippe Gambette grâce à qui le logiciel Densidées a pu voir le jour. Merci pour sa grande patience et son aide précieuse.

Nous souhaitons également remercier Agnès Witko pour son soutien méthodologique ainsi qu'Anne-Laure Charlois pour la rigueur apportée à notre travail statistique.

Merci à toutes les personnes qui ont accepté de nous rencontrer, et qui ont permis la réalisation de ce mémoire.

Un grand merci à nos familles pour nous avoir soutenues tout au long de nos études dans les moments difficiles comme dans les bons moments.

Un non moins grand merci à Thomas et Alexandre pour nous avoir supportées au sens large du terme ! Merci pour votre patience, votre soutien, vos voitures et les dépannages !

Enfin, merci à tous nos amis, merci à la solidarité et à la gaieté orthophonique...

SOMMAIRE

ORGANIGRAMMES	2
1. <i>Université Claude Bernard Lyon1</i>	2
2. <i>Institut Sciences et Techniques de Réadaptation FORMATION ORTHOPHONIE</i>	3
REMERCIEMENTS	4
SOMMAIRE.....	5
INTRODUCTION.....	7
PARTIE THEORIQUE.....	9
II. CONSEQUENCES DU VIEILLISSEMENT COGNITIF SUR LE LANGAGE.....	10
1. <i>Le vieillissement cognitif normal</i>	10
2. <i>Retentissements du vieillissement cognitif sur le langage.....</i>	11
3. <i>Facteurs influençant l'évolution du langage.....</i>	12
4. <i>Prise en compte du niveau socioculturel dans l'étude de langage.....</i>	12
III. CONSEQUENCES DU VIEILLISSEMENT COGNITIF PATHOLOGIQUE SUR LE LANGAGE : LA MALADIE D'ALZHEIMER	13
1. <i>La maladie d'Alzheimer.....</i>	13
2. <i>Conséquences de la maladie d'Alzheimer sur le langage</i>	15
IV. LA DENSITE DES IDEES	16
1. <i>Définition</i>	16
2. <i>Paramètres pouvant influencer le score de densité des idées</i>	16
3. <i>Evaluation de la densité des idées dans le discours.....</i>	17
4. <i>Mesure de la densité des idées</i>	19
5. <i>Un exemple d'analyse automatique de l'anglais, Computerized Propositional Idea Density Rater : CPIDR 3.2 (Brown, Snodgrass, Kemper, Herman, & Covington, 2008)</i>	23
PROBLEMATIQUE ET HYPOTHESES.....	25
I. PROBLEMATIQUE	26
II. HYPOTHESES.....	27
1. <i>Hypothèses générales.....</i>	27
2. <i>Hypothèses opérationnelles</i>	27
PARTIE EXPERIMENTALE	28
I. POPULATION.....	29
1. <i>Sujets sains.....</i>	29
2. <i>Sujets atteints de la maladie d'Alzheimer</i>	31
II. PROTOCOLE EXPERIMENTAL	31
1. <i>Objectif.....</i>	31
2. <i>Elaboration du protocole expérimental</i>	31
3. <i>Protocole de passation.....</i>	33
4. <i>Matériel.....</i>	34
III. CONDITIONS D'EXPERIMENTATION	34
1. <i>Passation.....</i>	34
2. <i>Lieu d'expérimentation</i>	35
3. <i>Durée de la passation</i>	35
IV. ANALYSE DES DONNEES.....	35
1. <i>Création du logiciel d'analyse automatique de la densité des idées : Densidées</i>	35
2. <i>Procédure d'analyse de nos corpus</i>	38
PRESENTATION DES RESULTATS.....	40
I. ETUDE DE LA POPULATION GENERALE	41
1. <i>Méthodologie statistique</i>	41
2. <i>Effet de l'âge.....</i>	41
3. <i>Effet du sexe.....</i>	43
4. <i>Effet du niveau socioculturel.....</i>	44
II. ETUDE DE L'ECHANTILLON DE PATIENTS ATTEINTS DE LA MALADIE D'ALZHEIMER	46

SOMMAIRE

1. <i>Méthodologie</i>	46
2. <i>Effet de la pathologie</i>	46
III. ETUDE DE L'EFFET DU TYPE DE DISCOURS SUR LA DENSITE DES IDEES	48
1. <i>Méthodologie</i>	48
2. <i>Impact du type de discours sur la DI</i>	48
3. <i>Impact du niveau socioculturel sur les scores de densité des idées obtenus pour chaque type de discours</i>	50
IV. SYNTHÈSE DES RESULTATS	51
DISCUSSION DES RESULTATS	53
I. VALIDATION DES HYPOTHESES DE TRAVAIL	54
1. <i>Effet de l'âge</i>	54
2. <i>Effet du sexe</i>	56
3. <i>Effet du niveau socioculturel</i>	56
4. <i>Effet de la pathologie</i>	57
5. <i>Effet du type de discours</i>	58
II. LIMITES ET INTERETS DE CETTE ETUDE	60
1. <i>Limites</i>	60
2. <i>Intérêts</i>	63
III. APPORT CLINIQUE	64
IV. PERSPECTIVES ET FAISABILITE	65
1. <i>Perspectives cliniques</i>	65
2. <i>Autres perspectives</i>	66
CONCLUSION	67
BIBLIOGRAPHIE	69
ANNEXES	75
ANNEXE I : GRILLE DE POITRENAUD (KALAFAT, HUGONOT-DIENER, & POITRENAUD, 2005)	76
ANNEXE II : FICHE DE RENSEIGNEMENTS	77
ANNEXE III : DENSIDEES - CALCUL AUTOMATIQUE DE LA DENSITE DES IDEES DANS UN CORPUS ORAL	78
ANNEXE IV : EXEMPLES DE CORPUS	88
ANNEXE V : SCORES DE DENSITE DES IDEES	94
TABLE DES ILLUSTRATIONS	97
1. <i>Liste des Tableaux</i>	97
2. <i>Liste des Figures</i>	98
TABLE DES MATIERES	99

INTRODUCTION

La plupart des pays occidentaux, et notamment la France, connaissent un vieillissement de la population lié à la baisse de la natalité mais également à l'augmentation de l'espérance de vie. Du fait de ce vieillissement, l'incidence des maladies neurodégénératives telle que la maladie d'Alzheimer augmente de façon exponentielle. La maladie d'Alzheimer représente 70% des syndromes démentiels, et touche 800 000 patients en France, avec 165 000 nouveaux cas par an. La prise en charge médicale et sociale de la dépendance induite par cette pathologie constitue ainsi un enjeu majeur de santé publique et les travaux de recherche, dédiés à une meilleure connaissance de cette maladie, se multiplient.

La dégradation linguistique est un élément caractéristique de la maladie d'Alzheimer. Aujourd'hui, la diminution progressive des capacités langagières est reconnue comme une importante manifestation clinique de la maladie. Les troubles du langage, qui apparaissent durant la première phase de la maladie, peuvent donc avoir une valeur diagnostique et servir de marqueurs de la progression de la maladie (Cardebat, Aithamon, & Puel, 1995). L'étude de la dimension linguistique de la maladie d'Alzheimer prend ainsi tout son sens et pousse la recherche à explorer de nouvelles perspectives.

En 1986, le docteur David Snowdon lance une importante étude épidémiologique consacrée au vieillissement et à la maladie d'Alzheimer, « The Nun study ». Son originalité est d'être longitudinale ; elle porte sur une communauté de nonnes qui ont vécu plus de 50 ans dans un environnement totalement similaire. Cette étude a recensé certains facteurs de risques de développer la maladie d'Alzheimer et a montré un lien entre compétences langagières observées à l'adolescence et probabilité de développer cette maladie au cours du vieillissement. Deux indicateurs linguistiques ont été pressentis comme indices pré morbides (Kemper, Greiner, Marquis, Prenovost, & Mitzner, 2001) : la complexité syntaxique et la densité des idées (i.e. qualité informative des propositions langagières).

Suite à la rencontre avec des linguistes et à la lecture des différents articles consacrés à cette étude, nous nous sommes particulièrement intéressées à la densité des idées. En effet, ce critère est peu étudié en français. Il nous semble toutefois intéressant de le prendre en compte dans le cadre de l'orthophonie puisqu'il renvoie à la qualité informative du discours.

Les auteurs de « The Nun Study » ont démontré une corrélation positive entre la densité des idées relevée dans les écrits autobiographiques des postulantes à l'entrée dans les ordres et leurs capacités cognitives à des âges compris entre 75 et 93 ans. De plus, l'examen clinique et neuropathologique d'un certain nombre de cas a confirmé que les critères cliniques de la maladie d'Alzheimer étaient plus souvent présents chez les nonnes qui avaient montré dans leur jeunesse une densité d'idées moins élevée (Snowdon, 1997 ; Snowdon, Greiner, Kemper, Nanayakkara, & Mortimer, 1999 ; Snowdon, Geiner, & Markesbery, 2000). Ainsi, une densité des idées faible durant l'adolescence signifierait un facteur de sensibilité accrue au déclin cognitif lié à la maladie d'Alzheimer. D'autres études ont également montré que l'apparition de la maladie serait reliée à une dégradation de la densité des idées (Kemper et al., 2001 ; Riley, Snowdon, Desrosiers, & Markesbery, 2005 ; Chand & Bonnici, 2007). De même, la DI pourrait permettre de mesurer l'avancée de la maladie (Kemper et al., 2001 ; Chand & Bonnici., 2007).

Il semble donc que l'étude de la densité des idées pourrait permettre d'élargir le champ de recherche sur la maladie d'Alzheimer, en ouvrant d'autres possibilités d'aide au diagnostic précoce et à la prise en soin anticipée de cette maladie, via l'analyse linguistique des productions des patients.

L'idée serait de distinguer les sujets ayant une évolution cognitive normale de ceux présentant une évolution pathologique grâce à la mesure de la densité des idées. Pour cela, il s'agit de déterminer au préalable des normes quant à l'évolution de ce critère lors du vieillissement cognitif normal. Cependant, nous ne pouvons étudier cette évolution qu'à l'aide d'une étude transversale. Nous allons donc devoir faire un état des lieux des variables agissant sur le vieillissement cognitif normal afin de contrôler au mieux ces paramètres et d'évaluer le plus objectivement possible l'effet du vieillissement sur la densité des idées. La description de profils normaux permettra ensuite de mieux départager l'effet du vieillissement de celui de la pathologie.

Une telle étude nécessite un regard pluridisciplinaire. Pour mener à bien ce projet de recherche, nous avons sollicité des chercheurs impliqués dans le domaine du vieillissement cognitif, autour d'un groupe de recherches montpelliérain dirigé par notre maître de mémoire : « Langage et Maladie d'Alzheimer ». Ces chercheurs viennent de différents domaines : linguistique (M. Barkat-Defradas, chercheur CNRS ; F. Gayraud, maître de conférences à l'université de Lyon ; H.R Lee, doctorante en sciences du langage), orthophonie (C. Cadilhac, Directrice de l'école d'orthophonie de Montpellier et maître de conférence en sciences du langage à l'université de Montpellier ; C. Fritsch et F. Vezolle, étudiantes en quatrième année d'orthophonie à Montpellier), neuropsychologie (L. Lefebvre, chargé de cours à l'université de Mons ; M.C. Gely-Nargeot), médecine (M. Lepape et P. Nkuka-Chesnel, médecins praticiens au centre hospitalier de Coutances), et informatique (P. Gambette, doctorant à l'université de Montpellier).

Afin de proposer un travail de recherche original et novateur, nous avons décidé d'étudier la densité des idées en langue française, les travaux menés jusqu'ici ne concernant que la langue anglaise. De plus, nous mesurerons ce critère à partir de discours oraux. En effet, la situation de production d'un discours oral spontané est plus naturelle et plus écologique que la situation de production écrite. De plus, le langage écrit se détériore rapidement avec l'avancée de la maladie d'Alzheimer, un discours oral sera plus facile à recueillir.

Dans un premier temps, nous étudierons les conséquences du vieillissement cognitif normal sur le langage et nous déterminerons les facteurs susceptibles d'influencer les productions langagières. Dans un second temps, nous identifierons les altérations linguistiques propres à la maladie d'Alzheimer ainsi que les mécanismes qui les sous-tendent. Enfin, nous définirons la densité des idées et sa méthode d'analyse. Tous ces éléments nous permettront de procéder à l'étude de la densité des idées chez une population de sujets sains. À partir de ces premières données, nous mesurerons l'incidence de la maladie d'Alzheimer sur la densité des idées pour un groupe de patients. Les résultats obtenus seront discutés à l'aide de la littérature anglo-saxonne.

Chapitre I
PARTIE THEORIQUE

II. Conséquences du vieillissement cognitif sur le langage

1. Le vieillissement cognitif normal

1.1. Définition

Le système cognitif humain est un système complexe de traitement de l'information capable d'acquérir, de conserver, d'utiliser et de transmettre des connaissances. Celui-ci va subir les conséquences du vieillissement cérébral, lié au vieillissement des neurones. En effet, dès la naissance, l'être humain dispose d'une réserve cérébrale constituée d'une importante quantité de neurones, laquelle va augmenter pendant la croissance jusqu'à l'adolescence. À l'adolescence, les circuits redondants vont être supprimés et la réserve va pouvoir continuer à augmenter à partir de cellules neuronales souches. Cependant, à partir d'un certain âge, une diminution du nombre et de la taille des neurones et une perte d'efficacité des réseaux neuronaux sont communément observées (Buée & Maurage, 2008).

C'est grâce à l'étude de pathologies cérébrales acquises en neuropsychologie et grâce aux modèles tirés de la psychologie cognitive expérimentale que l'on a pu comprendre une partie du fonctionnement cérébral. En effet, depuis une dizaine d'années, de nombreux modèles de psychologie cognitive tentent d'expliquer le déclin cognitif lié au vieillissement *via* l'étude des structures et processus de traitement par lesquels le cerveau assure la gestion des comportements humains. Elle s'attache ainsi à différents phénomènes mentaux tels que la perception, l'attention, la mémoire ou encore le raisonnement.

Le vieillissement cognitif normal pourra donc être défini « *en terme de déclin, avec l'âge, de la performance à diverses épreuves censées mesurer le fonctionnement cognitif* » (Van der Linden & Hupet, 1994, p.11). Ce déclin peut s'exprimer par la fragilisation des capacités cognitives ou encore l'installation de stratégies nouvelles.

1.2. Processus cognitifs fragilisés

L'intelligence fluide est celle que tout individu met en place pour résoudre un problème nouveau : elle s'exprime en terme de rapidité des processus mentaux, d'activation / inhibition des informations et de capacité de raisonnement. L'intelligence cristallisée correspond, quant à elle, à l'ensemble des connaissances et des expériences acquises au cours de la vie. Kaufman, Reynolds et Mc-Lean, expliquent qu' « *au cours du vieillissement normal seule l'intelligence fluide décline alors que la composante cristallisée demeure particulièrement préservée* » (1989, p.237).

Ainsi certaines opérations cognitives vont être fragilisées lors du vieillissement cérébral. D'une part un ralentissement de la vitesse de traitement des informations sera observé (Van der Linden & Hupet, 1994).

D'autre part les ressources attentionnelles vont diminuer. L'attention sélective serait notamment moins performante en raison d'un déclin de l'inhibition ; les tâches d'alternance et d'attention divisée montreraient une difficulté à se désengager d'une tâche en cours pour exécuter une nouvelle tâche ou pour partager l'attention entre deux tâches concurrentes (Bherer, 2008).

Enfin, certains systèmes de mémoire vont être fragilisés. Les effets du vieillissement sont particulièrement marqués en mémoire de travail, elle permet le maintien temporaire et la manipulation d'une information pendant la durée de la réalisation d'une tâche cognitive (Baddeley, 1986). La mémoire épisodique qui correspond aux événements personnellement vécus dans un contexte spatial et temporel spécifique sera également fragilisée. En revanche, notons que l'effet du vieillissement sur l'organisation et le fonctionnement de la mémoire sémantique (i.e. connaissances factuelles et encyclopédiques) paraît mineur (Taconnat & Isingrini, 2008).

2. Retentissements du vieillissement cognitif sur le langage

Le langage est la fonction cognitive probablement la mieux préservée lors du vieillissement. Cependant, la production de langage spontané entretient des liens étroits avec d'autres fonctions cognitives telles que l'attention, la mémoire et le raisonnement (Cohen, 1988). Certaines fragilisations cognitives dues au vieillissement cérébral auront donc des répercussions sur le langage. Ainsi avec l'avancée en âge, un déclin du langage est observé chez les personnes saines (Valdois, Joannette, Poissant, Ska, & Dehaut, 1990).

2.1. Au niveau lexical

La richesse lexicale et la diversité du vocabulaire des sujets âgés sont comparables à celles des jeunes sujets. Certaines études ont même observé que ces capacités linguistiques sont meilleures chez les sujets âgés que chez les sujets jeunes (Van der Linden & Hupet, 1994).

En revanche, en production orale, le phénomène du « *mot sur le bout de la langue* » est communément observé (i.e. le sujet est sûr de connaître le mot recherché sans pour autant parvenir à le restituer, Fayol, 1997 ; Gillioz, Zellner, & Keller, 2008). On note également une augmentation du nombre de termes vagues, de paraphrasies sémantiques et du nombre de pauses dans le discours des sujets âgés par rapport aux sujets jeunes (Nef & Hupet, 1992b).

2.2. Au niveau morphosyntaxique

La syntaxe se caractérise par un déclin graduel de la complexité syntaxique. On considère ici qu'elle correspond au nombre de propositions subordonnées par rapport au nombre de mots d'un discours (Nef & Hupet, 1992b). En effet, dans la littérature, il n'existe pas de définition unanime de la complexité syntaxique. Le déclin de la complexité syntaxique serait dû à la diminution, voire la disparition de certaines formes morphosyntaxiques complexes. Selon Nef et Hupet, à l'oral, plus les sujets sont âgés et moins les structures syntaxiques sont diversifiées. L'écrit quant à lui serait moins affecté que l'oral.

2.3. Au niveau textuel

Les structures textuelles d'un texte peuvent être analysées à différentes hauteurs. Ainsi la microstructure relève de l'organisation des phrases entre elles dans une perspective interphrastique tandis que la macrostructure relève de l'organisation des unités textuelles de tailles variées (paragraphe, chapitre, texte). Selon Duong, Ska, Poissant, & Joanne, 2000, la microstructure est généralement préservée tandis que la macrostructure est atteinte. En effet, les personnes âgées ont tendance à s'éloigner du sujet traité ce qui se traduit par une diminution de la cohérence globale. Celle-ci est décrite comme le résultat du maintien approprié de thèmes dans le discours (McNamara, Kintsch, Butler-Songer, & Kintsch, 1996).

Il existerait ainsi un effet du vieillissement sur la production de texte. Pour mettre en avant cet effet, il faut envisager « *l'influence des caractéristiques des sujets à savoir leur âge et leur niveau de scolarité sur la qualité des narrations rapportées* » (Duong et al., 2000, p.139) ainsi que les caractéristiques de la tâche.

3. Facteurs influençant l'évolution du langage

Les fonctions cognitives sont dépendantes de la vie psychique du sujet et de l'organisation fonctionnelle qu'il a mise en place au cours de sa vie : capacités mnésiques, connaissances encyclopédiques, capacités de manipulation, etc. Ainsi, chaque sujet développera des capacités différentes selon ses activités sociales, professionnelles et le support anatomique dont il dispose. D'importantes différences interindividuelles seront donc présentes dans le vieillissement cognitif normal.

D'autre part, parallèlement au vieillissement cognitif, des phénomènes de compensation s'installent pour faire face aux exigences de l'environnement. Les réseaux neuronaux sont réorganisés afin d'assurer une efficacité cognitive optimale face au déclin des capacités cognitives. Le phénomène de plasticité cérébrale continue donc à s'exprimer dans le vieillissement normal : les modifications des performances langagières chez la personne âgée pourront ainsi être dues à la fragilisation de certains processus cognitifs, mais également à la mise en place de phénomènes de compensation (Le Rouzo, 2008).

4. Prise en compte du niveau socioculturel dans l'étude de langage

Ces phénomènes de compensation pourraient se mettre en place notamment par une pratique relativement intensive d'activités professionnelles ou de loisirs au cours de la vie adulte (Cabeza, 2002). Sachant que les personnes ayant un haut niveau socioculturel (NSC) bénéficieraient d'un environnement social et professionnel plus stimulant, nous pouvons considérer qu'une personne de haut NSC mettra plus efficacement en place ces phénomènes. Ceci permettra un meilleur maintien des performances verbales. Ainsi l'appartenance à un NSC joue sur l'évolution des performances langagières au cours du vieillissement.

De plus, nous savons que pour objectiver au mieux la dégradation linguistique d'un sujet, il faut considérer son niveau langagier antérieur (e.g. sur le plan clinique, une personne

ayant un haut niveau de langage pourra se retrouver dans la norme d'un test de langage alors qu'une dégradation linguistique pourrait être quantifiable). Or une personne ayant un haut NSC est susceptible d'avoir un haut niveau de langage. En effet, il a été démontré que le NSC influe fortement sur la performance linguistique des sujets (Dujardin & Lemaire, 2008). Berstein, 1975, explique qu'il existerait deux types de « codes linguistiques » : les classes « favorisées » utiliseraient un « code élaboré » tandis que les classes « défavorisées » emploieraient un « code restreint ». On comprend ainsi que pour mesurer le niveau de langage d'un patient, il est préférable de considérer son niveau socioculturel, paramètre sur lequel on pourra également s'appuyer pour estimer son niveau langagier antérieur.

Ainsi, selon Croisile, 2007, les personnes de haut NSC compenseraient davantage leurs difficultés grâce à un meilleur maniement de la langue et une meilleure utilisation de leurs ressources intellectuelles, en cas de maladie elles pourraient faire illusion plus longtemps. La différence entre vieillissement normal et début de vieillissement pathologique est donc difficile à établir. Il s'agira de contrôler au mieux la variable NSC, afin de tenter de déceler le plus précisément possible le début du vieillissement pathologique.

III. Conséquences du vieillissement cognitif pathologique sur le langage : la maladie d'Alzheimer

1. La maladie d'Alzheimer

1.1. Définition

C'est dans la huitième édition du *Handbook of Psychiatry*, dirigée par Kraëpelin (1910), qu'Aloïs Alzheimer décrit le cas de Madame Deter Auguste avant de donner son nom à la maladie dont souffre cette patiente. Internée en 1901, à l'asile de Frankfort, A. Deter décèdera en 1906 à l'âge de 56 ans. L'originalité d'Alzheimer fut de faire abstraction de l'histoire psychiatrique de sa patiente pour insister sur les liens entre modifications cérébrales et pathologie mentale. La découverte *post mortem* des lésions caractéristiques de la maladie confirma les intuitions d'Alzheimer.

Aujourd'hui, la maladie d'Alzheimer (MA) est la principale cause de démence chez les personnes âgées de plus de 75 ans ; elle représente 70% des syndromes démentiels (étude Paquid in expertise collective, 2007, p.363). Avec le vieillissement de la population, sa prévalence est en constante augmentation et sa prise en charge constitue un enjeu majeur de santé publique. Cette maladie se caractérise essentiellement par des déficits cognitifs et elle évolue irrémédiablement vers un syndrome démentiel avec perte d'autonomie du sujet. L'identification et la compréhension de ces déficits sont primordiales pour la détection précoce de la maladie et sa prise en charge.

Les travaux menés dans ce domaine ont mis en évidence que, si la maladie touche l'ensemble des domaines du fonctionnement cognitif, tous les processus au sein de ces domaines ne sont pas systématiquement altérés (Dujardin & Lemaire, 2008).

1.2. Physiopathologie

La MA est une démence neurodégénérative affectant le cortex cérébral. Elle est la conséquence de différents phénomènes physiopathologiques (Sellal & Kruczek, 2007).

Deux types de lésions s'observent dans la MA :

- Des lésions extracellulaires : les plaques séniles, dépôts de protéine bêta-amyloïde toxiques pour les neurones qui entraînent la mort des neurones qui les entourent.
- Des lésions intracellulaires : les dégénérescences neurofibrillaires, dégénérescences des filaments situés dans les neurones. Les protéines Tau qui les constituent deviennent anormales.

Ces lésions vont entraîner essentiellement la mort des neurones de la région hippocampique qui sous-tend la mémoire épisodique ainsi que la mort de ceux des cortex associatifs qui gèrent les fonctions cognitives et comportementales.

1.3. Facteurs de risque

La MA est une pathologie multi-factorielle. Parmi les différents facteurs de risque qui ont été mis en évidence lors des études épidémiologiques, comme l'expertise collective de 2007 (chap. 12), l'âge reste sans conteste un facteur net : les courbes de prévalence et d'incidence montrent une augmentation exponentielle de la fréquence de la maladie avec l'âge.

Le sexe est aussi à prendre en compte. La prévalence de la MA est plus élevée chez la femme, même en tenant compte de la moindre longévité masculine. Pour les femmes, le risque relatif de développer la maladie serait de 1,5 à 2 par rapport aux hommes.

Le niveau d'éducation, la profession ainsi que le type et la quantité d'activités occupationnelles sont déterminants dans le risque de développer une MA. En effet un haut niveau socioculturel ainsi que des activités intellectuelles riches favorisent l'entretien cognitif. Par conséquent, une personne ayant un haut niveau socioculturel aura moins de risques d'être atteint de la MA.

Par ailleurs il existe d'autres facteurs de risque tels que les antécédents familiaux de MA, des facteurs médicaux, des risques cardio-vasculaires et des facteurs nutritionnels.

1.4. Systèmes de mémoires altérés par la maladie d'Alzheimer

La première préoccupation des personnes présentant cette pathologie concerne majoritairement l'efficacité de leur mémoire. La mémoire épisodique est considérée comme la première déficiente. Ce déficit est la conséquence de l'atteinte hippocampique progressive. L'altération de ce système engendre un défaut de mémorisation de l'information nouvelle, du passé récent : le malade connaît des difficultés pour stocker l'information.

On note aussi un dysfonctionnement de la boucle phonologique avec un ralentissement de la répétition subvocale. Cette altération témoigne d'un dysfonctionnement de la mémoire de travail. La mémoire sémantique est également altérée dès le premier stade de la maladie. En revanche, la mémoire procédurale est souvent préservée. Elle correspond aux aptitudes cognitives et motrices difficilement verbalisables (i.e. savoir-faire).

L'évolution de la pathologie est marquée entre autre, par l'aggravation de l'oubli à mesure, l'effritement des souvenirs anciens et la perte progressive de la fonction langagière. Si la MA est le prototype des maladies de la mémoire, une plainte récurrente renvoie à la sensation accrue de « *mot sur le bout de la langue* ». Mais, avec la progression de la pathologie, d'autres troubles du langage s'installent compromettant ainsi l'interaction du sujet avec son environnement.

2. Conséquences de la maladie d'Alzheimer sur le langage

Dans l'évolution de la maladie, des troubles du langage s'ajoutent assez rapidement aux troubles de la mémoire épisodique : le langage pouvant être considéré comme un épiphénomène de la mémoire. Ainsi, les troubles du langage peuvent parfois constituer l'un des symptômes les plus précoces de la maladie (Sellal & Kruczek, 2007). Ils ne se limitent pas aux manifestations de surface mais résultent de déficits sous-jacents. Concernant la production orale, trois stades d'évolution des troubles langagiers ont été proposés.

Au stade 1, on peut constater dans le discours spontané un manque du mot (Cardebat et al., 1995) plus ou moins facilement contourné dans la conversation par des circonlocutions, des termes vagues ou génériques. Les aspects syntaxiques et phonologiques du langage demeurent pour leur part relativement bien préservés (Cardebat et al., 1991).

Au stade 2, les symptômes présents au stade précédent s'aggravent : on observe un manque du mot plus manifeste ainsi qu'une augmentation du nombre de paraphasies sémantiques. On retrouve de nombreuses persévérations qui peuvent être imputées aux troubles de l'inhibition. On remarquera aussi une hypospontanéité qui s'aggravera avec l'avancée de la maladie, les phrases sont plus courtes, peu de mots sont employés. Le discours devient incohérent (Glosser et al., 1998). De plus, la compréhension orale devient difficile, notamment en ce qui concerne les inférences.

Au stade 3, le trouble devient global : la production orale est totalement désintégrée (Sellal & Kruczek, 2007), le débit est réduit, et le discours est parasité de paraphasies sémantiques, phonémiques et de persévérations. Le discours continue à s'appauvrir qualitativement et quantitativement, le contenu informatif se réduit de plus en plus jusqu'à la vacuité (Mentis et al., 1995 ; Barkat-Defradas et al., 2008). La compréhension orale est fortement altérée.

Le langage écrit, quant à lui, est marqué par des troubles précoces de l'orthographe, notamment sur les mots irréguliers. De telles erreurs doivent évidemment toujours être interprétées en fonction des compétences orthographiques antérieures du sujet, mais ces erreurs sont souvent jugées surprenantes ou grossières par l'entourage. À un stade plus avancé, l'écriture fait apparaître une perturbation du graphisme et de l'agencement

spatial : alternance de lettres minuscules et capitales, trouble de l'agencement des lettres dans un mot, erreurs dans les jambages, mauvaise répartition spatiale des lettres. Puis, au stade le plus évolué, la conjonction de troubles langagiers, praxiques et visuo-spatiaux rend l'écriture progressivement illisible (Sellal & Kruczek, 2007).

IV. La densité des idées

1. Définition

Le concept de densité des idées (DI) est issu des travaux de Kintsh et Keenan (1973) et Kintsh (1974). Il a ensuite été repris dans le cadre de la grande étude longitudinale sur les nonnes (Snowdon et al., 1996). La DI rend compte de l'efficacité individuelle en expression et correspond donc à la qualité informative des propositions langagières. C'est une mesure de la quantité d'informations qui est transmise dans un énoncé par rapport au nombre total de mots. Elle se situe donc à l'interface entre la linguistique et la psychologie.

La DI est évaluée grâce à une mesure du ratio entre le nombre d'unités significatives (nombres d'idées) et le nombre total de mots. Un score élevé reflète une bonne qualité informative du discours : l'information est véhiculée à partir d'un nombre réduit de mots (i.e. économie d'expression). Un score faible, quant à lui, reflète une expression vague et répétitive : un plus grand nombre de mots sera utilisé pour exprimer les idées essentielles.

2. Paramètres pouvant influencer le score de densité des idées

La mesure de la DI part du postulat que le langage est le résultat d'opérations cognitives complexes. Il est reconnu que la DI est un indicateur pertinent des fonctions intellectuelles des sujets. Ainsi un haut score en DI serait corrélé à un haut niveau d'éducation combiné à des activités de loisirs riches et stimulantes intellectuellement (Kemper et al., 2001).

Les recherches s'opposent quant à l'effet du vieillissement normal sur la DI. Certaines ont prouvé qu'il existe un déclin de la DI avec l'avancée en âge (Kemper et al., 2001) tandis que d'autres attestent qu'elle serait stable (Chand & Bonnici, 2007).

Selon Riley et al. (2005) la MA conduit à un déclin rapide de la DI. Il a été démontré que la DI est fortement corrélée avec l'avancement de la MA au niveau du langage écrit (Chafe & Tannen, 1987, Mitzner et al. 2003) et du discours oral (Chand & Bonnici, 2007). Selon Snowdon, Geiner et Markesbery (2000), une DI faible au cours de la vie adulte serait associée à un risque plus important de développer une démence et donc une MA. En effet, on trouve plus de lésions neuropathologiques caractéristiques de la MA chez des personnes ayant une faible DI (Riley, Snowdon, Desrosiers, & Markesbery, 2005). En outre, Snowdon et al. (1999), avancent l'hypothèse qu'une DI faible durant l'adolescence signifierait un développement cognitif et neurologique non optimal et donc un facteur de sensibilité accrue au déclin cognitif lié à la MA.

Enfin, plusieurs auteurs ont noté une asymétrie homme / femme face à la langue. Ainsi, Labov, 1998, a observé que les femmes sont plus sensibles que les hommes au modèle de prestige : elles utiliseraient moins de formes linguistiques stigmatisées qu'elles considèreraient comme fautives en discours dirigé. Toutefois peu d'études ont recherché spécifiquement l'effet du sexe sur le langage et parmi celles l'ayant recherché la majorité n'en a pas trouvé (Bryden, 1979). On peut cependant se demander s'il existe une différence de score de DI entre les hommes et les femmes.

3. Evaluation de la densité des idées dans le discours

3.1. Le discours

Selon Duong et al. (2000) depuis une quinzaine d'années, l'étude du discours et notamment des aspects pragmatiques a surgi dans de nombreux domaines : psychologie, neuropsychologie et orthophonie. La psychologie cognitive, n'ayant accès qu'aux productions des sujets pour étudier les mécanismes qui sous-tendent la pensée humaine, considère qu'elles portent en leur sein les marques des mécanismes qui les ont engendrées. Le langage est à la fois le support et le produit de la pensée. Autrement dit, les structures langagières reflètent les structures cognitives et l'analyse du discours revient à modéliser les phénomènes cognitifs sous-jacents.

La production discursive est une situation de communication qui met en jeu des processus linguistiques (i.e. représentation des idées par le langage) mais également des processus cognitifs (i.e. organisation des idées). Ainsi, l'étude de la production du discours permet de comprendre l'interaction entre les processus linguistiques et cognitifs dans le vieillissement normal, mais également de mieux cerner les profils psychologiques en offrant des données normalisées quant aux profils normaux. En ce sens, les productions discursives sont adaptées pour la mesure de la DI.

3.2. Les mécanismes impliqués dans la production du discours

La plupart des théories psycholinguistiques suggèrent que la production du discours oral passe par trois étapes principales de traitement. En 1989, Levelt propose ce qu'il appelle « le modèle général de production de la parole », modèle décrit ci-dessous.

La première étape (conceptualisation) est celle de la préparation conceptuelle du message préverbal. Ce message contient un ou plusieurs concepts auxquels correspondent des mots qui doivent être récupérés dans le lexique mental.

La seconde étape est celle de la formulation ou lexicalisation du message préverbal. Cette étape d'accès au lexique est elle-même subdivisée en deux sous étapes : l'étape d'encodage grammatical (ou de sélection lexicale) et l'étape d'encodage phonologique. L'étape d'encodage grammatical consiste à récupérer les informations sémantiques et syntaxiques relatives aux mots que l'on désire exprimer. Il est ensuite nécessaire de récupérer les informations phonologiques et morphologiques correspondant à ces mots lors de l'étape d'encodage phonologique.

Lors de la troisième étape ces intentions latentes doivent être converties en unités linguistiques grâce à l'articulation de phonèmes.

En temps réel, ces étapes de production ne sont pas successives mais se chevauchent. La planification du message oral va donc se traduire par la présence de pauses, hésitations et ruptures de la fluidité de la parole.

Le plus souvent, lorsque nous parlons, le processus de planification ne se limite pas aux unités de « bas niveaux » : les phonèmes et les mots (Holmes, 1984). En effet, l'activité de production orale est caractérisée - en aval - par la production d'unités de langage plus larges constituant le discours.

Pour produire un discours, le locuteur a donc à sa disposition un ensemble d'idées, de pensées ou encore d'images mentales. Elles vont devoir être converties en un énoncé qui se déroulera de manière linéaire, en obéissant à des règles syntaxiques d'emplacement relatives à l'ordre des mots (Matlin, 2001). À cela viendra se surimposer la planification des aspects suprasegmentaux de l'énoncé (débit, rythme, prosodie, intonation).

L'activité de production orale sous-tend ainsi des capacités cognitives et motrices d'une grande complexité (Levelt, 1994). Pour générer une phrase, une maîtrise des limites de la mémoire et des ressources attentionnelles est indispensable (Bock, 1995 ; Jou & Harris, 1992). Les fonctions exécutives, quant à elles, permettent de planifier, de débiter puis d'énoncer la phrase.

3.3. Le discours oral spontané

3.3.1. Intérêts du discours oral spontané pour l'étude de la densité des idées

Lorsque la psycholinguistique s'est développée, les aspects de production du langage ont été peu étudiés avec seulement 5% des publications sur ce thème (Levelt, 1994). Ceci est probablement dû aux difficultés rencontrées pour obtenir des données comparables entre elles. En effet, la durée du discours, le maintien du ou des thèmes proposés et le genre discursif sont difficilement contrôlables.

Ainsi, les recherches antérieures sur la DI étaient focalisées sur le langage écrit. Cependant, le discours oral spontané fournit des informations précieuses sur le fonctionnement cognitif : certains phénomènes oraux tels que les énoncés inachevés ou abandonnés, les substitutions, les omissions, les ratages, les corrections, les reformulations, les répétitions, les interjections et les pauses remplies permettent d'observer plus aisément les indices de la planification du discours et les aspects dynamiques de conceptualisation en cours. L'analyse du discours oral apparaît donc comme particulièrement informative pour l'évaluation des performances communicatives des individus. De plus, la situation de production d'un discours oral spontané est plus naturelle et plus écologique que la situation de production écrite (Lee, Barkat-Defradas, Gayraud, 2009). Pour finir avec l'avancée de la maladie d'Alzheimer, le langage écrit se

dégrade rapidement. Aussi, si l'on désire investiguer la densité des idées à un stade avancé, il semble intéressant d'étudier ce critère à partir de discours oraux.

Par conséquent, l'évaluation de la DI à partir de discours oraux, dans le cadre du vieillissement normal et / ou pathologique, constitue un sujet d'étude original et novateur.

3.3.2. Type de discours et production orale

a. Le discours narratif

Le discours narratif se réfère aux tâches dans lesquelles les sujets s'expriment sur leur vie, leurs activités, leurs opinions. En 1992b, Nef et Hupet, expliquent que le langage narratif spontané traduit l'habileté linguistique telle qu'elle se manifeste dans la vie quotidienne. Son analyse permet de dépasser la prise en compte figée d'éléments linguistiques isolés et d'appréhender les structures et le fonctionnement transphrastique d'un acte linguistique se rapprochant d'une situation de communication naturelle. De plus, les contraintes inhérentes à cette tâche étant réduites, elle permet au sujet de ne pas rester focalisé sur son objectif (Schiffrin, 1994).

b. Le discours descriptif

Le langage descriptif, par exemple décrire une image représentant une scène quotidienne, a l'avantage de contrôler les stimuli inducteurs de discours et par conséquent les réponses des sujets. Il facilite l'organisation de la pensée et du discours. Les données recueillies sont relativement homogènes et d'autant plus faciles à analyser. En outre, cette situation fait moins appel aux affects et à la pudeur que des questions concernant la vie personnelle des sujets : le support visuel peut être une aide pour se sentir à l'aise (la tâche est moins complexe) ainsi que pour organiser sa pensée (notion de « support »).

Dans le cas du discours descriptif, l'émetteur et le récepteur ont tous deux le même référent : l'image. Ainsi, l'émetteur peut se permettre d'être moins informatif contrairement au discours narratif, pour lequel il s'agira de renseigner au mieux le récepteur et d'exposer clairement et efficacement le contexte. Ainsi, on peut supposer que le discours narratif sera plus emprunt de nouvelles informations que le discours descriptif. Le discours narratif permettrait alors d'obtenir des scores de DI plus élevés que le discours descriptif.

4. Mesure de la densité des idées

4.1. L'analyse prédicative

La méthode utilisée pour mesurer la DI est l'analyse prédicative ou analyse propositionnelle (désormais AP). L'AP est un outil méthodologique et conceptuel,

développé initialement par Kintsch (1973) et diffusé largement dans le domaine de la psycholinguistique appliquée. Kintsch a développé cette analyse pour mesurer la compréhension de texte supposant que cette activité était contrôlée par les propriétés de la mémoire de travail. L'auteur fait l'hypothèse que le traitement d'un texte s'opère par cycle et que la quantité d'informations traitée par le sujet est variable en fonction de sa mémoire de travail.

4.1.1. Les fondements de l'analyse prédictive

L'AP a été d'abord élaborée pour l'étude du contenu sémantique de textes. En effet, les recherches de Kintsch et Keenan (1973), et Kintsch (1974), affirment que les propositions sont les unités impliquées dans la compréhension et la mémorisation des textes. Inspiré par les travaux de logiciens comme Fillmore (1968), Kintsch propose une méthode pour représenter le sens des textes sous la forme d'une hiérarchie de « propositions sémantiques » (Rouet, 2001).

Ce modèle d'analyse sera ensuite appliqué à l'étude du discours puisqu'il permet d'en décrire le contenu sémantique. La nature de la prédication est de dire quelque chose à propos de quelque chose ; ainsi la prédication revient à ajouter de l'information sémantique.

L'AP repose sur le postulat que la forme dominante de la représentation cognitive du langage est de nature propositionnelle. La proposition est l'unité sémantique de base car elle constitue la plus petite unité sémantique du discours. En effet, un mot isolé seul ne suffit pas à créer une idée, c'est l'ensemble de propriétés et de relations s'y rapportant qui permettent d'appréhender et de produire la signification. Ainsi, une proposition est définie comme un couple formé d'un prédicat (i.e. élément qui porte l'information verbale à propos du sujet) et d'un ou plusieurs arguments (i.e. ce qui est dit à propos du prédicat). En conséquence, le nombre de propositions concorde avec le nombre de prédicats.

Le prédicat est une propriété ou une relation, typiquement un verbe, un adjectif, un adverbe ou une conjonction (Kintsch, 1974 ; Turner & Green, 1977). C'est une unité dite « non saturée » car elle nécessite d'autres éléments pour constituer une idée complète. L'argument peut être, un objet, un être ou d'autres propositions auxquels s'applique la propriété ou la relation. C'est une unité dite « saturée » : elle se suffit à elle-même (Lee, Gambette, Maillé, Thuillier, soumis).

Une proposition peut également devenir argument d'une autre proposition, ce qui permet de représenter la signification de phrases, voire de textes entiers sous la forme de hiérarchies propositionnelles. La position d'une proposition dans la hiérarchie déterminera son importance : les propositions les plus importantes sont celles qui introduisent des arguments nouveaux ; les moins importantes sont celles qui leur sont subordonnées (Turner & Greene, 1977).

Par conséquent, au sein du discours, l'unité significative correspondra à la proposition (i.e. le prédicat et ses arguments). La relation entre les propositions au niveau de la phrase constituera la microstructure. Et enfin, l'organisation globale des propositions entre elles, au niveau du discours formera la macrostructure.

4.1.2. Exemples de modèles d'analyse prédicative

Afin de réaliser une AP, il s'agira de procéder à deux grandes étapes : extraire les prédicats puis assigner les arguments et leurs rôles sémantiques aux prédicats correspondants.

Ainsi dans l'exemple « Pierre se promène », la question qui se pose pour trouver le prédicat est : *qu'est-ce qui est dit de Pierre ?* Réponse : *qu'il se promène*. D'où l'écriture propositionnelle suivante :

P1. SE PROMENER (Pierre)

Selon la théorie logique, on écrit usuellement les prédicats en majuscules et les arguments en minuscules : « Pierre » est l'argument du prédicat « SE PROMENER ». L'ensemble forme une proposition (P1). Le prédicat « SE PROMENER » assigne à l'argument « Pierre » la propriété de sujet. L'ensemble forme une représentation sémantique activée dans le discours.

Il existe différentes méthodes d'AP ; ainsi le choix d'un modèle spécifique conduit à des traitements d'ordres différents (François, 1991).

- Considérons un exemple analysé selon le modèle descriptif de Kintsch et Van Dijk (1978). « *L'excellent livre que j'ai acheté hier sera certainement distingué par un prix* ».

P1.EXCELLENT (livre)

P2.ACHETER (je, livre)

P3.HIER (P2)

P4.DISTINGUER (prix, livre)

P5.CERTAIN (P4)

Comme vu précédemment, une proposition peut être argument d'un prédicat ; c'est ici le cas de P2 et P4.

- Considérons ce même exemple analysé selon le modèle de Le Ny (1989)

P1.EXCELLENT (livre)

P2.ACHETER (agent : je, objet : livre)

P3.HIER (2)

P4.DISTINGUER (instrument : prix, objet : livre)

P5.CERTAIN (4)

Le modèle de Le Ny prend en compte la catégorie grammaticale des arguments. En effet, l'actualisation des propositions dans le discours fait ressortir non seulement les connaissances lexicales mais également la connaissance du monde. Une méthode de description de la signification de l'énoncé peut fournir ainsi un moyen d'étudier le fonctionnement cognitif.

En effet, selon Ghiglione, Kekenbosch et Landré, 1995, p.49 « si l'on considère que la prédication qui s'exprime dans un message linguistique est une activité cognitive essentielle de l'homme et que, sous-jacent à la réalisation de surface, c'est-à-dire au mot, se trouve un concept, on peut estimer que l'analyse prédicative, outil de description sémantique, est pour le psychologue la transcription d'une activité cognitive ».

4.2. Application à la densité des idées

La DI correspond à la richesse d'un discours, c'est-à-dire au nombre de propositions (i.e. au nombre de prédicats) qu'il contient par rapport au nombre de mots (Snowdon et al., 1996 ; Kemper et al., 2001). On détermine le score de DI en utilisant la formule suivante :

$$\text{Score de DI} = \text{nombre de prédicats} / \text{nombre de mots}$$

Ainsi, si l'AP reflète l'activité cognitive, la DI quant à elle permet de la quantifier. En revanche, la DI ne s'intéresse pas à l'organisation dans la proposition.

Par ailleurs, la mesure de la DI a montré son utilité dans plusieurs domaines de la psycholinguistique appliquée, notamment la compréhension de texte (Kintsch et al., 1973 ; Kintsch, 1998), le vieillissement (Kemper et al., 2000), la MA (Snowdon, 1996) et le genre du discours (Covington, 2009).

4.3. Les limites de l'analyse prédicative

4.3.1. Limites de ces modèles pour une application à l'oral

L'AP a surtout été décrite et utilisée pour l'étude d'une production écrite. Ainsi, nous pouvons nous demander si les analyses qui existent sont suffisamment adaptées pour l'analyse d'un discours oral dont les caractéristiques sont très différentes de celles observées pour l'écrit. La transcription de l'oral à l'écrit est un passage nécessaire pour réaliser une AP. Or, le passage à l'écrit occulte bon nombre d'informations pertinentes pour l'accès au sens : le paraverbal (e.g. intonation, débit) et le non-verbal (e.g. posture, regard, gestes).

De plus, la construction du discours écrit est linéaire, contrairement à celle du discours oral. Un discours oral contient des répétitions, des énoncés inachevés ou abandonnés, des autocorrections, des marqueurs de pensée à voix haute, etc. D'autre part, certains éléments du discours oral sont difficilement analysables, tels que les éléments non verbaux (e.g. soupirs ou mimiques signifiantes comme une moue dubitative ou un froncement de sourcil de désaccord, etc.), les interjections ou les passages inaudibles. Ainsi, le passage de l'oral transcrit à l'écrit analysable va poser de nombreux problèmes (Chafe & Tannen, 1987). Afin de rendre le corpus analysable, il s'agira de mettre en place des règles spécifiques. Pour ce faire, un prétraitement des données doit être réalisé.

Ce prétraitement nécessite une interprétation de l'analyste. De ce fait, on risque de s'éloigner de la production originale et de gommer quelque peu la subjectivité des sujets,

tout en introduisant une certaine part de la subjectivité de l'analyste. Par conséquent, il est préférable que le transcripteur réalise lui-même le prétraitement.

4.3.2. Limites liées à l'analyste

Comme il a été souligné précédemment, il existe différentes méthodes d'AP. Ainsi, le choix d'un modèle spécifique conduit à des traitements d'ordres différents. Deux analystes ayant une méthode différente obtiendront deux scores différents pour un même corpus.

Mais, même si le modèle utilisé est identique, il est possible que nous obtenions autant de résultats d'analyses différents que d'évaluateurs. En effet, selon Ghiglione et al., 1995, p.45 « *l'analyse prédicative est une analyse essentiellement sémantique qui requiert une attention soutenue au sens véhiculé, ce qui rend parfois nécessaire le recours à des équivalents sémantique* ». En conséquence, chaque évaluateur fera des inférences sensiblement différentes, et les scores de DI pourront être variables pour un même corpus.

Il en va de même pour une personne non experte en AP. En effet, le français comprend certaines règles particulières : un même mot pourra être analysé différemment selon sa fonction au sein d'une phrase. Il sera donc extrêmement difficile pour quelqu'un qui ne maîtrise pas cette analyse d'obtenir un résultat correct et objectif.

Afin de limiter l'influence de l'analyste, il paraît essentiel de se tourner vers une analyse ne nécessitant pas (ou peu) son intervention. Une solution pourra être d'utiliser un logiciel d'analyse automatique permettant de calculer les scores de DI. De plus, si l'analyse prédicative est considérée à l'heure actuelle comme l'outil le plus complet pour étudier le processus cognitif *via* le langage, son analyse manuelle longue et fastidieuse fait souvent obstacle à sa mise en application qui plus est dans la pratique clinique. Aussi l'utilisation d'une procédure d'analyse automatisée permettra un gain de temps et sera d'autant plus objective.

5. Un exemple d'analyse automatique de l'anglais, Computerized Propositional Idea Density Rater : CPIDR 3.2 (Brown, Snodgrass, Kemper, Herman, & Covington, 2008)

5.1. Présentation du logiciel

CPIDR 3.2 permet de mesurer automatiquement la DI d'un discours oral ou écrit et a été réalisé pour l'analyse de la langue anglaise. CPIDR 3.2 part du postulat selon lequel la DI propositionnelle peut être mesurée en divisant la quantité d'information (i.e. nombre de prédicats) par le nombre total de mots (Snowdon et al., 1996). En effet, le calcul de la DI ne nécessite finalement pas de prendre en compte l'ensemble des prédicats et de leurs arguments, mais seulement de compter les prédicats. Ainsi, le logiciel étiquette chaque mot en tant que prédicat ou non-prédicat pour ensuite calculer le score de DI.

Dans le système de Kintsch, élaboré par Turner et Greene en 1977, une proposition correspond au verbe principal et ses arguments (sujet, objet, objet indirect, etc.). Les éléments descriptifs additionnels comme les adjectifs, les adverbes et autres modificateurs de la phrase sont considérés comme des propositions additionnelles.

Ainsi, d'après l'algorithme de base de CPIDR 3.2, seront considérés comme prédicats :

- les verbes, exceptés les auxiliaires ;
- les adjectifs ;
- les adverbes ;
- les prépositions ;
- les conjonctions ;
- les déterminants.

Par conséquent, pour déterminer si un mot est prédicat ou non, il s'agit de réaliser préalablement un étiquetage morphosyntaxique, étape réalisée dans le cas de CPIDR 3.2, grâce au logiciel MontyLingua (Liu, 2004). Cette étape est suivie d'un traitement à base de règles, destiné à corriger les erreurs d'étiquetage morphosyntaxique qui ont une influence sur le nombre de prédicats. Ces règles de réajustement ont été ajoutées afin d'obtenir, grâce à l'analyse automatique, des scores de DI reflétant plus justement ceux obtenus manuellement par les psycholinguistes (Brown, Snodgrass, Covington, Herman, & Kemper, 2007).

Dans le cas de l'analyse d'un discours oral, une transcription orthographique doit être réalisée. Le mode oral (i.e. speech mode) est sélectionné afin que le logiciel prenne en compte les particularités de l'oral. Ainsi les répétitions, les hésitations et les interjections sont rejetées : elles ne comptent pas comme des propositions.

Considérons l'analyse manuelle d'un même discours par plusieurs analystes. Les scores de DI mesurés seront relativement variables d'un analyste à l'autre, des différences de scores seront donc objectivables. De même, si l'on compare le score de DI obtenu par CPIDR 3.2 avec chaque score obtenu manuellement, on observe des différences. Cependant, la moyenne de ces différences est moins importante que chacune des différences observées entre les scores des étiqueteurs humains entre eux. Cette méthode est donc efficace puisque CPIDR 3.2 obtient généralement un meilleur accord avec l'ensemble des scores des analystes humains que les scores des analystes humains entre eux.

5.2. Limite de ce logiciel pour notre étude

Ce logiciel ne convient que pour l'étude de la langue anglaise. La grammaire française étant différente, les règles d'assignation des prédicats ne correspondent pas aux règles de CPIDR 3.2. Il est donc inutilisable en français. Ainsi, afin de réaliser cette étude, nous avons participé à la création d'un logiciel d'analyse automatique du français. Il sera détaillé par la suite.

Chapitre II
PROBLEMATIQUE ET HYPOTHESES

I. Problématique

L'évolution du critère de densité des idées pourrait constituer un indicateur pertinent d'aide au diagnostic précoce et d'évolution de la maladie d'Alzheimer. En effet, les études ont montré que chaque proposition sémantique ou idée produite requiert un certain nombre d'efforts d'opération cognitive. Ainsi une faible densité des idées dans la production langagière peut indiquer l'altération des capacités cognitives.

L'idée serait de distinguer les sujets ayant une évolution normale de ceux présentant une évolution pathologique et de pouvoir inférer sur l'état d'avancement du déclin. Pour ce faire, il conviendra de connaître le profil de l'évolution de la densité des idées dans le cadre du vieillissement normal, en tenant compte des variables sexe et niveau socioculturel. Cependant, les études antérieures qui se sont penchées sur l'effet du vieillissement cognitif normal sur la densité des idées, parviennent à des résultats contradictoires. De plus, il semblerait que les scores de densité des idées soient corrélés au niveau socioculturel. Enfin, les études actuelles ne font pas référence à l'impact du sexe sur la densité des idées. De cette réflexion émergent les questions suivantes :

Peut-on voir apparaître une diminution de la densité des idées lors du vieillissement normal ? Existe-t-il une corrélation entre la densité des idées et le niveau socioculturel ? Peut-on observer un impact du sexe sur la densité des idées ? Peut-on objectiver une diminution de la densité des idées chez des patients atteints de la maladie d'Alzheimer ?

En outre, selon Nef et Hupet (1992b), l'étude de l'expression orale spontanée s'est jusqu'ici fondée sur l'analyse de corpus, recueillis en demandant au sujet soit de décrire une ou plusieurs images, soit de raconter un évènement particulier de leur vie. Comme nous l'avons vu précédemment, la description d'image a l'avantage d'être une tâche standardisée, mais sa faiblesse est d'être spécifique et limitée. Au contraire, le récit d'un évènement constitue une tâche permettant de recueillir une production verbale plus riche et variée. Cependant, cette tâche est peu standardisée et implique des processus particuliers de récupération et planification susceptibles d'interférer avec les processus de production verbale. Tenant compte de ces divers points, nous nous demandons si la densité des idées d'un même sujet va varier en fonction du type de tâche effectuée.

II. Hypothèses

1. Hypothèses générales

Nos hypothèses générales sont les suivantes :

- Le vieillissement cognitif normal entraîne une diminution de la DI.
- Le sexe n'influence pas la DI.
- Le niveau socioculturel est corrélé à la DI.
- Les patients atteints de la MA ont une DI plus faible que des personnes saines appariées en âge, sexe et niveau socioculturel.
- La DI d'un même sujet sain varie en fonction du type de discours produit.

2. Hypothèses opérationnelles

Nos hypothèses opérationnelles sont les suivantes :

- Les personnes saines de 65 à 74 ans obtiennent des scores de DI plus élevés que les personnes saines ayant entre 75 et 85 ans, dans une tâche de langage oral spontané.
- Dans la population saine, les femmes obtiennent le même score moyen de DI que les hommes ayant le même âge et le même niveau socioculturel, dans une tâche de langage oral spontané.
- Les personnes saines ayant un bas niveau socioculturel obtiennent des scores de DI inférieurs à celles ayant un niveau socioculturel élevé, dans une tâche de langage oral spontané.
- Un groupe de 11 patients atteints de la MA obtient des scores de DI plus faibles qu'un groupe de 11 personnes saines appariées en âge, sexe et niveau socioculturel, dans une tâche de langage oral spontané.
- Un discours oral de type narratif est susceptible de faire émerger un score de DI plus élevé qu'un discours oral de type descriptif chez un même sujet sain.

Chapitre III
PARTIE EXPERIMENTALE

I. Population

1. Sujets sains

1.1. Méthode de recrutement

Des associations ou clubs regroupant un nombre important de personnes âgées ont été sélectionnés. Dans un premier temps, nous leur avons envoyé une lettre d'information afin de leur expliquer notre projet de recherche puis nous les avons contactés par téléphone.

L'ensemble de la population âgée saine regroupe:

- des membres de clubs du troisième âge ;
- des membres de clubs sportifs : club de yoga et club de piscine ;
- des adhérents de différentes associations : des associations de conscrits, l'association d'accueil des nouveaux arrivants d'Ecully, l'association « La ligue contre le cancer », l'association « France Alzheimer » ;
- des proches d'étudiantes en orthophonie ;
- des proches de notre entourage amical et familial.

Les rencontres se sont déroulées de juillet à octobre 2009.

Nos travaux ont fait l'objet d'une demande d'autorisation auprès du Comité de protection des personnes (CPP) de Montpellier par notre maître de mémoire. Après consultation du comité d'éthique, il n'a pas été nécessaire de déposer un dossier auprès du CPP, notre étude étant considérée comme "purement observationnelle".

1.2. Répartition de la population

1.2.1. Critères d'inclusion

La population générale est constituée de 40 sujets sains volontaires de langue maternelle française. Afin de conclure à l'existence de différences de performances, il s'agira de contrôler l'influence possible de différentes variables (Van Der Linden & Hupet, 1994). Nous avons retenu 3 critères.

- Le niveau socioculturel.

Il a été évalué par la grille de Poitrenaud (Kalafat, Hugonot-Diener, & Poitrenaud, 2005). Cette grille permet d'attribuer au niveau socioculturel un score compris entre 1 et 4. Ce score dépend du niveau d'étude (cf. annexe 1, p.77) :

- niveau 1 : pas de diplôme ou CAP adulte ;
- niveau 2 : certificat d'études primaires (CEP), ou CAP + CEP, ou jusqu'à la fin d'une classe de 4ème ;
- niveau 3 : études jusqu'à fin 3ème avec ou sans le brevet des collèges, ou fin terminale (sans le baccalauréat) ;
- niveau 4 : baccalauréat et plus.

Nous avons sélectionné notre population selon deux niveaux socioculturels. Le premier correspond à un niveau inférieur ou égal au certificat d'étude, soit un score de 1 ou 2. Le second est supérieur à 2 ans après le baccalauréat, soit un score de 4. Sachant que l'écart entre les deux niveaux socioculturels est important, on fait l'hypothèse que le contraste sera d'autant plus marqué. Notons aussi que les activités occupationnelles et professionnelles pouvant avoir une incidence sur le langage ont été prises en compte pour établir le niveau socioculturel (e.g. une personne, devenue chef d'entreprise qui ne possède que le certificat d'étude ne sera pas sélectionnée au sein de notre population).

- L'âge.

Deux classes d'âge sont retenues : 65-74 ans et 75-85 ans. Les sujets ont passé le Mini Mental State Examination, MMSE (Folstein & McHugh, 1975) dans le but d'écarter tout éventuel déficit mnésique. Tous les volontaires sains retenus ont un MMS égal à 30/30.

- Le sexe.

20 hommes et 20 femmes sont répartis de façon équivalente dans chaque classe d'âge et pour chaque niveau-socioculturel.

Les informations concernant les sujets sont récapitulées dans le Tableau 1.

	65- 74 ans (n=20)		75-85 ans (n=20)	
	Hommes	Femmes	Hommes	Femmes
Niveau socioculturel élevé (score de 4)	N = 5	N = 5	N = 5	N = 5
Niveau socioculturel bas (score de 1 ou 2)	N = 5	N = 5	N = 5	N = 5

Tableau 1: Répartition de la population en fonction de l'âge, du sexe et du niveau socioculturel

1.2.2. Critères d'exclusion

Trois critères d'exclusion sont retenus :

- Personnes domiciliées en maison de retraite ou foyer logement ;
- Score au MMS inférieur à 30/30 ;
- Personnes encore en activité professionnelle.

2. Sujets atteints de la maladie d'Alzheimer

Nous avons sélectionné 11 patients diagnostiqués comme étant atteints de la maladie d'Alzheimer (MA). Ils sont appariés en âge et niveau socioculturel à 11 sujets contrôles provenant de la population saine.

Les informations concernant les sujets sont récapitulées dans le Tableau 2.

	Patients MA	Sujets sains
Nombre	n = 11	n = 11
Âge moyen	78,2 (\pm 7.96)	78.18 (\pm 5.15)
Sexe	8F + 3H	7F + 4H
Niveau socioculturel	Haut : 4 Bas : 7 NSC bas moyen (selon les critères de Poitrenaud) = 2.27 (\pm 1.35)	Haut : 4 Bas : 7 NSC bas moyen (selon les critères de Poitrenaud) = 2.27 (\pm 1.35)
Score MMSE	MMS= 22 Stades légers (n=9) Stades modérés (n=2) 18 < MMS < 26	MMS= 30

Tableau 2: Composition des populations de sujets sains et des patients atteints de la maladie d'Alzheimer

II. Protocole expérimental

1. Objectif

L'objectif est de construire une tâche permettant d'étudier la densité des idées (DI) dans le discours oral. Sachant que les caractéristiques de la tâche influent sur la production, nous nous sommes penchées sur l'élaboration de l'outil le plus pertinent pour capturer cet indice.

2. Elaboration du protocole expérimental

2.1. Situations de communication sélectionnées

La communication correspond à tout moyen verbal ou non verbal utilisé par un individu pour échanger des idées, des connaissances, des sentiments avec un autre individu (Brin, Courrier, Lederlé, & Masy, 2004). En pratique trois types de communication peuvent être différenciés : la communication gestuelle, la communication verbale et la communication non-verbale. Dans le cadre de notre étude, nous nous sommes intéressées à la communication verbale, étant donné que nous étudions des discours oraux. La situation

de communication peut être définie, quant à elle, par l'objet de l'énoncé qui correspond aux différents types de discours (Dubois, Giacomo, Guespin, Marcellesi, Marcellesi, & Mével, 1994).

Concernant notre étude, deux situations de communication ont été envisagées : le discours narratif et le discours descriptif. La tâche que nous voulons créer doit être suffisamment complexe et nécessiter la mise en jeu de ressources attentionnelles et cognitives suffisantes pour que des différences se manifestent avec l'âge et le niveau socioculturel. Nous avons fait l'hypothèse que le discours narratif permettrait peut-être de mieux explorer la DI et serait susceptible d'analyses plus fines (Lee, Barkat-Defradas, & Gayraud, 2009). En effet, une description d'image va être moins souple qu'un discours semi-dirigé puisqu'elle facilite l'organisation de la pensée et du discours et limite les digressions. Le matériel verbal recueilli lors d'une description d'image serait donc moins riche.

Nous choisissons donc de provoquer un discours narratif semi-dirigé faisant appel à la mémoire autobiographique : il s'agira de poser une question au sujet sur un événement de sa vie. Afin que les corpus soient comparables, il faudra éviter de conduire une interview fondée sur des questions / réponses mais proposer plutôt un protocole limitant au maximum l'intervention de l'expérimentateur. De plus, pour mesurer l'effet de la tâche (i.e. narration vs. description) sur les productions discursives des sujets, nous avons décidé de réaliser également une description d'image

En conséquence, les deux types de discours seront analysés ensemble pour calculer les scores de DI puis séparément pour comparer les deux types de tâche.

2.2. Choix des thèmes de production

2.2.1. Choix des questions

Au départ nous avons pensé poser la question « *Quel est le plus beau jour de votre vie ?* ». En effet, toujours dans l'idée de faciliter la tâche aux personnes atteintes de la MA, il sera plus aisé, pour un patient au stade léger, de raconter un événement très ancien, souvenir plus facilement accessible : dans la MA, les événements anciens sont relativement bien préservés en début de maladie (Dujardin et Lemaire, 2008). De plus, cette question ne consiste pas à rapporter un événement routinier, le sujet devra faire preuve de réflexion quant à l'organisation de son discours. Enfin, cette question est fortement chargée émotionnellement. Plus la charge affective est importante plus elle laissera des traces, nous considérons à priori qu'elle nous permettra d'obtenir une quantité de discours suffisante pour mener à bien notre analyse comparative. Nous avons donc réalisé un pré-test à partir de cette seule question sur un nombre restreint de personnes de notre entourage. Mais il s'est avéré que les corpus recueillis étaient beaucoup trop courts.

Nous avons donc décidé de poser deux autres questions afin que le corpus soit de taille suffisante pour l'analyse. Ces trois questions ont été choisies avec le groupe de travail « Langage et Maladie d'Alzheimer ».

- *quel est le plus beau jour de votre vie ?*
- *racontez-moi une dispute ou une bagarre qui vous est arrivée et qui vous a marqué.*
- *quel est votre plus mauvais souvenir ?*

2.2.2. La description d'image

Nous avons choisi la description d'une image plutôt que celle d'une série d'images car selon Duong et al. (2000), le support d'une seule image force les sujets à explorer plus en profondeur les informations contenues dans le stimulus. Inversement, la présentation d'une série d'images est plus contraignante et réduit la probabilité que les sujets aillent au-delà des informations apparentes qui différencient deux images successives. Le discours qui en résulte est donc plus déclaratif et moins complexe au niveau sémantique. Le choix de l'image s'est fait avec l'ensemble des personnes du groupe « Langage et Maladie d'Alzheimer » : il s'agit de l'image du voleur de biscuit tirée de Boston Diagnostic Aphasia Examination (Goodlass & Kaplan, 1983).

3. Protocole de passation

L'ordre de passation du protocole est le suivant :

- 1- Pouvez-vous me raconter le plus beau jour de votre vie ?
- 2- Pouvez-vous me raconter une bagarre ou une dispute qui vous est arrivée et qui vous a marqué ?
- 3- Pouvez-vous me raconter votre plus mauvais souvenir ?
- 4- Pouvez-vous me décrire cette image ?
- 5- Remplir la fiche de renseignements (cf. annexe 2, p.77) : questionnaire renseignant sur l'âge ; le niveau d'éducation ; la (les) profession(s) exercée(s) ; la date d'arrêt d'exercice et enfin les activités pouvant avoir une incidence sur les performances langagières (lecture, mots croisés, vie associative, politique, etc.) ;
- 6- Passation du MMSE (Folstein & McHugh, 1975) ;
- 7- Lecture de la note d'information et signature de l'autorisation d'exploitation de données personnelles.

Les productions semi-dirigées obtenues ont été enregistrées numériquement (à 22 KHz, 16 bits, mono) à l'aide d'un enregistreur numérique (iPhone) posé devant le sujet.

4. Matériel

Matériel nécessaire à une expérimentation :

- protocole expérimental ;
- image du voleur de biscuit (Goodlass & Kaplan, 1983) ;
- fiche de renseignements ;
- MMSE (Folstein & McHugh, 1975) ;
- Appareil enregistreur ;
- Autorisation d'exploitation de données personnelles.

III. Conditions d'expérimentation

1. Passation

1.1. Sujets sains

En premier lieu, nous expliquons que notre mémoire est l'aboutissement de nos années d'études en orthophonie. Son objectif est d'étudier le langage oral et plus précisément la DI chez des personnes saines de 65 à 85 ans. Nous leur notifions aussi que nous allons comparer la DI entre notre population saine et quelques personnes atteintes de la MA.

Notre but est de provoquer une production de discours à partir de trois questions sur des évènements de leur vie et d'une description d'image. Afin de ne pas empiéter sur la vie privée de la personne, nous lui indiquons qu'elle n'est pas dans l'obligation de répondre à toutes les questions si elles lui semblent trop intimes. Il est aussi d'emblée précisé que l'épreuve n'est en aucun cas un test visant à évaluer leurs performances intellectuelles mais qu'elle consiste à étudier le langage de « tous les jours ». Nous leur expliquons que nous aimerions qu'ils nous racontent en détails certains évènements de leur vie. Enfin nous leur précisons qu'ils seront enregistrés mais que nous respecterons formellement leur anonymat.

Nous avons fait en sorte que la rencontre se fasse dans un contexte le plus naturel et écologique possible. En effet, même si notre but était d'obtenir des monologues et donc d'intervenir le moins possible, nous nous sommes permis de répondre lorsque le besoin de relancer la conversation se faisait sentir (i.e. fonction phatique) afin d'obtenir un discours argumenté de taille suffisante pour l'analyse.

1.2. Sujets atteints de la maladie d'Alzheimer

Les entretiens avec les personnes atteintes de la MA ont été réalisés par les deux étudiantes en orthophonie de Montpellier du groupe « Langage et Maladie d'Alzheimer ». Elles réalisent également leur mémoire de fin d'études sous la direction de Mme Barkat-Defradas. Le protocole et les conditions de passation sont les mêmes que pour les sujets

sains. Cependant la question « Pouvez-vous me raconter une bagarre ou une dispute qui vous est arrivée et qui vous a marqué ? » ne leur a pas été posée.

2. Lieu d'expérimentation

Les sujets sains, appartenant à la population générale, ont été rencontrés à leur domicile ou au sein de l'établissement de leur association. Ils sont domiciliés dans le Rhône, l'Ain, le Jura, les Pyrénées-Atlantiques et le Puy de Dôme.

Les entretiens avec les sujets atteints de la MA ont été réalisés à Montpellier, au sein de la maison de retraite médicalisée « Les Jardins de Sofia ».

3. Durée de la passation

La rencontre dure en moyenne une heure dont 5 à 20 minutes d'enregistrement.

IV. Analyse des données

Comme il a été vu précédemment, il est intéressant de calculer la DI à partir d'un logiciel automatique. Nous avons donc choisi de participer à la réalisation et au développement d'un programme informatique permettant le calcul automatique de la DI en français.

Ce programme est en libre accès et téléchargeable sur le site :

<http://code.google.com/p/densidees>.

Il est agrémenté d'un manuel d'utilisation répertoriant les sites nécessaires à son utilisation et les règles utilisées pour le calcul de la DI.

1. Création du logiciel d'analyse automatique de la densité des idées : Densidées

Ce logiciel a été élaboré par Hye Ran Lee, doctorante en sciences du langage à l'Université de Montpellier et Philippe Gambette, doctorant en informatique. Sa création a débuté en novembre 2009 et nous y avons collaboré (Lee, Gambette, Maillé, Thuillier, soumis, 2010, cf. annexe 3, p.78).

Il s'inspire du logiciel CPIDR qui part du postulat que pour calculer la DI il suffit de comptabiliser les prédicats. En effet, le nombre de prédicats rend compte du nombre de propositions. Il est donc inutile de recenser les arguments.

Comme nous l'avons décrit précédemment, un prédicat correspond à un verbe, à un adverbe, à un adjectif, à une préposition, à une conjonction, etc. Ainsi, il s'agira de réaliser préalablement l'étiquetage morphosyntaxique de chaque mot. Cette étape sera réalisée par le logiciel Treetagger. Ensuite, selon l'étiquetage morphosyntaxique de chaque mot, le logiciel Densidées, déterminera pour chacun des mots s'il est prédicat ou

non prédicat. Pour finir, il calculera le score de DI en divisant le nombre de prédicats recensés dans l'extrait étudié par le nombre total de mots.

1.1. Extraction des règles d'identification des prédicats

Dans un premier temps, nous avons participé à l'extraction des règles nécessaires à l'identification des prédicats (les verbes sont prédicats, les adverbes sont prédicats, etc.). Par la suite, des règles d'ajustement ont été ajoutées afin de corriger les erreurs d'étiquetage morphosyntaxique qui ont une influence sur le nombre de prédicats. 35 règles d'ajustement ont été proposées, elles sont intégralement décrites dans le manuel d'utilisation de Densidées (version 1.2).

Voici quelques exemples de règles d'ajustement mises en place dans le logiciel Densidées.

<p>Règle 208 – Comparatif : « que » n'est pas prédicat après « autant », « moins », « pire », « plus » Règle 301 – verbes de liaisons (« apparaître », « être », « sembler », « devenir », « paraître », « rester », « demeurer ») non prédicats si suivis d'un adjectif ou d'un adverbe.</p>

Figure 1 : Exemples de règles de Densidées 1.2

1.2. Création de deux règles particulières à l'analyse de l'oral

Ce logiciel présente un mode oral et un mode écrit qu'il s'agira de sélectionner en fonction des données à analyser. Le mode oral permettra de prendre en compte les caractéristiques spécifiques du langage oral qui ne sont pas traitables d'un point de vue automatique.

En effet, la construction orale est non linéaire et les phénomènes oraux particuliers doivent être pris en compte. Les mots répétés successivement (e.g. « le le le plus beau jour de ma vie ») sont inhérents à l'oral : le logiciel les rejettera donc, ils ne seront pas comptabilisés ni en terme de prédicats ni en nombre de mots. En revanche, la règle n'est pas généralisable aux idées répétées au sein du discours, qui doivent être comptabilisées en nombre de mots mais pas en nombre de prédicats puisqu'elles n'apportent pas d'idées nouvelles. Il en sera de même pour les éléments du discours oral ne pouvant pas être considérés comme porteurs de sens ou comme nouvelles idées (e.g. interjections). C'est donc l'expérimentateur qui devra juger de ces phénomènes afin de permettre au logiciel de les traiter correctement.

Pour ce faire un prétraitement manuel a été déterminé. Pour prendre en compte ce prétraitement, deux règles supplémentaires ont été ajoutées :

- les mots mis entre parenthèses ne sont pas comptabilisés comme des prédicats mais ils sont comptés dans le nombre total de mots.
- les mots encadrés de crochets ne sont comptés ni en tant que mots ni en tant que prédicats.

Voici les règles qui doivent être appliquées lors du prétraitement manuel.

- Doivent être mis entre parenthèses :
 - Les éléments non porteurs de sens ou d'idées nouvelles :
 - phrases inachevées, ratages ;
 - idées répétées au sein du discours.
 - Les mots ne pouvant être correctement étiquetés pour éviter qu'ils ne soient identifiés en tant que prédicats :
 - noms propres ;
 - mots polysémiques utilisés comme interjections (e.g. « bon », etc.).
- Doivent être mis entre-crochets les éléments qui ne sont en aucun cas la conséquence de l'élaboration du discours :
 - les mots fragmentés (e.g. la /gu/ pour la guerre) ;
 - les pauses remplies non lexicales (e.g. « pff », « ouf », « mm », etc) ;
 - les passages inaudibles.

1.3. Fiabilité et validité du logiciel

Afin de proposer des améliorations au logiciel, nous avons relevé les erreurs commises par celui-ci lors de l'analyse de nos données. Pour cela, nous avons vérifié manuellement l'étiquetage de nos corpus et l'assignation des prédicats. Le recensement des erreurs a permis de créer de nouvelles règles pour y remédier. Par la suite, chacune des nouvelles règles ajoutées ou des modifications apportées ont subi un test de fiabilité. Ainsi de nouvelles versions plus performantes ont été créées ; la dernière en date étant Densidées 1.2.

Pour vérifier la pertinence du dernier logiciel, une corrélation entre l'analyse manuelle et l'analyse automatique réalisée avec Densidées 1.2 a été effectuée sur les quarante corpus des sujets sains : on note un coefficient de corrélation de 0,972, là où CPIDR obtenait pour l'anglais 0,942. Pour une évaluation plus fine, le taux de faux négatifs (i.e. mots non comptés comme prédicats alors qu'ils le sont) et le taux de faux positifs (i.e. non-prédicats étiquetés comme prédicats) ont été déterminés : respectivement 2,7% et 3,1%. Cependant, comme la formule de calcul de la DI fait simplement intervenir le nombre total de propositions, ces deux types d'erreurs se compensent. Ainsi, le taux d'erreurs moyen obtenu est d'environ 0,5% sur le nombre de prédicats.

Finalement, le logiciel Densidées 1.2 fournit actuellement des scores de DI tout à fait satisfaisants pour l'étude de l'oral. Ce logiciel est d'autant plus intéressant qu'il élimine le biais de l'analyste et qu'il réduit considérablement le temps d'analyse. En effet, des tentatives préalables d'étiquetage manuel ont fait apparaître des différences d'étiquetage des prédicats de 5 à 10% entre les analystes. Par conséquent, bien que Densidées ne permette pas une analyse qualitative fine du discours, il donne un résultat rapide et objectif de la DI.

2. Procédure d'analyse de nos corpus

Cette procédure a été répétée pour chacun des corpus analysés.

➤ Etape 1. Transcription

La transcription est réalisée à partir du logiciel Transcriber. Elle consiste en une transcription orthographique standard et non phonétique (cf. exemples en annexe 4, p.88). Afin que les corpus soient comparables statistiquement entre eux nous les avons tronqués à 300 mots.

➤ Etape 2. Prétraitement manuel

Il s'agit de marquer certaines caractéristiques spécifiques à l'oral, lesquelles sont décrites précédemment. Elles sont en effet inexploitable d'un point de vue automatique. Voici un exemple de phrase prétraitée manuellement :

[Pfff] (bon) j'étais ingénieur chez (Merlin-Gérin) (on faisait) on vendait du gros outillage électrique à à à la SNCF

Justifications de ce prétraitement :

- « Pfff » : pause remplie non lexicale (0 prédicat, 0 mot).
- « bon » : Interjection (0 prédicat, 1 mot). Il ne doit pas être étiqueté comme adjectif et donc identifié en tant que prédicat.
- « Merlin-Gérin » : Nom propre (0 prédicat, 1 mot).
- « on faisait » : phrase inachevée (0 prédicat, 2 mots).
- « à à à » : cette répétition successive est automatiquement écrasée par le logiciel donc « à » ne sera pris en compte qu'une fois.

➤ Etape 3. Etiquetage morphosyntaxique

L'étiquetage du texte est réalisé à partir du logiciel TreeTagger. L'interface web de ce logiciel est disponible à l'adresse : <http://cental.fltr.ucl.ac.be/treetagger/>

Le texte transcrit est copié dans le logiciel qui effectue par la suite l'étiquetage morphosyntaxique de chaque mot : il en donne la nature grammaticale.

➤ Etape 4. Calcul de la DI par le programme Densidées

Pour commencer, il s'agit de sélectionner le mode oral. Puis, il suffit de coller le texte étiqueté par TreeTagger dans le cadre de gauche et cliquer sur « calculer ».

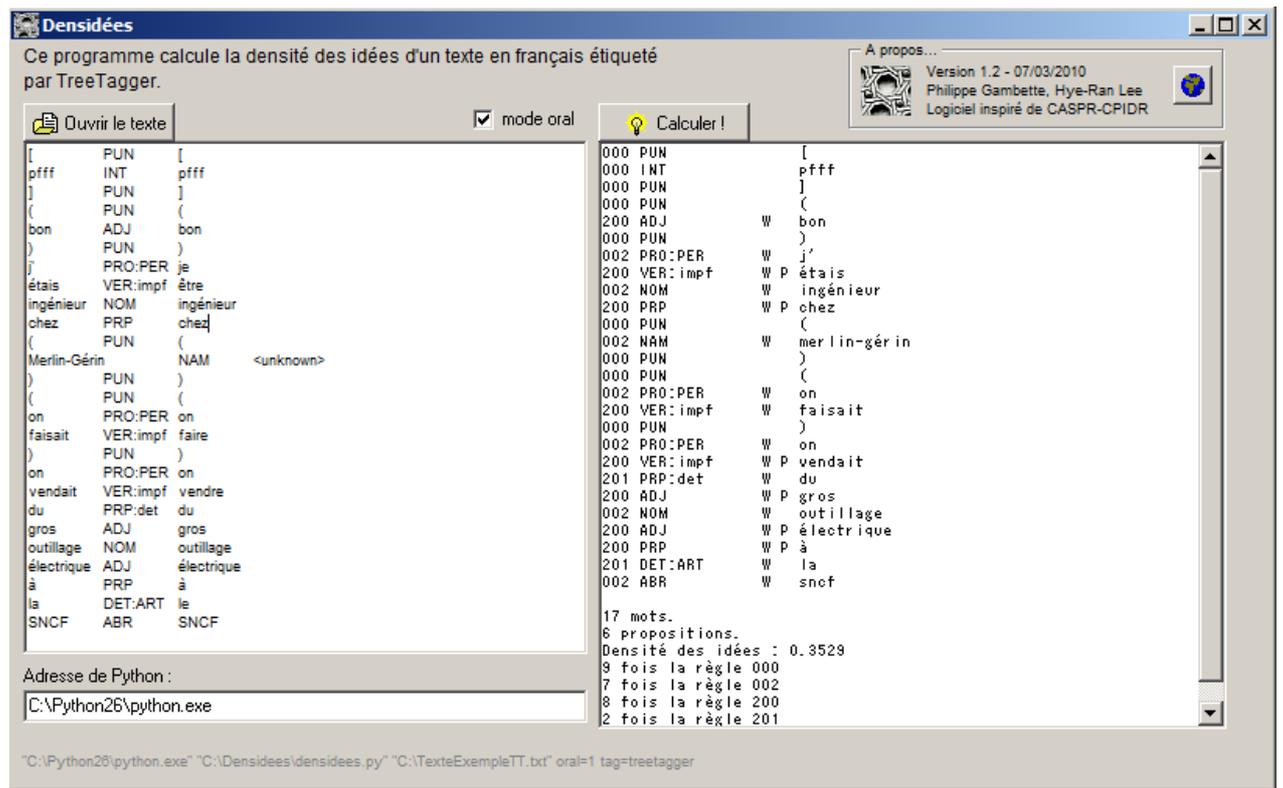


Figure 2: L'interface graphique de Densidées sous windows

Le cadre de gauche présente le texte étiqueté par Treetager.

Le cadre de droite affiche le résultat de l'analyse par Densidées : W signifie que le mot analysé rentre dans le nombre total de mots et P indique un prédicat. En bas du cadre figurent :

- le nombre total de mots ;
- le nombre total de prédicats (i.e. propositions) ;
- le score de DI ;
- les règles qui ont été utilisées pour identifier les prédicats, et le nombre de fois où elles ont été employées.

Ainsi, nous avons pu attribuer un score de DI à chaque corpus étudié.

Chapitre IV
PRESENTATION DES RESULTATS

I. Etude de la population générale

Les scores de densité des idées (DI) obtenus par les sujets appartenant à la population saine sont présentés dans l'annexe 5, p.94.

1. Méthodologie statistique

L'ensemble des données a été traité à l'aide du logiciel SPSS 12.0 (Illinois, Inc). Un résultat sera considéré comme statistiquement significatif si $p < 0,05$.

Les tests utilisés sont des tests non paramétriques. Etant donné l'échantillon restreint de participants dans chaque groupe d'âge, de sexe et de niveau socioculturel (N=20), il a été jugé préférable d'utiliser la moyenne des rangs plutôt que la moyenne des résultats. Cette moyenne est moins affectée par les valeurs aberrantes. Ainsi, on classe l'ensemble des individus par ordre croissant de valeur de DI et on attribue un rang à chacun des individus. Le rang numéro 40 est attribué au score le plus haut ; et le rang numéro 1 au score le plus faible. Puis, pour chaque variable, est effectuée la moyenne des rangs obtenus par les participants du groupe.

C'est à partir de ces données que sera réalisée l'analyse statistique à l'aide du test de Mann-Withney. Un effet de l'âge, du sexe ou du niveau socioculturel (NSC) sur la DI pourra être ainsi mis en exergue.

2. Effet de l'âge

Le but est ici d'observer si la DI varie de façon significative lors du vieillissement cognitif normal. Nous comparons donc les scores moyens de DI obtenus chez les personnes saines ayant entre 65 et 74 ans et chez celles ayant entre 75 et 85 ans.

Dans un premier temps, il est souhaitable d'avoir des indications sur les moyennes des scores obtenus par chaque groupe d'âge (cf. Tableau 3).

	Age	N	Moyenne	Ecart-type	Erreur standard moyenne
DI	65-74 ans	20	0,41120	0,042830	0,009577
	75-85 ans	20	0,41460	0,026757	0,005983

Tableau 3 : Statistiques de groupe : moyenne des scores de densité des idées obtenus pour chaque groupe d'âge

Les moyennes de score de DI obtenues par les deux groupes sont relativement proches. Ainsi, le groupe comprenant les personnes les plus jeunes obtient une DI moyenne de 0,4112 tandis que celui comportant les personnes les plus âgées présente une DI moyenne de 0,4146 (cf. Figure 3).

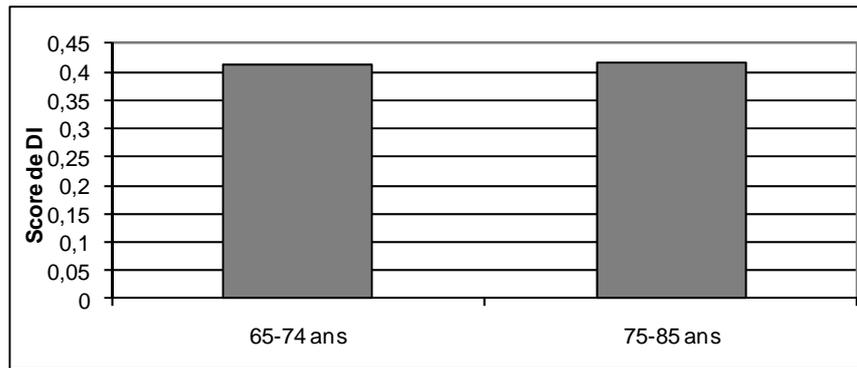


Figure 3 : Score moyen de densité des idées en fonction de l'âge

Les résultats obtenus pour les deux groupes sont ensuite comparés grâce au test de Mann-Whitney. Il s'agit d'un test non paramétrique de comparaison de moyennes pour des échantillons indépendants, il se base sur les rangs. Le rang moyen pour chacun des 2 groupes d'âge est calculé (cf. Tableau 4).

	Age	N	Rang moyen	Somme des rangs
DI	65-74 ans	20	19,78	395,50
	75-85 ans	20	21,23	424,50
	Total	40		

Tableau 4: Rang moyen de chaque tranche d'âge

Dans un second temps, ces 2 rangs moyens sont comparés. Le test informe donc sur la différence de position des individus des 2 groupes par rapport aux autres. Il s'agit ensuite de déterminer si cette position est statistiquement différente (cf. Tableau 5).

	DI
U de Mann-Whitney	185,500
W de Wilcoxon	395,500
Z	-0,392
Signification asymptotique (bilatérale)	0,695
Signification exacte [2*(signification unilatérale)]	0,698 ^a

a. Non corrigé pour les ex aequo.

b. Critère de regroupement : age

Tableau 5 : Résultats obtenus au test^b pour les deux groupes d'âge

La différence observée entre les scores obtenus par les personnes ayant entre 65 et 74 ans et celles ayant entre 75 et 85 ans n'est pas statistiquement significative ($p=0,698$).

La DI n'est donc pas affectée par le facteur âge lors du vieillissement cognitif normal.

3. Effet du sexe

Il s'agit de vérifier si les hommes et les femmes ont une DI équivalente pour une même tranche d'âge et un même niveau socioculturel ou bien si l'on peut observer une différence statistiquement significative entre les deux sexes.

Les moyennes des scores obtenus par chaque groupe sont tout d'abord calculées (cf. Tableau 6).

	Sexe	N	Moyenne	Ecart-type	Erreur standard moyenne
DI	Féminin	20	0,41565	0,030222	0,006758
	Masculin	20	0,41010	0,040343	0,009021

Tableau 6 : Statistiques de groupe : moyenne des scores de densité des idées obtenus pour chaque sexe

Les moyennes des scores de DI obtenues par chaque groupe diffèrent peu. Les femmes ont une DI moyenne de 0,4156 tandis que les hommes ont une DI moyenne égale à 0,4101 (cf. Figure 4).



Figure 4 : Score moyen de densité des idées en fonction du sexe

Les scores obtenus par les deux groupes sont ensuite comparés à l'aide du test de Mann-Whitney (cf. Tableau 7).

	Sexe	N	Rang moyen	Somme des rangs
DI	Féminin	20	21,93	438,50
	Masculin	20	19,08	381,50
	Total	40		

Tableau 7 : Rang moyen de chaque sexe

Le rang moyen de chaque groupe se situe aux alentours de 20/40. Ces rangs moyens sont donc très proches. Il s'agit ensuite de les comparer (cf. Tableau 8).

	DI
U de Mann-Whitney	171,500
W de Wilcoxon	381,500
Z	-0,771
Signification asymptotique (bilatérale)	0,441
Signification exacte [2*(signification unilatérale)]	0,445 ^a

a. Non corrigé pour les ex aequo.

b. Critère de regroupement : sexe

Tableau 8 : Résultats obtenus au test^b pour les deux sexes

La différence observée entre les deux groupes n'est pas statistiquement significative ($p=0,445$). Les hommes et les femmes ont donc une DI équivalente pour un même âge et un même NSC.

Il n'y a pas d'impact du sexe sur la DI.

4. Effet du niveau socioculturel

Le but est d'observer si le NSC a un effet significatif sur les scores de DI obtenus par les participants. Les scores de DI des personnes ayant un haut NSC (niveau 4) doivent donc être comparés à ceux des personnes ayant un faible NSC (niveau 1 ou 2).

La DI moyenne de chaque groupe représentant un NSC est tout d'abord calculée (cf. Tableau 9).

	Niveau	N	Moyenne	Ecart-type	Erreur standard moyenne
DI	Bas niveau	20	0,39065	0,027515	0,006153
	Haut niveau	20	0,43510	0,027572	0,006165

Tableau 9 : Statistiques de groupe : moyenne des scores de densité des idées obtenus pour chaque niveau socioculturel

Les 20 personnes ayant un haut NSC obtiennent une moyenne de score en DI plus élevée que celles ayant un bas NSC. Elles obtiennent respectivement une moyenne de DI de 0,4351 et 0,3906 (cf. Figure 5).

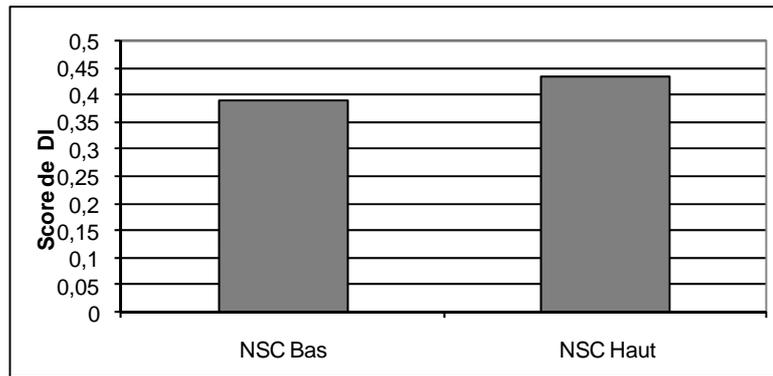


Figure 5 : Score moyen de densité des idées en fonction du niveau socioculturel

Le test de Mann-Withney est alors utilisé pour comparer statistiquement les deux groupes (cf. Tableau 10).

	Niveau	N	Rang moyen	Somme des rangs
DI	Bas niveau	20	12,40	248,00
	Haut niveau	20	28,60	572,00
	Total	40		

Tableau 10 : Rang moyen de chaque niveau socioculturel

Les rangs moyens de chaque groupe sont éloignés. Le groupe haut niveau obtient un rang moyen de 28,6/40 alors que le groupe bas niveau obtient un rang moyen de 12,4/40. Ces deux rangs moyens sont comparés dans le Tableau 11.

	DI
U de Mann-Whitney	38,000
W de Wilcoxon	248,000
Z	-4,384
Signification asymptotique (bilatérale)	0,000
Signification exacte [2*(signification unilatérale)]	0,000 ^a

a. Non corrigé pour les ex aequo.

b. Critère de regroupement : niveau

Tableau 11 : Résultats obtenus au test^b pour les deux niveaux socioculturels

La différence observée entre les scores de DI des deux groupes est très significative ($p < 0,001$). Les personnes ayant un haut NSC ont une DI significativement plus élevée que celles ayant un bas NSC.

Il existe donc bien un impact du NSC sur la DI.

II. Etude de l'échantillon de patients atteints de la maladie d'Alzheimer

Les scores de DI des patients atteints de la maladie d'Alzheimer (MA) et des sujets sains appariés sont récapitulés en annexe 5, p.95.

1. Méthodologie

L'ensemble des données a été traité à l'aide du logiciel SPSS 12.0 (Illinois, Inc). Un résultat sera considéré comme statistiquement significatif si $p < 0,05$.

Les échantillons de population étudiés ici sont appariés entre eux. Le nombre d'individus qu'ils contiennent étant faible, les scores de DI seront analysés à l'aide d'un test non paramétrique basé sur les rangs mais pour des échantillons appariés. Le test utilisé sera le test de Wilcoxon. Ce test utilise l'information sur la direction des différences de scores de DI entre paires d'individus.

2. Effet de la pathologie

Un groupe de 11 patients atteints de la MA a été apparié en âge et en NSC avec un groupe de 11 personnes saines. Ainsi, chaque sujet malade a été apparié à un sujet sain issu de la population générale.

Bien que les échantillons soient restreints, 11 sujets chacun, la DI moyenne de chaque groupe est calculée (cf. Tableau 12).

	N	Moyenne	Ecart-type	Minimum	Maximum
DI (sain)	51	0,40329	0,041250	0,289	0,491
DI (MA)	11	0,36845	0,044123	0,289	0,420

Tableau 12 : Statistiques descriptives: moyenne des scores de densité des idées obtenus pour le groupe « sain » et le groupe de personnes atteintes de la maladie d'Alzheimer

La DI moyenne des sujets atteints de la MA est inférieure à celle des sujets sains. En effet, elle est égale à 0,3685 alors que celle des sujets sains est de 0,4412. Il semblerait donc que la pathologie ait un effet sur les scores de DI.

Afin de le vérifier statistiquement, un test de Wilcoxon est réalisé sur les deux échantillons appariés (cf. Tableau 13).

		N	Rang moyen	Somme des rangs
DI (MA) – DI (sain)	Rangs négatifs	9 ^a	6,61	59,50
	Rangs positifs	2 ^b	3,25	6,50
	Ex aequo	0 ^c		
	Total	11		

a. DI (MA) < DI (sain)

b. DI (MA) > DI (sain)

c. DI (MA) = DI (sain)

Tableau 13 : Rang moyen du groupe « sain » et du groupe de personnes atteintes de la maladie d’Alzheimer

Pour 9 paires d’individus sur 11, la DI du sujet malade est inférieure à celle du sujet sain apparié (cf. Figure 6).

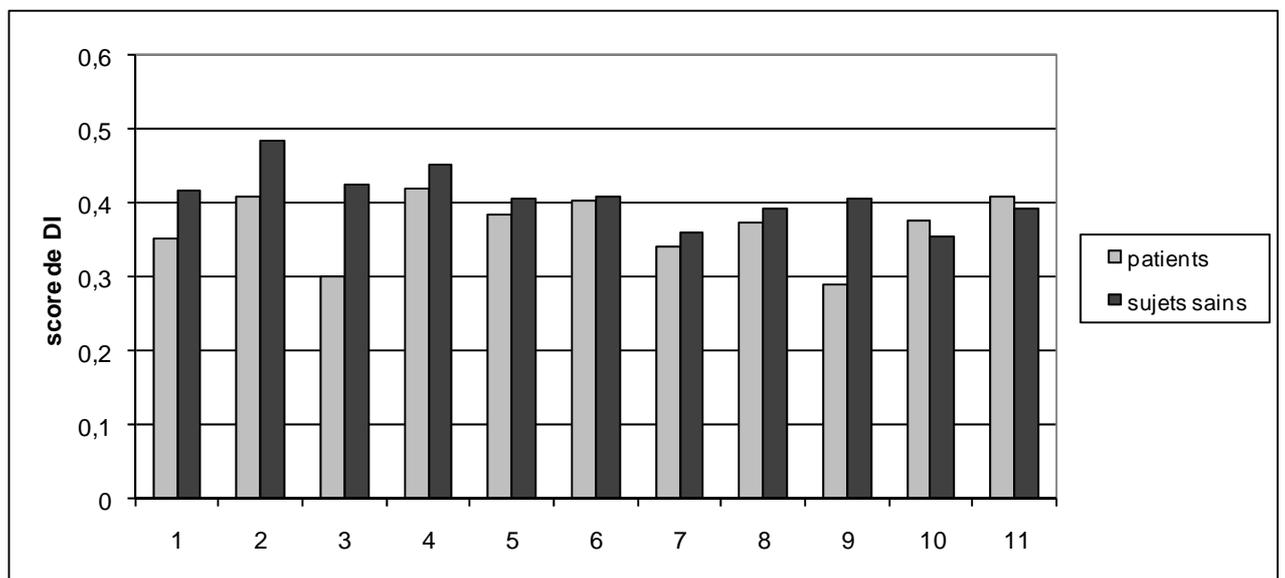


Figure 6 : Scores de densités des idées obtenus par les 11 sujets atteints de la maladie d’Alzheimer et leurs sujets sains appariés

Les scores de DI observés entre le groupe de sujets sains et le groupe de sujets atteints de la MA sont comparés dans le Tableau 14.

	DI MA – DI
Z	-2,357 ^a
Signification asymptotique (bilatérale)	0,018

a. Basée sur les rangs positifs.

b. Test de Wilcoxon

Tableau 14: Résultats obtenus au test^b pour le groupe « sain » et le groupe de personnes atteintes de la maladie d'Alzheimer

La différence entre les scores de DI observée entre le groupe de sujets sains et le groupe de sujets atteints de la MA est significative ($p=0,018$).

La maladie d'Alzheimer a donc un impact sur le score de DI, il diminue dans le cadre de cette pathologie.

III. Etude de l'effet du type de discours sur la densité des idées

Les scores de DI des sujets sélectionnés pour l'étude de chaque type de discours sont présentés en annexe 5, p.96.

1. Méthodologie

La méthodologie utilisée est la même que celle vue précédemment pour l'étude de l'effet de la pathologie. Les scores de DI seront analysés à l'aide d'un test non paramétrique basé sur les rangs mais pour des échantillons appariés : le test de Wilcoxon.

2. Impact du type de discours sur la DI

Les sujets appartenant à la population saine ont réalisé deux tâches de nature différente. Ces tâches étaient destinées à recueillir un corpus afin de mesurer la DI de chaque individu selon le type de tâche. L'une correspond à une production de discours narratif et l'autre à l'élaboration d'un discours descriptif. L'objectif est ici d'observer une tendance : le score moyen de DI sera-t-il influencé par le type de discours (i.e. narratif vs descriptif) pour un même groupe d'individus ?

Pour étudier cela, un échantillon de 8 personnes a été sélectionné. En effet, la taille des discours descriptifs devait être suffisante pour qu'ils soient comparables statistiquement aux discours narratifs. La DI de ces personnes a été calculée une première fois à partir du discours narratif, puis une seconde fois à partir du discours descriptif.

Les moyennes des scores de DI obtenus pour chaque type de discours ont été calculées (cf. Tableau 15).

	N	Moyenne	Ecart-type	Minimum	Maximum
Description	8	0,3650	0,03341	0,31	0,41
Narration	8	0,4069	0,05874	0,29	0,48

Tableau 15 : Statistiques descriptives : moyenne des scores de densité des idées obtenus pour les discours descriptif et narratif

La moyenne des scores de DI obtenue lors de l'analyse du discours descriptif (DI=0,3650) est inférieure à celle obtenue pour le discours narratif (DI=0,4069) (cf. Figure 7).

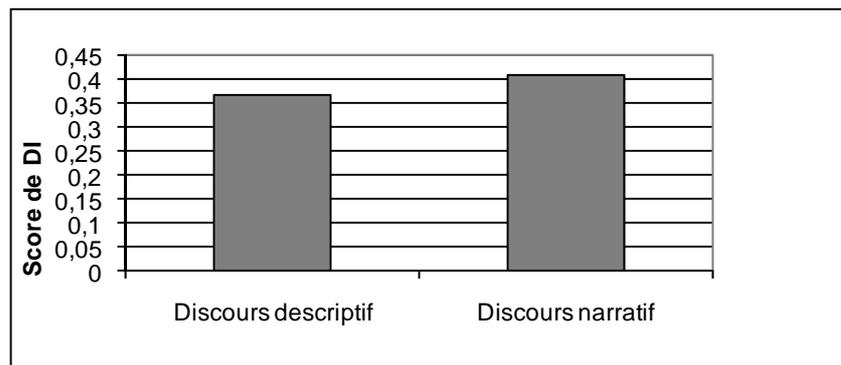


Figure 7 : Score moyen de densité des idées en fonction du type de discours

Le test utilisé pour voir si les scores correspondant aux deux types de discours sont significativement différents est le test de Wilcoxon (cf. Tableau 16).

		N	Rang moyen	Somme des rangs
Narration – description	Rangs négatifs	2 ^a	3,50	7,00
	Rangs positifs	6 ^b	4,83	29,00
	Ex aequo	0 ^c		
	Total	8		

a. narration < description

b. narration > description

c. narration = description

Tableau 16: Rangs négatifs et positifs moyens pour les discours narratif et descriptif

Six individus sur huit ont un rang positif. Le score de DI qu'ils obtiennent lors de l'analyse du discours narratif est supérieur à celui qu'ils obtiennent lors de l'analyse du discours descriptif.

La différence de DI observée entre les deux types de discours est comparée statistiquement dans le Tableau 17.

	narration – description
Z	-1,540 ^a
Signification asymptotique (bilatérale)	0,123

a. Basée sur les rangs négatifs.

b. Test de Wilcoxon

Tableau 17 : Résultats obtenus au test^b pour les discours narratif et descriptif

Il n’y a pas de différence statistiquement significative ($p=0,123$) entre les deux types de discours ; cependant la DI moyenne obtenue par les participants est supérieure pour le discours narratif. L’échantillon utilisé étant vraiment de petit effectif, il serait intéressant d’effectuer l’analyse à partir d’un échantillonnage supérieur.

3. Impact du niveau socioculturel sur les scores de densité des idées obtenus pour chaque type de discours

En revanche, si l’on compare les deux types de discours, il semble que la DI des textes narratifs soit corrélée avec le NSC.

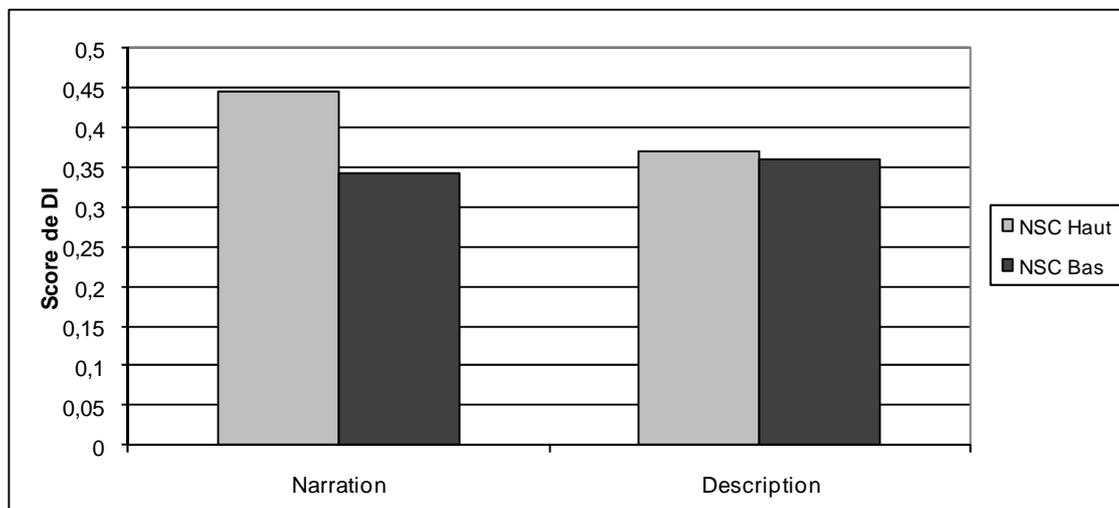


Figure 8 : Score moyen de densité des idées en fonction du type de discours produit et du niveau socioculturel

Ainsi, lorsqu’on analyse le discours narratif, les moyennes obtenues sont respectivement de 0,4462 pour les niveaux 4 et 0,3425 pour les niveaux 1 et 2. Les moyennes des scores obtenues en description par les deux groupes sont, quant à elles, relativement proches : les hauts niveaux obtiennent une DI moyenne de 0,371 et les bas niveaux une DI moyenne de 0,359 (cf. Figure 8).

Il semblerait donc que le discours de type narratif soit plus propice que le discours descriptif pour faire émerger des différences de DI entre des NSC différents.

Toutefois, la taille restreinte des échantillons et la longueur réduite des textes analysés ne permettent pas de conclure catégoriquement sur cette intuition.

IV. Synthèse des résultats

Au vu de ces résultats, nous pouvons conclure que les facteurs âge et sexe n'ont pas d'impact sur la DI. Le NSC, quant à lui, influence très significativement les scores de DI. Une personne ayant un haut NSC obtiendra ainsi des scores de DI plus élevés qu'une personne ayant un faible NSC (cf. Figure 9).

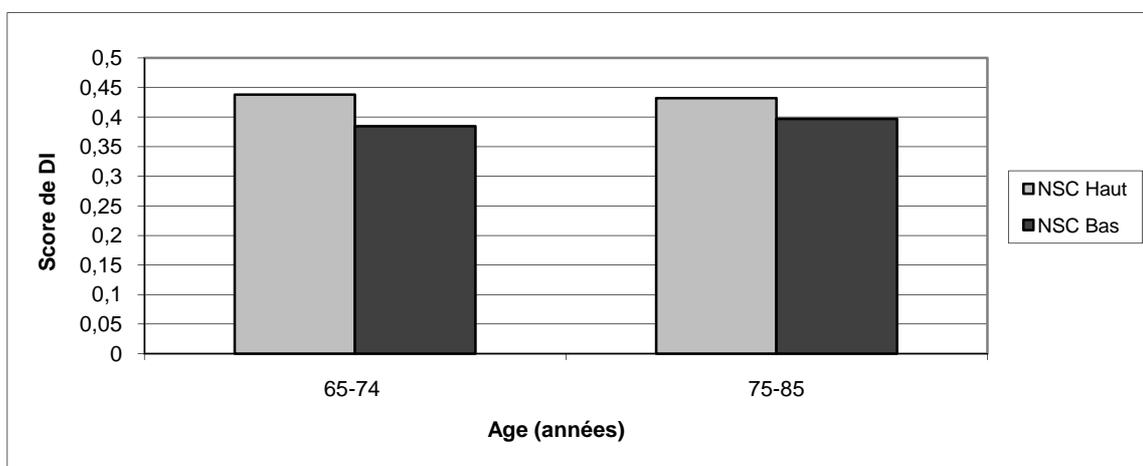


Figure 9 : Score moyen de densité des idées en fonction de l'âge et du niveau socioculturel

De plus, la MA a un impact sur les scores de DI puisque les personnes atteintes de cette maladie ont une DI plus faible que des personnes saines ayant le même âge et le même NSC (cf. Figure 10).

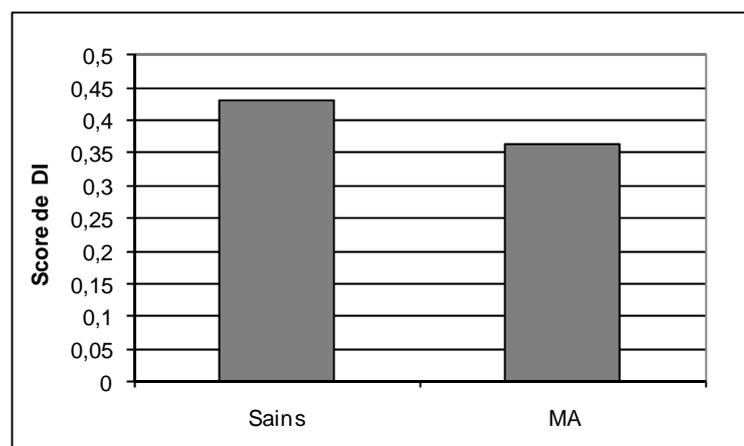


Figure 10 : Score moyen de densité des idées chez les sujets sains et chez les sujets atteints de la maladie d'Alzheimer

Le type de discours, quant à lui, n'influence pas les scores moyens de DI obtenus. Cependant, il serait intéressant de vérifier ce résultat sur un échantillon de taille plus importante car six individus sur huit ont une DI plus élevée lorsqu'ils produisent un discours narratif que lorsqu'ils produisent un discours descriptif.

Notons enfin qu'il semblerait que le discours narratif fasse émerger des scores de DI différents pour deux niveaux socioculturels distincts alors que le discours descriptif mettrait moins en relief ces disparités.

Ainsi, lors de la production d'un discours descriptif, une personne ayant un haut NSC obtiendrait une DI moyenne proche d'une personne de bas NSC. En revanche, lors de la production d'un discours narratif, la DI moyenne serait plus élevée pour la personne de haut NSC que pour celle ayant un bas NSC.

Chapitre V
DISCUSSION DES RESULTATS

I. Validation des hypothèses de travail

1. Effet de l'âge

La densité des idées (DI) reste stable entre 65 et 85 ans dans le cadre du vieillissement cognitif normal. L'hypothèse selon laquelle la DI diminuerait avec l'âge n'est donc pas validée.

Cette hypothèse avait été émise car, bien que le langage soit une fonction cognitive bien préservée lors du vieillissement, la production de langage spontané est liée à d'autres fonctions cognitives telles que l'attention, la mémoire ou encore les fonctions exécutives. En effet, la production d'une phrase nécessite une maîtrise des limites de notre mémoire et de nos ressources attentionnelles. Les fonctions exécutives, quant à elles, permettent de planifier, de débiter puis de produire un énoncé. Ces fonctions cognitives étant fragilisées lors du vieillissement cognitif normal, nous avons supposé que cette altération aurait un impact sur la DI.

Cependant, comme nous l'avons évoqué précédemment, les études se contredisent quant à l'effet du vieillissement sur la DI : nos résultats infirment ceux de Kemper et al., 2001 mais corroborent ceux de Chand et Bonnici (2007) qui attestent que la DI serait stable lors de l'avancée en âge.

Ainsi, Kemper et al. ont observé une diminution de la DI lors du vieillissement cognitif normal. Ils ont alors supposé qu'elle serait influencée par des changements morphosyntaxiques liés à la fragilisation de la mémoire de travail (MDT). En effet, certaines études ont démontré que les capacités de MDT seraient liées à la compétence des sujets à produire une syntaxe complexe et donc des énoncés multi propositionnels (Kemper, 1988). La production de ces structures requiert d'importantes ressources de traitement qui ne seraient pas disponibles chez les personnes âgées en raison de la limitation de leurs capacités de MDT et de leurs difficultés d'inhibition. Cependant, ce type d'explication demande à être encore exploré. Il serait intéressant, par exemple, d'envisager d'étudier la corrélation existant entre les scores de DI, la complexité syntaxique et les scores aux tests d'empans mnésiques (endroit et envers).

En ce qui concerne la description d'image, une étude de Cooper (1990) a montré, qu'il n'y aurait pas de corrélation entre l'âge et la production (i.e. total des mots et durée totale de production), entre l'âge et la complexité syntaxique (i.e. nombre de propositions subordonnées par 100 mots), entre l'âge et la fluidité (i.e. répétitions, énoncés inachevés, pauses remplies et vides, interjections, autocorrections), entre l'âge et la concision, et enfin entre l'âge et l'informativité (i.e. nombre d'informations pertinentes fournies). Ces éléments vont dans le sens des résultats obtenus puisqu'ils tendent à légitimer la stabilité de la DI lors de vieillissement cognitif normal, dans le cadre d'un discours descriptif.

De plus, nous savons que lors du vieillissement cognitif, des phénomènes de compensation se mettent en place pour faire face aux exigences de l'environnement, afin de préserver les performances. Les réseaux neuronaux seraient réorganisés pour garder

une efficacité cognitive optimale face aux déclinés cérébraux. Ces changements entraîneraient une modification qualitative du discours.

Ainsi, les récits des personnes âgées témoignent d'une plus grande élaboration narrative. Ces récits, tout en étant composés de phrases simples, présentent davantage d'épisodes et d'épisodes enchâssés (Nef & Hupet, 1992b). Leur structure narrative serait plus complexe avec des épisodes narratifs multiples interconnectés et l'évaluation des éléments narratifs exposés.

D'autre part, selon Nef et Hupet (1992b), les adultes de 50 et 60 ans produiraient des structures qu'on ne retrouverait pas chez les adultes de 70 et 80 ans. Les sujets les plus âgés n'utiliseraient pas de structures embranchées « à gauche » (i.e. constructions dans lesquelles la proposition subordonnée précède ou interrompt la principale) mais des structures embranchées à droite (i.e. en fin d'énoncé) ; ce qui leur permettrait de maintenir une longueur d'énoncé comparable à celle des adultes plus jeunes. Ainsi, s'il est vrai que lors du vieillissement la complexité syntaxique décline, ce n'est pas forcément le signe d'une diminution du nombre de prédicats employés. Ce déclin s'explique par la diminution de formes morphosyntaxiques complexes, mais ces dernières peuvent être remplacées par d'autres formes. Or, la mesure de la DI ne prend pas en compte ni la complexité ni la diversité des formes morphosyntaxiques ; elle se contente de les comptabiliser. Par exemple, la conjonction de subordination « puisque » sera identifiée comme un prédicat au même titre que la conjonction de subordination « étant donné que ». Pourtant, la deuxième forme est plus complexe. Par conséquent, le discours des sujets âgés serait aussi dense conceptuellement que celui des plus jeunes mais moins complexe morphosyntaxiquement.

Enfin, il a été démontré que chez les personnes âgées, au niveau textuel, la microstructure est généralement préservée tandis que la macrostructure est atteinte. Or, quand on mesure la DI, on ne fait aucune référence à la macrostructure textuelle et donc à la cohérence globale du discours : ce phénomène n'a donc pas d'incidences sur les scores de DI.

Avant de conclure, il est nécessaire de noter que, si l'analyse d'extraits de langage oral a permis d'étudier l'expression spontanée des personnes âgées, les critères d'études sont plus ou moins précis et très variables d'une étude à l'autre. Les résultats sont donc très hétérogènes et il est possible de trouver des discordances chez un même auteur en fonction du protocole expérimental réalisé. De plus, les études ont quasiment toutes été réalisées chez des populations anglophones ; or l'usage de formes grammaticales dites complexes peut évoluer différemment d'une langue à l'autre. Il n'est donc pas aisé de dégager un tableau parfaitement clair et univoque concernant l'évolution de la DI avec l'âge dans la littérature. Cependant, il est probable qu'une analyse syntaxique qualitativement plus fine de nos corpus aurait révélé que certaines constructions syntaxiques sont employées par les deux groupes d'âge, tandis que d'autres sont davantage employées par l'un ou l'autre des groupes.

Il est donc justifiable, au vu des éléments recueillis dans la revue de littérature, que les sujets appartenant à notre population aient une DI stable lors du vieillissement cognitif normal.

2. Effet du sexe

L'hypothèse selon laquelle le sexe n'influence pas la DI est validée. En effet, dans la population saine, les femmes obtiennent le même score moyen de DI que les hommes ayant le même âge et le même niveau socioculturel.

Pourtant, l'effet du sexe sur la latéralisation du langage est une donnée bien admise : il suggère une certaine bilatéralisation des fonctions linguistiques chez la femme tandis que l'homme présenterait une latéralisation du langage sur l'hémisphère gauche (Rondal & Séron, 2000).

Au vu de cet élément, on aurait pu croire en des performances langagières différentes selon le sexe. Pourtant, les données sont loin de montrer un franc effet du sexe. En effet, peu d'études ont recherché spécifiquement l'effet du sexe et parmi celles l'ayant recherché, la majorité n'en a pas trouvé (Bryden, 1979). Ainsi, aucun impact du sexe sur la DI n'est objectivable chez le sujet sain.

3. Effet du niveau socioculturel

D'après les résultats obtenus, nous pouvons confirmer l'hypothèse selon laquelle le niveau socioculturel (NSC) influence le score de DI. Les personnes ayant un haut NSC ont une DI plus élevée que celles ayant un bas NSC. Ce résultat rejoint celui de Kemper et al. (2001) qui expliquent que la DI est corrélée au NSC.

D'un point de vue qualitatif, nous avons constaté au sein des corpus des personnes de bas NSC, une utilisation plus restreinte des adjectifs et des adverbes (i.e. prédicats), que dans ceux des personnes de haut NSC. De plus, leurs phrases sont moins élaborées syntaxiquement et on relève un plus grand nombre de phrases inachevées (i.e. uniquement comptabilisées dans le nombre de mots). Ces observations corroborent les propos de Bernstein (1975) selon lesquels il existerait deux types de « codes linguistiques ». Il caractérise le discours oral des « classes défavorisées » comme « pauvre ». Il explique que les phrases sont plus courtes et plus simples d'un point de vue grammatical et syntaxique (usage simple et répétitif des conjonctions, usage rare des propositions subordonnées et usage rigide et limité des adjectifs et des adverbes). Il ajoute que les personnes de bas NSC auraient tendance à produire plus souvent des phrases inachevées.

Il est important de préciser que selon Kemper et al. (2001) la DI est associée non seulement au NSC mais aussi au vocabulaire et aux connaissances générales ; compétences d'autant plus efficaces que l'environnement est stimulant. En effet, les activités sociales et culturelles permettent de développer des habiletés verbales performantes : « les capacités linguistiques d'un individu dépendent directement de son expérience psychologique et sociale. Comme les individus occupent des positions sociales inégales, ils acquièrent des expériences différentes, de même qu'ils maîtrisent des codes sociolinguistiques divers » (Bachmann, Lindenfeld, & Simonin, 1981, p.93). Ainsi, il ne suffit pas de prendre en compte le niveau d'étude pour l'analyse des habiletés linguistiques, il s'agira aussi de se renseigner sur la vie sociale du sujet : « *la société et la culture ne sont pas présentes avec la langue et à côté de la langue, mais présentes dans la*

langue » (Baylon, 1991, p.31). Ceci explique d'autant plus les hauts scores de DI obtenus par les personnes de haut NSC.

Comme nous l'avons vu précédemment, chaque sujet développera des capacités différentes selon ses activités sociales, professionnelles et le support anatomique dont il dispose. Ainsi, nous pouvons généraliser notre résultat pour supposer qu'une personne ayant une DI efficiente présente de bonnes capacités cognitives, le langage étant le résultat des opérations cognitives. La DI devient non seulement un critère de mesure des habiletés linguistiques mais plus largement un indice de l'état cognitif et neurologique des individus (Kemper et al., 2001).

4. Effet de la pathologie

L'hypothèse selon laquelle les patients atteints de la maladie d'Alzheimer (MA) auraient une DI plus faible que des personnes saines appariées est validée. Ces résultats confirment les travaux de Kemper et al. (2001) ; Riley et al. (2005) ; Chand et Bonnici (2007). La MA est associée à une dégradation linguistique se manifestant également par une diminution de la densité des idées.

La MA est essentiellement caractérisée par des déficits cognitifs. La DI étant une mesure générale du développement cognitif et neurologique (Kemper et al., 2001), on s'attend à ce qu'elle soit affectée par les troubles cognitifs liés à cette maladie. Ainsi, si les atteintes langagières constituent l'un des symptômes les plus précoces de la maladie, ils résultent de déficits sous-jacents. Ces déficits impliquent notamment les fonctions exécutives, la MDT et la mémoire sémantique qui ont des rôles très importants pour l'élaboration du discours oral.

La production d'un discours oral est une opération particulièrement complexe quant au nombre d'informations à traiter et à leur gestion. Elle est contrainte par un système de traitement de l'information limité, généralement associé à la MDT (Bourdin, 1999). En effet, la MDT est à l'origine de la coordination des fonctions de traitement qui permettent de gérer simultanément les étapes du discours et le maintien temporaire des informations. Ainsi, elle permet la manipulation des unités du langage, indispensable à l'élaboration d'un discours cohérent.

Cependant, les capacités de la MDT sont limitées et elle restreindrait le nombre de relations, en particuliers les relations hiérarchiques, qui peuvent être formulées en une fois. Pour ce faire, elle imposerait des limites sur la complexité syntaxique. De plus, elle comporte une composante exécutive (Baddeley, 1986), elle est donc impliquée dans les processus d'attention et d'inhibition des interférences. Ces processus seraient régis par l'administrateur central qui permet également la mise à jour des informations et la flexibilité mentale.

Les capacités de la MDT vont donc influencer la gestion de la production langagière. Or on sait qu'elle est précocement atteinte dans la MA. Ce dysfonctionnement précoce pourrait donc être responsable de la diminution de la DI. Les patients auraient des difficultés à gérer simultanément les étapes du discours et le maintien des informations pertinentes.

Ceci entraînerait une diminution du nombre d'informations traitées parallèlement pour réduire les surcharges de la MDT. Cette difficulté pourrait être à l'origine de la diminution du nombre de propositions émises par rapport au nombre de mots. De plus, les difficultés d'inhibition des interférences et de maintien de l'attention seraient également impliquées dans la dégradation linguistique.

La MA entraîne également une altération des fonctions exécutives. Les fonctions exécutives permettent de planifier l'énoncé, d'inhiber les informations non pertinentes pour le discours mais également de maintenir l'attention tout au long de l'énonciation. Etant donné que les patients ont des difficultés à planifier ce qu'ils veulent dire, et à inhiber ou ignorer les informations non pertinentes, leurs discours vont être parasités de phrases inachevées, d'hésitations et d'éléments digressifs non informatifs qui vont contribuer à faire baisser la DI. De plus, le manque du mot, les difficultés d'attention et la tendance à la persévération idéique associés à la MA, favorisent l'émergence de phrases inachevées.

En conclusion, les corpus des patients atteints de la MA comportent des caractéristiques très représentatives de ces déficits. Chaque fois qu'ils génèrent une phrase, ces locuteurs ont des difficultés à maîtriser les limites de leur mémoire et de leurs ressources attentionnelles. Ceci se traduit qualitativement, au sein des corpus que nous avons recueillis, par une proportion importante d'hésitations, de répétitions et d'énoncés inachevés (Gayraud, Lee, Barkat-Defradas, à paraître). Ces phénomènes, propres à la MA, vont avoir un impact négatif sur la DI.

5. Effet du type de discours

5.1. Impact du type de discours sur la DI

L'hypothèse selon laquelle la DI d'un même sujet varierait en fonction du type de tâche proposé n'est pas validée. Cependant, une tendance est objectivable : les scores de DI obtenus lors du discours narratif sont généralement supérieurs à ceux obtenus lors de la description d'image. Nous pensons donc qu'un échantillon de taille plus grande aurait peut-être permis de valider statistiquement cette tendance.

Afin de comprendre les différences observées selon le type de tâche, nous allons présenter les processus spécifiques mis en œuvre dans ces deux types de discours ainsi que leurs implications dans les variations des scores.

Le langage narratif est défini par le fait de raconter une histoire au cours de laquelle des personnes accomplissent des actions qui s'enchaînent, par des modifications successives de situations. Ainsi, le narrateur se doit de représenter la succession d'actions dans le temps tout en respectant la linéarité du discours oral. Pour respecter la chronologie, il utilise de nombreux repères temporels (e.g. adverbes de temps, prépositions, conjonctions de subordination, etc.) mais aussi de nombreux verbes d'actions ; autant de termes qui seront comptabilisés comme prédicats. De plus, afin d'être exhaustif et informatif, l'émetteur devra renseigner au mieux le récepteur (e.g. où, quand, comment, qui) et par conséquent utiliser bon nombre de repères spatio-temporels (e.g. adverbes de lieu et de

temps), d'adjectifs qualificatifs, d'adjectifs démonstratifs et possessifs. Tous ces termes sont également des prédicats.

En revanche, dans le cas du discours descriptif, l'émetteur et le récepteur ont le même référent : l'image. Ainsi, l'émetteur peut se permettre d'être moins informatif en se contentant de pointer l'image (i.e. deixis), situation fréquemment rencontrée dans nos entretiens. De plus, le discours descriptif donne des renseignements sur les circonstances (e.g. atmosphère, ressenti) et sur les personnages de l'image. On retrouvera alors surtout des verbes d'état (non comptés comme prédicats), des verbes de perception et quelques indices spatiaux. Par ailleurs, notons que l'image choisie est en noir et blanc et comporte peu de détails graphiques. Elle limite donc l'utilisation d'adjectifs qualificatifs, notamment les adjectifs de couleurs.

Ainsi, nous comprenons que dans le cas du discours narratif, on rencontre beaucoup plus d'idées nouvelles (i.e. prédicats) : des précisions devront être apportées pour rendre compte efficacement du contexte (e.g. marqueurs spatiaux, marqueurs temporels, etc.). Le discours descriptif, quant à lui, contiendra préférentiellement des marqueurs spatiaux et se verra moins complet à cause du référent image : le nombre de prédicats étant plus faible, le score de DI est logiquement moins élevé.

Par conséquent, même si nous ne pouvons pas affirmer catégoriquement que le type de tâche influe sur la DI, nous pouvons le supposer. Il faut donc toujours proposer la même tâche aux sujets que l'on entend comparer.

5.2. Effet du niveau socioculturel sur les scores de densité des idées obtenus pour chaque type de discours

Suite aux résultats obtenus, nous avons décidé d'approfondir l'étude du type de discours afin de déterminer si l'un des deux serait plus adapté pour mettre en exergue des différences de performances. Pour ce faire, nous nous sommes penchées sur l'influence du NSC sur la DI selon le type de discours. Nous avons observé que le discours narratif permettrait de mieux mettre en évidence la différence de scores de DI entre les hauts et les bas niveaux.

Ainsi, nous allons tâcher de comprendre en quoi le discours narratif serait plus propice à faire émerger des différences de scores entre les NSC, par rapport au discours descriptif.

Le discours descriptif permet de contrôler la production des individus car il génère le même matériel lexical pour tous. Ainsi, que ce soit dans le discours des hauts ou des bas NSC, on retrouvera sensiblement le même vocabulaire, les mêmes indices spatiaux, les mêmes adjectifs ; et donc approximativement le même nombre de prédicats.

Le discours narratif est, quant à lui, moins contraignant : il ne limite pas l'émetteur à une quantité d'information donnée, n'induit pas un vocabulaire précis, permet les digressions et l'apport de nouvelles informations. Le registre de langue et la manière de s'exprimer, propres à chaque personne, auront donc un rôle prépondérant sur le score de DI. De plus, la production de ce type de discours met plus en jeu la planification : l'émetteur doit gérer le parallélisme entre ordre chronologique et linéarité du discours. On rencontrera ainsi de

nombreux retours en arrière pour préciser une information antérieure ; phénomène qui peut se traduire par des phrases inachevées et des hésitations. Nous avons observé que ces phénomènes étaient plus rares chez les personnes de hauts NSC, on suppose donc qu'elles gèrent mieux ces contraintes.

Le discours narratif est également régi par de nombreuses règles qu'il s'agit de respecter pour parvenir à une communication efficace. En effet, l'émetteur se doit d'être exhaustif afin de donner le maximum d'informations sur le contexte pour être compris par le récepteur (e.g. utilisation d'un vocabulaire précis), mais aussi pour éviter de se répéter. De plus, il devra attirer l'attention du récepteur, évaluer les connaissances de son interlocuteur et respecter, de ce fait, la règle de pertinence (l'énonciation doit être maximale appropriée au contexte). Enfin, les formulations doivent être accessibles à l'émetteur. On peut donc supposer que la production du discours narratif dépendra à la fois de la maîtrise de ces règles mais aussi de l'efficacité des fonctions cognitives, notamment de la capacité à traiter les inférences.

Au vu de ces éléments, nous pouvons supposer que la production d'un discours narratif, est plus complexe que celle d'un discours descriptif. Elle nécessite la mise en jeu de ressources attentionnelles et cognitives plus importantes. En effet, la tâche de discours narratif implique des processus de récupération d'information en mémoire et de planification, auxquels la description d'image est moins sensible. La production d'un discours narratif nécessite un coût cognitif important qui permet la mise en relief de différences de performances.

On peut donc supposer que le discours narratif est plus sensible pour identifier des différences de DI dans le discours oral, notamment dans le cadre de la pathologie.

II. Limites et intérêts de cette étude

1. Limites

1.1. Entretien avec les patients atteints de la maladie d'Alzheimer

Un des biais de notre étude est que nous n'avons pas réalisé les entretiens des personnes atteintes de la maladie d'Alzheimer. Cependant, nous avons écouté intégralement les enregistrements afin de réaliser les transcriptions nous-mêmes. Ainsi, nous nous sommes imprégnées des discours des patients afin de réaliser le plus objectivement possible le prétraitement des données.

1.2. Sélection de la population saine

Nous avons sélectionné les personnes appartenant à la population, en fonction de leur âge, leur sexe, leur NSC, et enfin le score obtenu au MMS. Nous n'avons donc pas pris en compte les profils psychologiques (voire psychiatriques) de ces personnes ; cela pourrait induire un biais dans nos résultats. En effet, la dépression peut entraîner un appauvrissement de la langue ou encore restreindre l'appétence au langage. Marazziti,

Consoli, Picchetti, Carlini et Faravelli (2010), expliquent que les personnes dépressives présentent un ralentissement cognitif associé à des perturbations des fonctions exécutives. En effet, elles connaîtraient des difficultés de planification, de maintien de l'attention et une moindre flexibilité mentale. Ainsi, ces perturbations engendreraient des difficultés de raisonnement, de concentration et de formulation des idées. Comme nous l'avons vu précédemment, toutes ces fonctions ont un rôle prépondérant pour la conceptualisation des idées ; leur affection pourrait donc influencer la DI. Cependant, il est important de préciser que l'ampleur de ces déficits est corrélée avec la sévérité de la dépression. Par conséquent, nous pensons qu'il aurait été pertinent d'ajouter une investigation des antécédents psychiatriques dans notre protocole.

Il aurait également été intéressant de rajouter une tranche d'âge beaucoup plus jeune au sein de notre population. En effet, la littérature expose que la structure du récit narratif varie entre des personnes jeunes et des personnes âgées. Comparer les scores de DI obtenus chez des personnes saines de 25 à 35 ans et chez celles ayant plus de 65 ans aurait pu constituer un apport à cette étude. Cela aurait permis d'observer si des différences se manifestent dans le cadre du vieillissement non pathologique. Cependant, il est difficile d'apparier et contrôler la variable niveau socioculturel entre une population jeune *versus* une population âgée.

1.3. Protocole de passation

Nous avons choisi, avec le groupe « Langage et Maladie d'Alzheimer », de poser la question « *Pourriez-vous me raconter une dispute ou une bagarre qui vous est arrivée et qui vous a marqué ?* ». Au fur et à mesure des entretiens, nous avons constaté que cette question n'était pas adaptée à la population âgée. Une grande majorité des personnes interrogées a refusé d'y répondre. En effet, elles nous répondaient « *je ne me suis jamais bagarré* », « *je n'aime pas les histoires* » ou encore « *je suis contre la violence* ». Cette question les ramenait souvent à des choses difficiles à évoquer, notamment des disputes familiales. Quand bien même nous leur précisions qu'elles pouvaient raconter une dispute banale, par exemple avec un commerçant, elles montraient une certaine réticence et expliquaient qu'elles n'avaient pas de souvenir à ce propos. De plus, la réponse à cette question peut entraîner la crainte d'être jugé. En effet, se disputer ou encore se battre véhicule une image négative dans la société.

D'autre part, la question concernant le plus mauvais souvenir, était destinée à faire évoquer un fait marquant de leur vie. Cette question étant difficile, nous leur en avons demandé un retour. La plupart nous ont répondu que l'évocation était douloureuse, mais que ça ne les avait pas dérangées outre mesure. Toutefois nous pensons à posteriori que demander aux sujets « *pourriez-vous me raconter **un** mauvais souvenir ?* » aurait facilité la tâche.

Pour finir, l'utilisation de l'image du « voleur de biscuit » comme support du discours descriptif, n'était peut-être pas le choix le plus adéquat. En effet, cette image a beaucoup surpris de par sa simplicité graphique. Les personnes interrogées l'ont qualifiée de banale et elle leur a semblé infantilisante. De ce fait, la description était souvent rapide et sommaire, l'image n'attisant pas la curiosité. La longueur des corpus recueillis a donc été très faible. Il aurait pu être intéressant de réaliser un pré-test pour comparer la qualité de la description obtenue à partir de l'image du « voleur de biscuit » avec celle obtenue pour

une image plus complexe graphiquement. Ainsi, nous aurions pu déterminer, au préalable, l'image la plus pertinente pour l'étude du discours.

1.4. Analyse des données

1.4.1. Choix de l'extrait étudié

Nous avons fait le choix de tronquer nos corpus à environ 300 mots pour procéder à l'analyse. L'extrait sélectionné commence par le premier mot du discours. Avec du recul, nous pensons que ce choix n'était pas judicieux, et ce pour deux raisons. Tout d'abord, les réponses des sujets débutent, quasi systématiquement, par une reprise de la question posée. Ainsi, les corpus de toutes les personnes interrogées, qu'elles soient saines ou non, commencent par la même phrase. La structure de la phrase émise ne reflète donc pas forcément leur densité des idées. Il aurait donc été intéressant de rejeter la première phrase pour chaque question posée.

De plus, certaines personnes mettent plus de temps à entrer dans la narration, elles utilisent de nombreuses circonlocutions avant d'être réellement informatives. Cette attitude est d'autant plus exacerbée qu'elles se retrouvent face à quelqu'un qu'elles ne connaissent pas, qu'elles sont enregistrées et que leurs réponses vont être étudiées.

Nous estimons donc, au vu de ces deux idées, qu'il est préférable de sélectionner un passage situé en milieu de discours pour l'analyse.

Par ailleurs, nous avons fait le choix de sélectionner un corpus assez long. Cependant, il aurait été intéressant d'étudier si un passage plus court aurait donné une DI représentative de celle obtenue pour 300 mots et s'il aurait permis de faire émerger les différences observées. Raccourcir le passage analysé réduirait ainsi les temps d'analyse et serait donc moins contraignant pour un clinicien.

1.4.2. Aspect qualitatif

Comme nous l'avons vu précédemment, l'analyse prédicative permet de quantifier la DI. Ainsi, les scores obtenus suite à son application, permettent de déterminer si la DI d'un sujet est conforme à son NSC ou si elle est altérée. Or, nous savons que ce score peut être chuté par une faible utilisation de prédicats par rapport au nombre total de mots et par une utilisation importante de phrases inachevées, de répétitions et d'hésitations.

Ainsi, nous avons constaté, à partir de nos résultats, qu'en fonction du NSC ou encore de la pathologie, ces phénomènes pouvaient avoir une proportion variable. Les sujets de bas NSC semblent faire plus de phrases inachevées que ceux de hauts NSC, ce qui s'accroît d'autant plus avec la pathologie. Cependant, nous n'avons pas quantifié le nombre exact de ces phénomènes pour chacun des corpus, et donc mesuré leurs proportions pour les variables NSC et pathologie. Ceci aurait permis de justifier plus objectivement les variations de scores de DI entre les différents groupes, en comparant la fréquence d'apparition de ces phénomènes.

Nous avons fait le choix de ne procéder qu'à une analyse quantitative de nos corpus. Nous avons estimé que compléter notre travail par une analyse qualitative aurait été trop coûteux en temps. Toutefois, a posteriori, nous considérons que la prise en compte de la dimension qualitative aurait été un apport pertinent pour l'analyse de nos résultats. Etant donné que cette étude pilote a déjà permis d'élaborer un logiciel automatique de quantification de la DI, il sera plus confortable d'y ajouter une analyse qualitative dans de futures études.

1.4.3. Limites inhérentes à l'étude de la densité des idées

La DI évalue la quantité d'informations transmises au sein d'un énoncé. Elle ne rend pas compte de la cohésion (effet de continuité et de progression sémantique produit dans un texte grâce à des indicateurs de surface de nature linguistiques). Elle ne reflète pas non plus la cohérence globale du discours. En outre, cette mesure ne permet pas d'affirmer que l'émetteur a été exhaustif et donc assez informatif pour faciliter la compréhension du récepteur. Ainsi, elle ne permet d'évaluer que partiellement la compétence communicative, elle n'en est qu'un indicateur. Ce critère linguistique ne permet donc pas de conclure catégoriquement quant à l'efficacité de la communication.

2. Intérêts

2.1. Intérêts du protocole expérimental

Les différentes tâches proposées sont relativement écologiques. En effet, elles se présentent sous la forme d'un entretien, les personnes ne se sentant pas en situation de test. Notre but était d'obtenir des discours narratif et descriptif, discours très fréquemment utilisés dans la vie quotidienne des sujets. Par ailleurs, les questions étant intimes, nous craignons que les personnes rencontrées soient réticentes pour y répondre. Cependant, les retours que nous en avons eu décrivent la rencontre comme agréable et non intrusive. Au fil de la discussion, les questions et les réponses s'enchaînaient naturellement et les personnes prenaient du plaisir à partager leur vécu.

De plus, le protocole est simple et pratique à reproduire. Il ne nécessite pas de matériel particulier et sa passation se fait aisément et rapidement.

2.2. Intérêts de l'étude

Malgré l'abondance et la richesse des travaux consacrés à la dimension linguistique de la MA, peu d'entre eux mettent en regard les productions discursives des patients avec les troubles cognitifs. Ainsi, l'originalité de cette étude est de mettre l'accent sur les discours oraux, faciles à recueillir avec la progression de la MA et susceptibles d'analyses linguistiques fines.

De plus, dans le cadre des maladies neurodégénératives, l'étude de critères linguistiques fins est en pleine effervescence. En effet, ces critères offrent des moyens supplémentaires pour la mesure de la dégradation linguistique. Or les aspects linguistiques de la MA ont

été encore peu étudiés en français. C'est pourquoi, notre étude peut être considérée comme originale et novatrice. De plus, nos résultats permettent une comparaison avec les données anglo-saxonnes ; comparaison d'autant plus intéressante que les résultats ne sont pas toujours concordants.

Cette étude a également servi de base de corpus pour la création d'un logiciel automatique de calcul de la DI. Or, ce calcul automatique est une condition nécessaire à sa mise en pratique clinique. Il donne un résultat rapide et objectif de la DI.

III. Apport clinique

Selon l'article 4 du décret de compétence des orthophonistes, la prise en charge (PEC) orthophonique consiste au « *maintien et adaptation des fonctions de communication chez les personnes atteintes de maladies neurodégénératives* ». Cependant, afin de rendre la prise en soins de la communication efficace, l'orthophoniste se doit d'évaluer, préalablement, les troubles fins de la communication au cours du « *bilan des troubles d'origine neurologique* ». L'étude de la DI prend ainsi tout son sens, elle permettrait d'évaluer un des critères de la communication : l'informativité des actes de langage. C'est donc un indicateur supplémentaire pour mesurer l'informativité du discours lors d'un bilan.

Les résultats significatifs obtenus attestent que la DI pourrait être un outil fiable pour mettre en évidence une altération cognitive, notamment chez des personnes ayant un haut NSC. En effet, nous avons démontré que les hauts NSC présentent une DI plus élevée que les bas NSC. Cette mesure pourrait donc contribuer à objectiver précocement l'altération de la DI d'un patient de haut niveau. En effet, il arrive parfois, qu'une personne ayant un haut niveau de langage se retrouve dans la norme d'un test classique du langage alors qu'elle se plaint de difficultés. Selon Croisile, 2007, les personnes de haut NSC pourraient faire illusion plus longtemps en cas de maladie. Elles seraient donc les seules à avoir conscience de leurs difficultés, leur propre entourage ne s'en rendant pas forcément compte. Ainsi, on comprend l'importance de tenir compte du NSC pour réaliser une évaluation des performances langagières. Pour ce faire, l'évaluation devra porter sur une tâche assez pertinente pour permettre de distinguer des différences de performances entre les différents NSC. Le critère de la DI présente ainsi l'avantage d'être un indice linguistique assez fin pour différencier les NSC ; elle permet de prendre en compte aussi bien les difficultés langagières d'un patient de haut NSC que celles d'un patient de bas NSC et de les objectiver par leur quantification. Notons toutefois que même si une diminution significative de la DI peut être un indicateur d'une altération cognitive ; elle ne permet pas à l'heure actuelle d'identifier précisément les fonctions altérées et les mécanismes perturbés.

Par ailleurs, un certain nombre de travaux ont montré l'importance d'une PEC des troubles cognitifs et de la communication par des orthophonistes, à condition que l'approche thérapeutique mise en œuvre soit adaptée à la pathologie, et qu'elle soit mise en place suffisamment tôt. L'aide apportée au patient sera d'autant plus bénéfique que la PEC aura débuté tôt dans l'histoire de la maladie. Ainsi, une PEC précoce permet d'anticiper sur les dégradations cognitives en contribuant à retarder le déclin, actuellement irrémédiable, des capacités cognitives et des capacités de communication des patients atteints de la MA. La DI est un critère altéré avec l'apparition de la MA mais

son repérage nécessite une investigation très fine. Manuellement cette investigation ne peut être réalisée que par un expert en linguistique et elle est très coûteuse en temps. L'utilisation de Densidées pourra dorénavant permettre à un orthophoniste de prendre en compte ce critère lors de l'évaluation du langage d'un patient. En effet, d'une part ce logiciel est gratuit et d'autre part son utilisation ne requiert aucune formation particulière. Il donne un résultat rapide, fiable et objectif de la DI.

Pour finir, les résultats obtenus attestent que la DI est un critère linguistique dégradé dans la MA. Ainsi, on peut supposer qu'elle est un indicateur fiable des capacités de MDT, de l'efficacité des fonctions exécutives - notamment de la planification - et de l'accès au lexique. En outre, comme nous l'avons vu précédemment, elle rend également compte de certaines compétences nécessaires à la communication puisqu'elle est un indicateur de l'informativité du discours. Au vu de ces idées, nous pouvons attester la légitimité de ce critère pour rendre compte à la fois de l'efficacité de certaines fonctions cognitives mais aussi pour détecter un appauvrissement de la valeur informative d'un discours dans le cadre d'une maladie neurodégénérative telle que la MA.

Dans les années à venir, les orthophonistes seront amenés à voir de plus en plus de patients atteints de la MA. On comprend ainsi l'importance d'investiguer plus profondément les phénomènes fins propres à cette maladie pour mener une PEC de qualité. Pour ce faire, l'analyse de la qualité informative du message pourrait être intéressante pour appréhender l'altération des capacités de communication. Par ailleurs, l'évaluation de la DI d'un discours est d'autant plus intéressante qu'elle se veut pragmatique et écologique. Cependant, comme nous allons le voir dans les perspectives, l'utilisation clinique de la DI requière une normalisation sur un plus grand nombre de personnes.

IV. Perspectives et faisabilité

1. Perspectives cliniques

La première perspective envisageable suite à cette étude est une amélioration du logiciel d'analyse automatique de la DI. Son utilisation nécessite une détection humaine des idées répétées (i.e. prétraitement) qui peut être contraignante pour un clinicien. Il pourrait être envisagé d'aborder ce problème de façon automatique afin de réduire ou même supprimer ce prétraitement. Cela faciliterait sa mise en pratique clinique. De même, les mots fragmentés pourraient être rejetés automatiquement par le logiciel.

À l'heure actuelle, le logiciel recense le nombre de fois où une règle a été utilisée pour identifier un mot en tant que prédicat ou non-prédicat. Il serait intéressant de compléter cette approche par règles, avec une partie statistique qui déterminerait la fréquence d'utilisation des mots, selon leur nature grammaticale (i.e. pourcentage de noms, de verbes, de pronoms etc.). Ceci pourrait permettre, à terme, d'identifier si une catégorie est massivement altérée ou non. Ainsi, ces observations pourront être utilisées par l'orthophoniste lorsqu'il tentera de maintenir et renforcer les capacités langagières d'un patient.

Bien que ce ne soit pas l'orthophoniste qui pose le diagnostic dans le cadre de pathologies neurodégénératives, il participe au diagnostic médical. Ce professionnel contribue à quantifier et à qualifier les dégradations linguistiques observées. Les résultats significatifs obtenus dans cette étude indiquent que la quantification de la DI pourrait avoir une valeur d'aide au diagnostic de la maladie d'Alzheimer ; résultats en adéquation avec la littérature. Cette étude est la première réalisée sur la langue française, elle donne donc une référence sur les scores moyens de DI obtenus dans une population saine selon deux NSC. Il pourrait donc être intéressant de la poursuivre : en l'élargissant à une population plus importante avec plus de tranches d'âge et de NSC, on pourrait obtenir une norme de DI. Cette norme servirait de référence pour quantifier la dégradation linguistique d'un sujet et déterminer à partir de quel seuil une DI réduite devient pathologique.

Pour finir, il serait intéressant d'étudier la DI moyenne de populations atteintes de la MA en distinguant les scores obtenus pour les trois stades de la maladie. On pourrait ainsi déterminer si le déclin de la DI est homogène ou non lors de l'avancement de la MA. La DI pourrait peut-être devenir un indicateur contribuant à évaluer l'avancement de la maladie. En effet, on sait que la dégradation des compétences linguistiques est associée à la sévérité de la MA.

2. Autres perspectives

Les personnes de bas NSC auraient plus tendance à développer des maladies de type neurodégénératives par manque de conservation des connaissances et à cause d'une réserve cognitive moindre (Croisile, 2007). Cette vulnérabilité différentielle à la MA pourrait être expliquée par l'existence d'une réserve cérébrale. En effet, cette réserve se construirait au cours d'une longue période de plasticité, permettant l'influence de stimulations éducatives inégales.

Ainsi, au moment de la perte d'un nombre de neurones et de synapses, une plus grande réserve cérébrale permettrait d'atteindre plus lentement le seuil au-dessous duquel la compétence mentale et les stratégies de remplacement ne sont plus possibles, laissant la symptomatologie clinique apparaître. Cependant les bases neurobiologiques du fonctionnement cognitif normal et pathologique sont encore loin d'être élucidées.

De plus, dans les dix dernières années, plusieurs études épidémiologiques, y compris l'étude Paquid (expertise collective, 2007), ont montré une prévalence plus élevée de la MA dans les catégories socioprofessionnelles les plus défavorisées. Cependant, au cours de la vie, il est délicat de distinguer les effets de l'éducation de ceux de l'environnement.

Ainsi, la grande étude longitudinale américaine menée sur les nonnes (Snowdon et al. 1997, 1999, 2000) a permis de conclure qu'une faible DI refléterait un faible NSC, mais également un risque accru de développer la MA. La mesure de la DI pourrait donc contribuer à la détermination des personnes à risque.

CONCLUSION

Notre travail est l'un des premiers à rapporter des résultats à propos de l'évolution de la densité des idées, chez la personne âgée saine et pathologique, à partir de l'analyse de discours oraux en langue française. Les résultats obtenus lors de cette étude permettent de confirmer qu'une faible densité des idées peut révéler une altération des capacités cognitives, notamment dans le cadre de la maladie d'Alzheimer. Par conséquent, l'étude de ce critère linguistique est un champ d'investigation prometteur pour élargir l'éventail des outils de diagnostic de la maladie d'Alzheimer mais aussi pour mesurer l'avancée de la maladie.

Cette étude a également permis de légitimer l'importance de la prise en compte de la variable socioculturelle pour toute étude relative au langage. En effet, les personnes ayant un haut niveau socioculturel ont une densité des idées plus élevée par rapport à celles ayant un bas niveau socioculturel. Ce résultat est en accord avec de nombreux travaux qui révèlent l'importance d'un environnement stimulant pour le développement et le maintien des capacités cognitives, et notamment des capacités linguistiques et communicationnelles. Afin d'objectiver et de quantifier la diminution de la densité des idées associée à la maladie d'Alzheimer, il est donc indispensable de prendre en compte le niveau initial du patient.

La mise au point de ce protocole permet d'élargir le champ de recherche sur la maladie d'Alzheimer, en ouvrant d'autres possibilités de détection précoce de cette maladie. En effet, jusqu'ici, malgré l'importance et la fréquence des troubles du langage dans la maladie d'Alzheimer, les échelles cognitives utilisées n'évaluent finalement que peu d'indices linguistiques fins. Cependant, la détermination et l'analyse d'indicateurs fiables et précoces de nature linguistique pourraient sans doute contribuer à affiner le diagnostic. En effet, la densité des idées n'est pas le seul critère linguistique affecté par la maladie d'Alzheimer : la littérature évoque également la complexité syntaxique ou encore la richesse lexicale. Ces troubles langagiers pourraient être mis en évidence par des protocoles spécifiques de production de discours oral spontané, du type de celui créé lors de ce mémoire. Ces protocoles se veulent pragmatiques et écologiques. Ils sont aujourd'hui considérés comme étant plus adaptés à la dimension fonctionnelle du langage et s'inscrivent ainsi dans la continuité des courants actuels.

Par ailleurs, une exploration des manifestations linguistiques fines contribuera à élaborer un programme de rééducation orthophonique spécifique par l'optimisation des performances préservées et l'utilisation de stratégies compensatoires. Pour ce faire, il est indispensable d'identifier plus précisément les fonctions cognitives qui sont atteintes lorsque l'on constate une altération de la qualité informative du discours. Une telle évaluation permettra d'appréhender au mieux la prise en soins des capacités de communication.

L'originalité de notre travail de recherche réside dans sa dimension pluridisciplinaire. Cette approche nous a fait prendre conscience de la richesse et du caractère nécessaire de la collaboration entre chercheurs et cliniciens. Les connaissances des linguistes sont nécessaires pour soulever les critères linguistiques pertinents, au même titre qu'une approche clinique orthophonique est indispensable à la compréhension de certaines manifestations langagières et communicationnelles de la maladie. De plus, le partenariat avec un informaticien a été indispensable à la création de Densidées.

CONCLUSION

La création d'un programme informatique permettant l'analyse du critère de la densité des idées a constitué une étape fondamentale de ce travail dans la mesure où elle a permis de réfléchir à l'exportation de l'outil Densidées dans le champ de la pratique clinique orthophonique.

Ce mémoire nous a donc sensibilisées à la recherche, apport d'autant plus intéressant qu'elle a toute sa légitimité en orthophonie. En effet, l'étude de la qualité informative du discours, telle qu'elle est capturée par la mesure de la densité des idées, est un premier pas vers l'appréhension des marqueurs plus fins de la communication. Ce travail a également renforcé notre intérêt pour la linguistique appliquée, champ d'investigation qui donne toute son étendue à l'étude du langage et de la communication.

Enfin nous pensons qu'il serait intéressant d'élargir l'étude de la densité des idées à d'autres pathologies. Ceci permettrait d'observer si sa dégradation est spécifique à la maladie d'Alzheimer ou si elle est généralisable à d'autres maladies, et dans quelle mesure.

BIBLIOGRAPHIE

Bachmann, C., Lindenfeld, J., & Simonin, J. (1981). *Langage et communications sociales*. Paris : Hatier.

Baddeley, A. (1986). *Working memory*. Oxford: Clarendon Press.

Barkat-Defradas, M., Martin, S., Rico Duarte, L., & Brouillet, D. (2008). Les troubles de la parole dans la maladie d'Alzheimer, 28èmes JEP, Avignon, 9-13 juin.

Baylon, C. (1991). *Sociolinguistique : société, langue et discours*. Paris : Nathan

Bourdin, B. (1999). Mémoire de travail et production langagière : comparaison de l'oral et de l'écrit chez les adultes et les enfants. *L'année psychologique*, 99, 123-148.

Belin, C., Ergis A.M., & Moreaud, O. (2006). *Actualité sur les démences : aspect clinique et neuropsychologique*. Marseille : Solal.

Bernstein, B. (1975). *Langage et classes sociales : codes sociolinguistiques et contrôle social*. Paris : minuit.

Bherer, L. (2008). L'attention. In K. Dujardin, & P. Lemaire (Eds), *Neuropsychologie du vieillissement normal et pathologique* (pp.29-44). Issy-les-Moulineaux, Elsevier Masson.

Bock, K. (1995). Sentence production : from mind to mouth. In J.L. Miller, & P.D. Eimas (Eds), *Speech, language and communication* (pp.181-216). San Diego, CA, Academic Press.

Bond-Chapman, S., Peterson Highley, A., & Thompson, J.L. (1998). Discourse in fluent aphasia and Alzheimer's disease: linguistic and pragmatic considerations. *Neurologistics*, 11, 1-2, 55-78.

Brin, F., Courrier, C., Lederlé, E., & Masy, V. (2004). *Dictionnaire d'orthophonie*. Isbergues : Ortho Edition.

Brouillet, D., & Syssau, A. (Eds.). (2000). *Le vieillissement cognitif normal. Vers un modèle explicatif du vieillissement*. Bruxelles : De Boeck Université.

Brown, C., Snodgrass, T., Kemper, S., Herman, R., & Covington, M. (2008). Automatic measurement of propositional idea density from part-of-speech tagging. *Behavior research Methods*, 40 (2), 540-545.

Bryden, M. (1979). Evidence for sexe differences in cerebral organization. In M. Witting, A. Peterson (Eds.), *Determinants of sex-related differences in cognitive functioning* (pp121-126). New York : academic.

Buée, L., & Maurage, C-A. (2008). Le vieillissement : des molécules, des cellules et des structures cérébrales en involution. In K. Dujardin, & P. Lemaire (Eds), *Neuropsychologie du vieillissement normal et pathologique* (pp11-126). Issy-les-Moulineaux, Elsevier Masson.

BIBLIOGRAPHIE

- Cabeza, R. (2002). Hemispheric asymmetry reduction in older adults : the Harold model. *Psychology and aging*, 17, 85-100.
- Cardebat, D., Demonet, J.F., Puel, M., Nespoulous, J.L., & Rascol, A. (1991). Langage et démences. In M. Habib, Y. Joanette, & M.Puel (Eds.), *Démences et syndromes démentiels : approche neuropsychologique* (pp.153-164). Paris, Masson.
- Cardebat, D., Aithamon, B., & Puel, M. (1995). Les troubles du langage dans les démences de type Alzheimer. In F. Eustache, & A. Agniel (Eds.), *Neuropsychologie clinique des démences : évaluation et prise en charge* (pp.213-223). Marseille, Solal
- Cardebat, D., & Joanette, Y., (1999). Perturbations discursives en pathologie du langage de la description... à l'interprétation. In X. Seron, & M. Jeannerod, *Neuropsychologie humaine* (pp.408-418). Bruxelles, Mardaga.
- Caza, N., & Moscovitch, M. (2005). Effects of cumulative frequency, but not of frequency trajectory in lexical decision times of older adults and patients with Alzheimer's disease. *Journal of memory and language*, 53, 3, 456-471.
- Chafe, W., & Tannen, D. (1987). The relation between written and spoken language. *Annual Review of Anthropology*, 16, 383-407.
- Chand, V., & Bonnici, L. (2007). Quantifying language degradation in Alzheimer's disease. *New Ways of Analyzing Variation 36* (NWA 36), University of Pennsylvania: Philadelphia, PA. 14 octobre, 2007.
- Cohen, G. (1988). Age differences in memory for texts : Production deficiency or processing limitation?. In L.L. Light, & M. Burke, (Eds.), *Langage, memory and aging*, (pp.171-190). New York, Cambridge University Press.
- Cooper, P.V. (1990). Discourse production and normal aging: Performance on oral picture description tasks, *Journal of Gerontology*, 45, 5, 210-214.
- Covington, M. (2009). Idea Density : a potentially informative characteristic of retrieved documents. Proceedings, IEEE SoutheastCon.
- Croisile, B. (2007). *Alzheimer et les maladies apparentées*. Paris : Larousse.
- Dubois, J., Giacomo, M., Guespin, L., Marcellesi, C., Marcellesi, J.B., & Mével, J.P. (1994). *Dictionnaire de linguistique et des sciences du langage*. Paris : Larousse.
- Dujardin, K., & Lemaire, P. (2008). *Neuropsychologie du vieillissement normal et pathologique*. Issy-les-Moulineaux : Elsevier Masson.
- Duong, A., Ska, B., Poissant, A., & Joanette, Y. (2000). Effet du vieillissement, de la scolarité et du stimulus sur la production de narrations. In D. Brouillet, & A. Syssau, (Eds.), *Le vieillissement cognitif normal. Vers un modèle explicatif du vieillissement* (pp.137-154). Bruxelles, De Boeck Université.

BIBLIOGRAPHIE

- Expertise collective (Eds). (2007). *Maladie d'Alzheimer. Enjeux scientifiques, médicaux et sociétaux*. Paris : Jouve, Inserm.
- Fayol, M. (1997). *Des idées au texte psychologie cognitive de la production verbale, orale et écrite*. Paris : Presses universitaires de France.
- Folstein, M.F., Folstein, S.E., & Mac Hugh, P.R. (1975). Mini Mental State: a practical method for grading the cognitive state of patients for the clinician. *J Psych Res*, 12, 189-198.
- Fontaine, R., & Toffart, L. (2000). Les prédicteurs de réserve cognitive chez la personne âgée. In D. Brouillet, & A. Syssau, (Eds). *Le vieillissement cognitif normal. Vers un modèle explicatif du vieillissement* (pp.274-286). Bruxelles, De Boeck Université.
- François, J. (1991). La pertinence linguistique des représentations propositionnelles de la sémantique cognitive. *Sémiotiques*, 1, 1, 69-80.
- Gayraud, F., Lee, H.R., & Barkat-Defradas, M. (à paraître). Syntactic and lexical context of pauses and hesitations in the discourse of Alzheimer patients and healthy elderly subjects. *Clinical Linguistics and Phonetics*.
- Ghiglione, R., Kehenbosch, C., & Landré, A. (1995). *L'analyse cognitivo-discursive*. Grenoble : Presses Universitaires de Grenoble.
- Gillioz, C & Zellner Keller, B (2008). Analyse du Mot sur le Bout de la Langue de 19 à 79 ans. *28ème Journée d'études sur la Parole*, Avignon.
- Goodlass, H., & Kaplan, E. (1983). *The assessment of aphasia and related disorders*. Philadelphia : Lea and Febiger.
- Jou, J. & Harris, R.J. (1992). The effect of divided attention on speech production. In M.W. Matlin (Eds.), *La cognition, introduction à la psychologie cognitive*. Paris, De Boeck Université.
- Kalafat, M., Hugonot-Diener, L., Poitrenaud, J. (2003). Étalonnage français du MMS version GRECO. *Revue de neuropsychologie*, 13(2): 209-236.
- Kaufman, A. S., Reynolds C.R., & Mc Lean J. E. (1989). Age and WAIS-R Intelligence in a national sample of adults in the 20- to 74-year age range: A cross-sectional analysis with educational level controlled. *Intelligence*, 13, 235-253
- Kemper, S. (1988). Geriatric psycholinguistics: Syntactic limitation of oral and written language. In L. Light & D. Burke (Eds.), *Language, memory, and aging*, (pp.58-76). New York, Cambridge University Press,
- Kemper, S., Greiner, L.H., Marquis, J.G., Prenovost K., & Mitzner, T.L. (2001). Language decline across the life span : findings from the nun study. *Psychology and Aging*, 16, 2, 227-239.

BIBLIOGRAPHIE

Kintsch, W., & Keenan, J.M. (1973). Reading rate and retention as a function of the number of the propositions in the base structure of sentences. *Cognitive Psychology*, 5, 257-274.

Kintsch, W. (1974). *The representation of meaning in memory*. Hillsdale, NJ: Erlbaum.

Kintsch, W., & Van Dijk, T.A. (1975). Comment on se rappelle et on résume des histoires. *Language*, 4, 98-116.

Kintsch, W. & Van Dijk, T. (1978). Toward a Model of Text Comprehension and Production. *Psychological Review*, 85, 363-394.

Labov, W. (1998). Vers une réévaluation de l'insécurité linguistique des femmes. In P. Singy (Ed.). *Les femmes et la langue : L'insécurité linguistique en question* (pp.25-35). Lausanne, Delachaux et Niestlé.

Laine, M., Laakso, M., Vuorinen, E., & Rinne, J. (1998). Cohérence and informativness of discourse in two dementia types. *Neurolinguistics*, 11, 1-2, 79-87.

Lee, H., Barkat-Defradas, M., Gayraud, F. (2009). Le vieillissement normal et pathologique du langage: étude comparative de l'analyse du discours oral. *6èmes journées internationales de Linguistique de Corpus*, Lorient, 10-12 septembre 2009.

Lee, H., Gambette, P., Maillé, E. & Thuillier, C. (soumis). Densidées : calcul automatique de la densité des idées dans un corpus oral. *Rencontre des Étudiants Chercheurs en Informatique pour le Traitement Automatique des Langues*, Québec, du 19 au 22 juillet 2010.

Le Rouzo, M.L. (2008). Le vieillissement psychologique : une vue d'ensemble. In M.L. Le Rouzo, *La personne âgée : Psychologie du vieillissement* (pp.13-63). Mercuès, Amphi psychologie.

Levelt, W.J.M. (1989). *Speaking : From intention to articulation*. Cambridge (MA) : The MIT Press.

Levelt, W.J.M. (1994). The skill of speaking. In P. Bertelson, P. Eelen, & G. Ydewalle, (Eds.), *International perspectives on psychological science* (Vol.1, pp.89-103). Hove, England, Erlbaum.

Liu, H. (2004). MontyLingua : An end-to-end natural language processor with common sense. Available at : web.media.mit.edu/~hugo/montylingua.

Matlin, M.W. (2001). *La cognition, introduction à la psychologie cognitive*. Paris : De Boeck Université.

Marazziti, D., Consoli, G., Picchetti, M., Carlini, M., & Faraveli, L. (2010). Cognitive impairment in major depression. *Européan Journal of Pharmacology*, 626, 83-86.

BIBLIOGRAPHIE

McNamara, D.S., Kintsch, E., Butler-Songer, N., & Kintsch, W. (1996). Are good texts always better? Interactions of text coherence, background knowledge, and levels of understanding in learning from text. *Cognition and Instruction*, 14(1), 1-43.

Nef, F., & Hupet, M. (1992a). Étude comparative du langage spontané d'adultes jeunes et âgés. *L'année Psychologique*, 92, 4, 511-525.

Nef, F., & Hupet, M. (1992b). Les manifestations du vieillissement normal dans le langage spontané oral et écrit. *L'année Psychologique*, 92, 3, 393-419.

Ortega, J. & Rémond-Bésuchet, C. (2007). Fluence verbale : de 60 à 89 ans une épreuve rapide et standardisée en fonction du niveau d'étude. *Glossa*. 101, 42-59.

Piolino, P., Desgranges, B., & Eustache, F. (2000). *La mémoire autobiographique : théorie et pratique*. Marseille : Solal.

Riegel, M., Pellat, J.C., & Rioul, R. (1999). *Grammaire méthodique du français*. Paris : Presses Universitaires de France.

Riley, K.P., Snowdon, D.A., Desrosiers M.F., & Markesbery W.R. (2005). Early life linguistic ability, late life cognitive function, and neuropathology : findings from the Nun Study. *Neurobiology of Aging*, 26, 341-347.

Rondal, J.A., & Seron, X. (2000). *Troubles du langage: bases théoriques, diagnostic et rééducation*. Bruxelles : Mardaga.

Sellal, F., & Kruczek, E. (2007). *Maladie d'Alzheimer*. Paris : Doin.

Signoret, J.L., & Hauw, J.J. (1994). *Maladie d'Alzheimer et autres démences*. Paris : Flammarion médecine sciences.

Snowdon, D.A., Kemper S.J., Mortimer J.A., Greiner L.H., Wekstein D.R., & Markesbery W.R. (1996). Linguistic ability in early life and cognitive function and Alzheimer's disease in late life. *The journal of the american medical association*, 275, 7, 528-532.

Snowdon, D.A. (1997). Aging and Alzheimer's disease: Lessons from the Nun Study. *Gerontologist*, 37, 150-156.

Snowdon, D.A., Greiner, L.H., Kemper, S., Nanayakkara, N., & Mortimer, J.A. (1999). Linguistic ability in early life and longevity: Findings from the Nun Study. In J.M. Robine, B. Forette, C. Franchesci, & M. Allard (Eds), *The paradoxes of Longevity* (pp.103-113). Amsterdam, Springer.

Snowdon, D.A., Greiner, L.H., & Markesbery, W.R. (2000). Linguistic Ability in Early Life and the Neuropathology of Alzheimer's Disease: Findings from the Nun Study, *Annals of the New York Academy of Sciences*, NYAS.

Taconnat, L., & Isingrini, M. (2008). In K. Dujardin, & P. Lemaire, *Neuropsychologie du vieillissement normal et pathologique* (pp.45-63). Issy-les-Moulineaux, Elsevier Masson.

BIBLIOGRAPHIE

Touchon, J., & Portet, F. (2002). *La maladie d'Alzheimer*. Paris : Masson.

Turner, K., & Green, E. (1977). The construction and use of a propositionnal text base. *Technical report*, Vol 63, University of Colorado.

Valdois, S., Joannette, Y, Poissant, A., Ska, B., & Dehaut, F. (1990). Heterogeneity in the cognitive profile of normal elderly. *Journal of Clinical and Experimental Neuropsychology*, 12, 587-596.

Van der Linden, M. (1999). Neuropsychologie des syndromes démentiels. In X. Seron, M. Jeannerod, *Neuropsychologie humaine* (pp.558-566). Bruxelles, Mardaga.

Van der Linden, M., & Hupet, M. (1994). *Le vieillissement cognitif*. Paris : Presses universitaires de France.

ANNEXES

Annexe I : Grille de Poitrenaud (Kalafat, Hugonot-Diener, & Poitrenaud, 2005)

Evaluation du niveau socio-culturel (d'après J. Poitrenaud)

Entretien Standardisé : Droitier Gaucher

- Quelle est (était) votre profession ?
- Avez-vous déjà exercé d'autres activités professionnelles au cours de votre carrière (enregistrer in extenso les réponses du sujet) ?
- À quel âge avez-vous commencé à travailler ?
- Jusqu'à quel âge avez-vous poursuivi des études ? (ou) jusqu'à quel âge êtes-vous allé(e) à l'école ?
- Quel est le diplôme le plus élevé que vous ayez obtenu ?
- Avez-vous passé des diplômes professionnels ou techniques ?
- Si oui, ce (ou ces) diplômes vous a (ont) il(s) aidé à obtenir votre emploi ou vous a (ont) il(s) permis de progresser dans la hiérarchie professionnelle ?
- Quel a été votre niveau de qualification professionnelle le plus élevé (manœuvre, ouvrier qualifié, agent de maîtrise, cadre moyen, cadre supérieur) ?

On ne posera pas cette question si le niveau de qualification était déjà évident d'après les réponses du sujet concernant sa vie professionnelle.

A. Détermination du niveau d'éducation scolaire | ____ |

1. Pas de diplôme, ou au maximum un CAP pour adultes.
2. CEP ou CAP + CEP
 CEP ou notion d'une scolarité secondaire (au maximum jusqu'à la fin d'une classe de 4^{ème})
 CEP ou CEP + études techniques courtes (au maximum jusqu'à la fin d'une classe de 4^e T)
3. Notion d'une scolarité allant de la fin d'une classe de 3^{ème} (avec ou sans le Brevet des Collèges) à la fin d'une classe de terminale (sans le Bac complet)
4. Réussite à un examen de niveau Bac, ou Bac et plus

B. Les indications du niveau socioprofessionnel peuvent conduire à améliorer le classement du sujet coté 1, 2 ou 3 dans l'échelle de niveau d'éducation scolaire.

Généralement l'amélioration est d'un échelon | ____ |

C. Niveau socioculturel effectivement atteint (score maximum 4) | ____ |

Annexe II : Fiche de renseignements

FICHE DE RENSEIGNEMENTS

Date de l'entretien :

Nom :

Prénom :

Sexe :

Age :

Niveau d'études :

Profession(s) exercée(s) :

Retraité(e) :

Année d'arrêt d'exercice de la profession :

Activités de loisirs :

Annexe III : Densidées - Calcul automatique de la densité des idées dans un corpus oral

RECITAL 2010, Montréal, 19-23 juillet 2010

Densidées : calcul automatique de la densité des idées dans un corpus oral

Hyeran Lee¹, Philippe Gambette², Elsa Maillé³, Constance Thuillier³

(1) Praxiling - Université Montpellier 3 / CNRS

(2) LIRMM - Université Montpellier 2 / CNRS

(3) ISTR - Université Claude Bernard Lyon 1

hlee1@univ-montp3.fr

Résumé. La densité des idées, qui correspond au ratio entre le nombre de propositions sémantiques et le nombre de mots dans un texte reflète la qualité informative des propositions langagières d'un texte. L'apparition de la maladie d'Alzheimer a été reliée à une dégradation de la densité des idées, ce qui explique l'intérêt pour un calcul automatique de la densité des idées. Nous proposons une méthode basée sur un étiquetage morphosyntaxique et des règles d'ajustement, inspirée du logiciel CPIDR. Cette méthode a été validée sur un corpus de quarante entretiens oraux transcrits et obtient de meilleurs résultats pour le français que CPIDR pour l'anglais. Elle est implémentée dans le logiciel libre Densidées disponible sur <http://code.google.com/p/densidees>.

Abstract. Idea density, which is the ratio of semantic propositions divided by the number of words in a text, reflects the informative quality of the sentences of a text. A decreasing idea density has been identified as one of the symptoms of Alzheimer's disease, which explains the interest in an automatic calculation of idea density. We propose a method based on part-of-speech tagging followed by adjustment rules inspired from the CPIDR software. This method was validated on a corpus of 40 transcribed conversations in French and obtains better results in French than CPIDR in English. It is implemented in the free software Densidées available at <http://code.google.com/p/densidees>.

Mots-clés : densité des idées, analyse prédicative, étiquetage sémantique, psycholinguistique

Keywords: idea density, propositional analysis, semantic tagging, psycholinguistics

HYERAN LEE, PHILIPPE GAMBETTE, ELSA MAILLÉ, CONSTANCE THULLIER

1 Introduction

Pour étudier les mécanismes qui sous-tendent la pensée humaine, la psychologie cognitive n'ayant accès qu'aux productions du sujet, considère qu'elles possèdent en leur sein les marques des mécanismes qui les ont engendrées. En ce sens, les productions discursives sont privilégiées car le langage est à la fois le support et le produit de la pensée. Autrement dit, les structures langagières peuvent refléter les structures cognitives. Ainsi, une certaine forme d'analyse du discours revient à modéliser des phénomènes cognitifs.

La fonction primitive du langage est la fonction référentielle (Jakobson, 1963), c'est-à-dire qu'il sert à transmettre à autrui des informations du monde réel par la symbolisation. Cette fonction est accomplie en véhiculant du sens. L'activité sémantique consiste donc à produire du sens dans l'intellect du récepteur, c'est-à-dire la formation de représentation mentale chez l'interlocuteur par l'intermédiaire du langage.

Depuis la logique aristotélicienne ainsi que dans la logique classique de Frege (1967, 1971), en passant par la théorie psychologique des réseaux sémantiques propositionnels d'Anderson (1976), les chercheurs se sont intéressés au traitement sémantique de l'information et ont tenté de définir la structure cognitive. Ils ont fait l'hypothèse que l'information dans la mémoire est organisée sous forme propositionnelle. En effet, un mot isolé seul ne suffit pas à créer une idée, c'est l'ensemble de propriétés et de relations s'y rapportant qui permet d'appréhender et de produire la signification psychologique. Ainsi, une proposition est formée d'une association de deux unités : le prédicat et l'argument.

Soit dans la phrase « Le chien poursuivait un chat dans le jardin », l'exemple emprunté de Le Ny (1989), les mots « chien », « poursuivre », « chat », « dans », « jardin » expriment les concepts qui sont des représentations permanentes en mémoire à long terme. Si « chien », « chat » et « jardin » renvoient aux objets, et relèvent d'une catégorie lexicale qui a pour fonction principale la désignation d'objets, en l'occurrence les substantifs, « poursuivre » et « dans » représentent des relations entre ces objets, et ont comme catégorie lexicale verbe et préposition respectivement. La signification de cette phrase peut donc être représentée comme montré en figure 1.

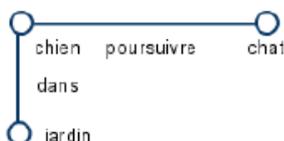


Figure 1 : Réseau propositionnel simple, représentant la signification de la phrase « Le chien poursuivait un chat dans le jardin » (Le Ny, 1989 : 205)

Cependant, le graphe ci-dessus pose le problème de l'insertion de « dans ». Ainsi, Le Ny propose une autre figure (voir figure 2) qui représente la « représentation sémantique » telle qu'elle existe dans l'intellect d'un sujet humain suite à la compréhension de cette phrase.

Les nœuds rectangulaires représentent les concepts génériques qui font références à des objets (« chien », « chat », « jardin »), à des événements (« poursuivre »), et à des relations dans l'espace (« dans ») ; les petits nœuds circulaires représentent l'instanciation de ces concepts, c'est-à-dire l'utilisation des mots qui expriment ces concepts ; et les arcs qui les relient.

DENSIDÉES : CALCUL AUTOMATIQUE DE LA DENSITÉ DES IDÉES DANS UN CORPUS ORAL

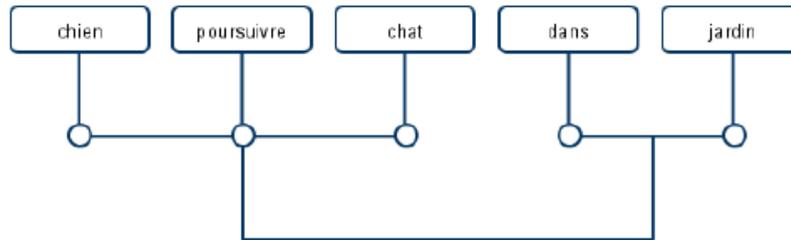


Figure 2 : Représentation sémantique de la signification de la phrase « Le chien poursuivait un chat dans le jardin » (Le Ny, 1989 : 206)

Selon Frege (1967), il y a des unités dites « saturées » qui se suffisent à elles-mêmes, comme les substantifs dans notre exemple. Il y a également des unités « non saturées », qui nécessitent d'autres éléments afin de constituer une idée complète, comme dans notre cas « poursuivre » et « dans ». On parle des *arguments* qui sont des entités référentielles pouvant correspondre à des êtres ou des objets, et des *prédicats* qui sont des unités requérant des arguments. Ainsi les prédicats assignent des propriétés aux arguments ou définissent la relation entre les arguments (Coirier et al., 1996).

En logique on écrit usuellement les prédicats en majuscules, et les arguments en minuscules :

P1. POURSUIVRE (chien, chat)

P2. DANS (P1, jardin)

On peut remarquer qu'un prédicat peut avoir comme argument d'autres propositions. On appelle *prédicat de premier rang* le prédicat qui n'implique que des arguments objets (comme dans P1) et *prédicat de rang supérieur* le prédicat qui implique d'autres propositions (comme dans P2).

En logique, la proposition est définie comme la plus petite unité porteuse de sens du discours à laquelle une valeur de vérité peut être attribuée. En revanche, en psychologie, elle est définie comme « la plus petite unité sémantique intégrée susceptible d'être traitée ou mémorisée » (Le Ny, 1987: 27). L'actualisation des propositions dans le discours fait donc ressortir non seulement la connaissance lexicale mais également la connaissance du monde. Une méthode de description de la signification de l'énoncé peut ainsi fournir un moyen d'étudier le fonctionnement cognitif.

À partir de ces théories logique et psychologique, Kintsch (1974) a développé une méthode d'analyse linguistique qui permet de modéliser la manière dont l'humain encode les informations, appelée analyse prédictive. Il part du postulat que la forme dominante de la représentation cognitive du langage est de nature propositionnelle. L'analyse prédictive permet d'extraire les propositions sémantiques dans le discours par la concaténation des unités élémentaires du sens. Selon Ghiglione et al., « Si l'on considère que la prédication qui s'exprime dans un message linguistique est une activité cognitive essentielle de l'homme et que, sous-jacent à la réalisation de surface, c'est-à-dire au mot, se trouve un concept, on peut estimer que l'analyse prédictive, outil de description sémantique des textes, est pour le psychologue la transcription d'une activité cognitive » (Ghiglione et al., 1995 : 49).

Si l'analyse prédictive reflète l'activité cognitive, la densité des idées quant à elle permet de la quantifier. Ainsi, en ramenant le nombre de propositions sémantiques au nombre de mots produits dans le discours, on peut mesurer la densité des idées d'un discours. La densité des idées permet donc de mesurer la qualité informative du langage. Une densité des idées élevée peut refléter l'aptitude d'un locuteur à exprimer efficacement ses idées ainsi que leur interrelation complexe. Par contre, une faible densité des idées dans

HYERAN LEE, PHILIPPE GAMBETTE, ELSA MAILLÉ, CONSTANCE THUILLIER

le discours peut révéler un discours peu efficient, du fait de l'utilisation d'un plus grand nombre de mots pour exprimer les idées essentielles.

Cette méthode d'analyse du discours a montré son utilité dans plusieurs domaines de la psycholinguistique appliquée : la compréhension du texte (Kintsch et al., 1973 ; Kintsch, 1998), la mémoire (Thorson et al., 1984), la maladie d'Alzheimer (Snowdon, 1996 ; Lee et al., 2009), la qualité de prise de note des étudiants (Takao et al., 2002), le vieillissement (Kemper et al., 2001), la schizophrénie (Covington et al., 2007), le genre du discours (Covington, 2009).

Si cette méthode est considérée à l'heure actuelle comme l'outil le plus complet pour étudier le processus cognitif via le langage, son analyse manuelle longue et fastidieuse fait souvent obstacle à sa mise en application dans la recherche.

Parallèlement à ces applications en psycholinguistique, plusieurs problématiques de traitement automatique des langues naturelles (traduction automatique, extraction d'informations, etc.) ont fait appel à l'analyse prédicative, avec l'objectif de représenter un texte en langue naturelle par une formule logique, en langage des prédicats du premier ordre. Divers formalismes sont par exemple présentés par François (1991).

Toutefois, cette approche s'est heurtée aux limites du formalisme de représentation, et de telles analyses prédicatives d'étiquetage sémantique sont actuellement utilisées en pratique uniquement sur des tâches très spécifiques et des corpus ciblés, comme dans le système présenté par Meurs et al. (1998). Sur des corpus plus généraux, la couverture des bases de données sémantiques (comme FrameNet) est trop faible, et l'analyse sémantique conduit à des taux d'erreurs importants. Ces erreurs sont amplifiées sur les corpus oraux du type de ceux qui nous intéressent dans le contexte du calcul de la densité des idées, du fait des contraintes qu'ils induisent : fragments, idées répétées, marqueurs discursifs, etc.

Le calcul de la densité des idées ne nécessite en fait pas de calculer l'ensemble des prédicats et de leurs arguments, mais seulement de compter les prédicats. Nous avons donc choisi d'éviter d'utiliser cette approche sémantique, et d'utiliser plutôt les travaux de Brown et al. (2008) sur la langue anglaise. Ceci dans le but de concevoir une approche du comptage des propositions par un ensemble de règles appliquées après un étiquetage morphosyntaxique du texte. En particulier, nous nous intéresserons à l'analyse du discours oral, même si l'outil que nous proposons est aussi destiné à l'écrit.

2 Calcul de la densité des idées

Pour calculer la densité des idées d'un texte en anglais, Brown et al. proposent le logiciel CPIDR (Brown et al., 2008) qui étiquette chaque mot du texte comme prédicat, ou bien comme non-prédicat. L'idée principale de l'étiquetage est qu'un prédicat correspond typiquement à un verbe, un adjectif, à un adverbe, une préposition, ou une conjonction. Ces unités constituent des prédicats car elles apportent une contribution sémantique forte sur les arguments. Ainsi, l'étiquetage morphosyntaxique est à la base du calcul approximatif de la densité des idées.

Cette étape d'étiquetage morpho-syntaxique, traitée dans le cas de CPIDR par le logiciel MontyLingua (Liu, 2004), est suivie d'un post-traitement à base de règles destiné à corriger les erreurs d'étiquetage morpho-syntaxique qui ont une influence sur le nombre de prédicats, à traiter le cas spécifique des corpus oraux (avec une gestion basique de certaines répétitions ou auto-corrrections), et enfin à ajuster le calcul du

DENSIDÉES : CALCUL AUTOMATIQUE DE LA DENSITÉ DES IDÉES DANS UN CORPUS ORAL
 nombre de prédicats. Cette méthode est efficace en anglais, puisque CPIDR obtient généralement un meilleur accord avec un ensemble d'étiqueteurs humains que les étiqueteurs humains entre eux.

Nous avons donc choisi de suivre les mêmes principes, en apportant une attention particulière au caractère oral de notre corpus, important à la fois pour nos objectifs d'utilisation de la densité des idées, mais aussi pour les contraintes qu'il implique. En effet, si le caractère éphémère du discours oral permet de voir clairement la dynamique de la conceptualisation en cours en fournissant des informations précieuses sur le fonctionnement cognitif, la construction orale non linéaire et les phénomènes oraux particuliers tels que les mots fragmentés, les énoncés inachevés, les substitutions, les omissions, les ratages, les corrections, les reformulations, les répétitions, les interjections, les habitudes du langage (gimmicks) et les pauses remplies rendent son analyse complexe. Ainsi, nous avons choisi de proposer un prétraitement manuel du corpus pour marquer certaines caractéristiques spécifiques à l'oral qui ne sont pas traitables d'un point de vue automatique. Nous avons utilisé les crochets [] pour marquer les mots fragmentés qui ne doivent pas être comptés ni comme un prédicat ni comme un mot ; et les parenthèses () pour entourer les mots qui doivent être intégrés dans le compte du nombre total de mots mais ne doivent pas être marqués comme prédicats, par exemple pour les phrases inachevées ou le marqueur discursif "bon" qui ne doit pas être traité comme un adjectif (donc un prédicat), mais qui est considéré comme un mot contrairement aux autres marqueurs discursifs non-lexicales (e.g. "bah", "hein", etc.). Selon Ghiglione et al. (1995) l'« analyse prédicative est une analyse essentiellement sémantique qui requiert une attention soutenue au sens véhiculé, ce qui rend parfois nécessaire le recours à des équivalents sémantiques. Ceci est particulièrement justifié dans un texte conversationnel souvent complexe à analyser parce qu'il est en général beaucoup plus elliptique qu'un texte écrit » (Ghiglione et al., 1995: 45). Dans cette étude, nous avons choisi d'éviter toute intervention interprétative de la part des examinateurs et de préférer l'analyse la plus objective possible des phénomènes oraux.

L'implémentation de ces principes pour le français nous a fait recourir à TreeTagger pour l'étiquetage morphosyntaxique du texte. C'est ensuite un ensemble de 35 règles d'ajustement que nous proposons pour déterminer si un mot est un prédicat ou non. Des exemples de règles sont fournis en figure 3, elles sont intégralement décrites dans le manuel d'utilisation de Densidées. Dans la mesure du possible les numéros de règles utilisés dans CPIDR ont été conservés dans Densidées. En outre, le logiciel Densidées est écrit en Python de façon commentée et très lisible. Il fonctionne en ligne de commande mais peut également être appelé depuis Windows par l'intermédiaire d'une interface graphique montrée en figure 4.

Règle 208 - Comparatif : "que" n'est pas proposition après "autant", "moins", "pire", "plus"

Règle 301 - Verbes de liaison ("apparaître", "être", "sembler", "devenir", "paraître", "rester", "demeurer") non proposition si suivis d'un adjectif ou d'un adverbe

Figure 3. Exemples de règles de Densidées

3 Résultats

Un corpus oral correspondant à des entretiens avec 40 sujets a été constitué. Pour le recueillir, nous avons mené un entretien individuel semi-dirigé auprès de 40 sujets volontaires. Ces sujets sont âgés de 65 à 85 ans, domiciliés dans le Rhône, l'Ain, le Jura, les Pyrénées-Atlantiques et le Puy de Dôme. Cet entretien dure environ 35 minutes dont 5 à 15 minutes pour l'enregistrement du discours oral. Une narration libre de l'évocation d'un souvenir personnel a été demandée pour la production du discours spontanée, et une description d'image « voleur du biscuit », tirée de Boston Diagnostic Aphasia Examination (Goodglass et

HYERAN LEE, PHILIPPE GAMBETTE, ELSA MAILLÉ, CONSTANCE THUILLIER al., 1983), pour un discours descriptif. Tous les entretiens ont été enregistrés et transcrits individuellement avec une transcription orthographique standard. Ce corpus a été tronqué de manière à garder environ 300 mots par transcription pour que les corpus soient comparables statistiquement.

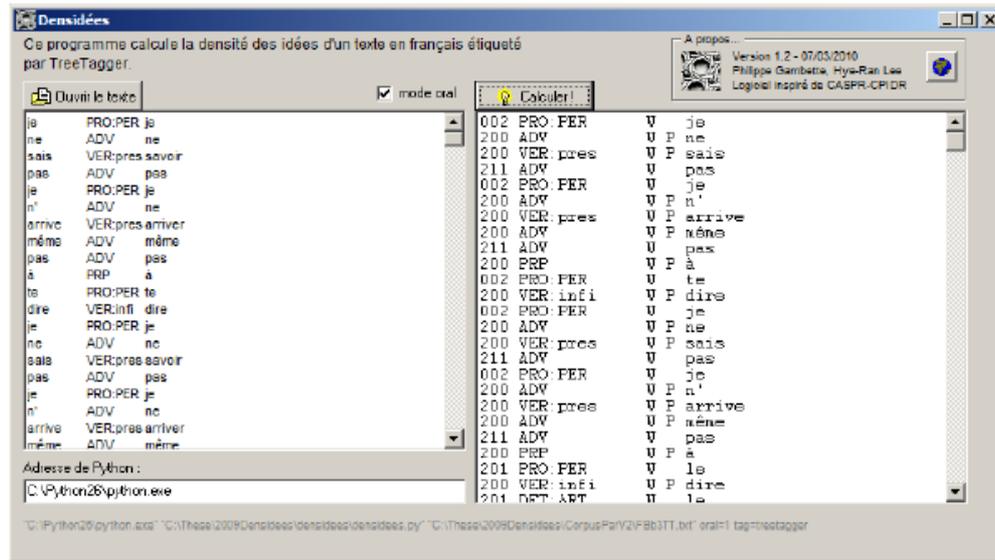


Figure 4 : L'interface graphique de Densidées sous Windows

Ce corpus a alors été parenthésé, et étiqueté manuellement en mots et prédicats par trois experts qui ont travaillé en commun, pour aboutir à une base totale de 13939 mots dont 5747 prédicats. On peut noter que des tentatives préalables d'étiquetage en parallèle ont fait apparaître des différences d'étiquetage des prédicats de 5 à 10% entre les experts.

Nous avons testé sur ce corpus la version 1.2 de Densidées, qui a fourni les résultats montrés dans la figure 5. On peut noter, entre la densité des idées calculée manuellement et automatiquement, pour les 40 textes, un coefficient de corrélation de 0.972, là où CPIDR obtenait 0.942. Pour une évaluation plus fine du logiciel, nous avons choisi de déterminer le taux de faux négatifs (prédicats non étiquetés comme telles) et de faux positifs (non-prédicats étiquetés comme prédicats) : respectivement 2.7% et 3.1%. Comme la formule de densité des idées fait intervenir le nombre total de prédicats, ces deux types d'erreurs se compensent, pour arriver à un taux d'erreur moyen de 0.5% sur le nombre de prédicats.

Le corpus a alors été séparé en une base de test (correspondant à 10 sujets pour assurer une variété dans les scores de densité des idées) de 3728 mots et 1548 prédicats et une base de validation de 10211 mots et 4199 prédicats. La base de test a été utilisée pour évaluer la pertinence de chaque règle, en testant l'effet de sa suppression.

Pour évaluer la qualité d'un étiquetage automatique, on calcule la F-mesure, qui se base sur la précision (proportion de prédicats corrects parmi les prédicats trouvés automatiquement) et la couverture (proportion des prédicats corrects trouvés par Densidées sur l'ensemble des prédicats corrects). En n'utilisant que la règle 200, qui étiquette les conjonctions, numéraux, déterminants, prépositions, adjectifs, adverbes et verbes comme prédicats, on obtient 29 faux négatifs et 1003 faux positifs, ce qui correspond à une F-mesure de 0,747 sur la base de test. Si l'on prend en compte l'ensemble des 35 règles de la version 1.2, on arrive à une F-mesure de 0,975.

DENSIDÉES : CALCUL AUTOMATIQUE DE LA DENSITÉ DES IDÉES DANS UN CORPUS ORAL

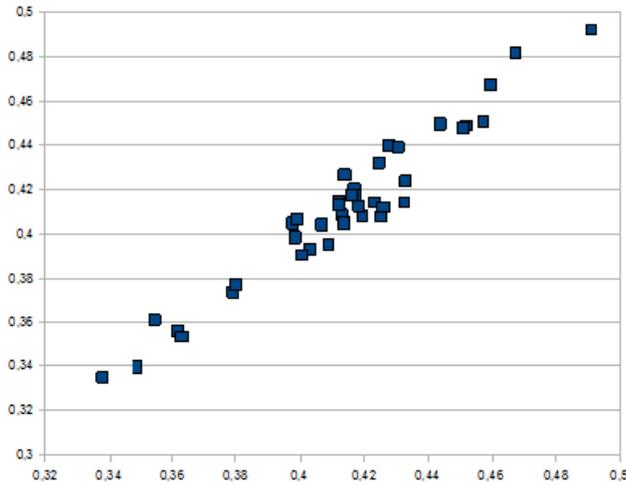


Figure 5. Représentation de la densité des idées calculée automatiquement en fonction de la densité des idées calculée manuellement.

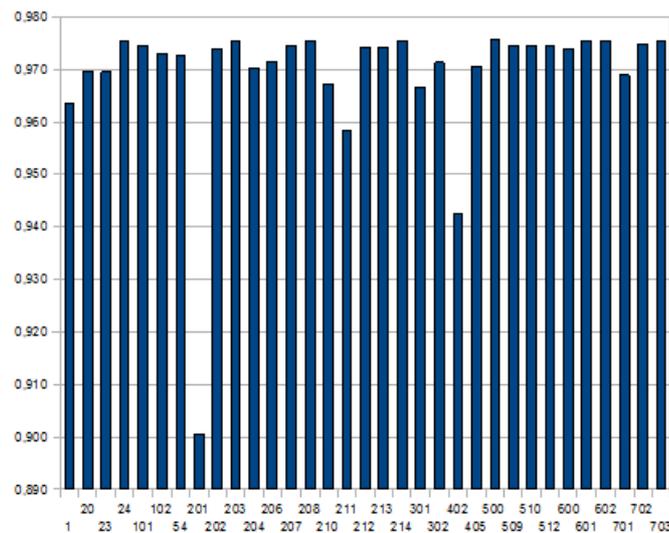


Figure 6. F-mesure obtenue après suppression de chaque règle de la version 1.2 de Densidées

La figure 6 illustre la dégradation de la F-mesure induite par la suppression de chaque règle. Ainsi, on constate par exemple qu'en retirant la règle 201 dédiée à l'étiquetage des déterminants "un", "une", "le", "la", etc. (qui ne sont pas des prédicats), la F-mesure décroît à 0,900. Inversement, la suppression de certaines règles (24, 203, 208, 214, 601, 602, 703) n'a aucun effet, voire améliore légèrement la F-mesure en supprimant un faux positif pour la règle 500.

HYERAN LEE, PHILIPPE GAMBETTE, ELSA MAILLÉ, CONSTANCE THUILLIER

Même si la suppression de ces règles semble n'avoir aucun effet sur le corpus de test, elle pourrait en avoir sur d'autres textes, ce qui explique que nous les laissons dans Densidées. En effet, certaines de ces règles sont plus adaptées pour des discours écrits (comme la 214 qui considère "si ... alors" comme un seul prédicat). D'autres, à l'inverse, sont prévues pour le discours oral, mais ciblées sur des marqueurs discursifs particuliers que les locuteurs n'utilisent pas nécessairement : la règle 602 considère par exemple que "donc" n'est pas un prédicat après les verbes "dire", "comprendre", et "aller".

Munis de ce score de qualité qu'est la F-mesure, nous proposons la méthodologie suivante pour l'ajout de nouvelles règles ou la modification de règles existantes dans Densidées. Nous utilisons la base de test à titre d'exploration pour évaluer l'évolution de la F-mesure suite à des modifications du programme, la base de validation sert quant à elle à valider la pertinence des modifications, en vérifiant que les modifications proposées à partir du corpus de test ne sont pas biaisées par les spécificités linguistiques de ce corpus. Par exemple, par rapport à la version 1.2, parmi toutes les modifications de règles testées, une seule (modification de la règle 301), raisonnable du point de vue linguistique, a permis d'améliorer la F-mesure en atteignant 0,978 sur le corpus de test. Sur la base de validation, cette modification a permis de passer d'une F-mesure de 0,969 à 0,972. Ainsi, elle sera intégrée dans la version 1.3 de Densidées.

4 Discussion

Un entretien oral de 300 mots nécessite environ 25 minutes de transcription, 10 minutes de parenthésage et 35 minutes d'étiquetage manuel des idées. Ainsi, Densidées permet de diviser par deux le temps total nécessaire à l'évaluation manuelle de la densité des idées. Il permet surtout de normaliser l'étiquetage en évitant les spécificités d'étiquetage des experts humains sur certains mots difficiles à étiqueter. Précisons que le temps de parenthésage est principalement dû à la relecture attentive du texte nécessaire à la détection des passages vides de sens ou correspondant à des idées répétées. Si les entretiens seraient réalisés dans l'optique exclusive de calculer la densité des idées, par exemple pour des applications médicales, la transcription obéit à des contraintes ciblées, et peut se faire plus rapidement en conjonction avec l'étiquetage. Par exemple les passages incompréhensibles placés entre crochets ne sont alors simplement pas transcrits.

Le taux d'erreur est tout à fait acceptable pour une utilisation dans le cadre de l'analyse du discours de patients atteints de la maladie d'Alzheimer (Lee et al., 2010). En effet, tout comme la densité des idées calculée de façon manuelle, il est possible de vérifier qu'une faible densité des idées selon Densidées est liée de façon statistiquement significative à la maladie d'Alzheimer.

5 Conclusion

Le logiciel Densidées fournit actuellement une approximation tout à fait satisfaisante de la densité des idées d'un discours oral transcrit selon la méthodologie que nous proposons ici. Cette méthodologie fait intervenir une détection humaine des idées répétées, et il pourrait être envisagé d'aborder ce problème de façon automatique. Toutefois, nous pensons que l'effort d'étiquetage humain des idées répétées pendant la transcription constitue un effort minime, et envisageons plutôt pour de prochaines versions du logiciel d'améliorer l'étiquetage morphosyntaxique sur lequel se basent les règles de Densidées, en faisant appel au logiciel Cordial au lieu de TreeTagger.

On peut également noter que l'étiquetage automatique des prédicats dépend fortement des habitudes langagières récurrentes des locuteurs. Ainsi, compléter cette approche par règles avec une partie statistique

DENSIDÉES : CALCUL AUTOMATIQUE DE LA DENSITÉ DES IDÉES DANS UN CORPUS ORAL (par exemple pour détecter des mots fréquents inattendus) pourrait aider à repérer ces tics, et proposer automatiquement de nouvelles règles adaptées.

Enfin, la densité des idées a d'autres applications citées plus haut, dont la mesure du niveau de technicité d'articles scientifiques en langue anglaise. Il serait intéressant de calculer la densité des idées de corpus écrits en français pour tenter d'identifier certains genres de textes associés à une densité des idées très élevée ou au contraire très basse. La densité des idées pourrait aussi s'ajouter aux paramètres pertinents pour choisir des textes en fonction de leur niveau de technicité en enseignement du français langue étrangère (Thomas, 2009).

Références

- ANDERSON, J. (1976). *Language, memory and thought*. Hillsdale, NJ: Erlbaum Associates.
- BROWN, C., SNODGRASS, T., KEMPER, S., HERMAN, R. & COVINGTON, M. (2008). Automatic measurement of propositional idea density from part-of-speech tagging. *Behavior Research Methods*, 40(2), 540-545.
- COIRIER, P., GAONACH, D., PASSERAULT, J.-M. (1996). *Psycholinguistique textuelle*. Paris: Armand Colin.
- COVINGTON, M. (2009). Idea Density: A potentially informative characteristic of retrieved documents. *Proceedings, IEEE SoutheastCon*.
- COVINGTON, M., RIEDEL, W., BROWN, C., HE, C. MORRIS, E., WEINSTEIN, S., et al. (2007). Does ketamine mimic aspects of schizophrenic speech? *Journal of Psychopharmacology* 21, 338-346.
- FRANÇOIS, J. (1991). Pertinence linguistique des représentations propositionnelles de la sémantique cognitive. *Sémiotiques*, 1(1).
- FREGE, G. (1967). *The basic laws of arithmetic*. Berkeley: University of California.
- FREGE, G. (1971). *Écrits logiques et philosophiques*. Paris : Presses Universitaires de France.
- GHIGLIONE, R., KEKENBOSCH, C., LANDRÉ, A. (1995). *L'analyse cognitivo-discursive*. Grenoble: Presses Universitaires de Grenoble.
- GOODGLASS, H., KAPLAN, E. (1983) *The assessment of aphasia and related disorders*. Philadelphia : Lea and Febiger.
- JAKOBSON, R. (1963). *Essais de linguistique générale*. Paris : Éditions de Minuit.
- KEMPER, S., GREINER, L., MARQUIS, J., PRENEVOST, K., MITZNER, T. (2001). Language decline across life span: findings from the Nun study, *Psychology and Aging* 16(2), 227-239.
- KINTSCH, W. (1874). *The representation of meaning in memory*. Hillsdale, NJ : Erlbaum.
- KINTSCH, W., KEENAN, J. (1973). Reading rate and retention as a function of the number of propositions in the base structure of sentences. *Cognitive Psychology* 5 (3), 257-274.

- HYERAN LEE, PHILIPPE GAMBETTE, ELSA MAILLÉ, CONSTANCE THULLIER
KINTSCH, W., KEENAN, J. (1978). Toward a model of text comprehension and production. *Psychological Review* 85, 363-394.
- LE NY, J.-F. (1987). Sémantique psychologique. In *Problèmes de psycholinguistique*. Bruxelles : Pierre Mardaga.
- LE NY, J.-F. (1989). *Science cognitive et compréhension du langage*. Paris : Presses Universitaires de France.
- LEE, H., BARKAT-DEFRADAS, M. (2009). La densité des idées : un modèle d'analyse du discours pertinent pour le diagnostic précoce de la maladie d'Alzheimer ? *Actes des 8ème Rencontres Jeunes Chercheurs en Parole*, Avignon, 16-18 Novembre.
- LEE, H., GAMBETTE P., BARKAT-DEFRADAS, M. (2010). Utilisation de l'analyse textuelle automatique dans la recherche sur la maladie d'Alzheimer. Poster au *Colloque international des jeunes chercheurs en Didactique des Langues et en Linguistique (CEDIL 2010)*.
- LIU, H. (2004). MontyLingua : An end-to-end natural language processor with common sense. Available at: web.media.mit.edu/~hugo/montylingua
- MEURS M.-J., DUVERT F., BÉCHET F., LEFÈVRE F., DE MORI R. (2008). Annotation en Frames Sémantiques du corpus de dialogue MEDIA. *Actes de TALN 2008*.
- SCHMID, H. (1994). Probabilistic Part-of-Speech tagging using decision trees. In D., Jones & H., Somers (Eds.), *New Methods in Language Processing*, 154-164, London : UCL Press.
- SNOWDON, D., KEMPER, S., MORTIMER, J., GREINER, L., WEKSTEIN, D., MAYKESBERY, W. (1996). Linguistic ability in early life and cognitive function and Alzheimer's disease in late life: findings from the Nun Study, *JAMA* 275, 528-532.
- TAKAO, A., PROTHERO, W., KELLY, G. (2002). Applying argumentation analysis to assess the quality of university oceanography students' scientific writing. *Journal of Geoscience Education* 50. 40-48.
- THOMAS, F. (2009). Modèles statistiques pour l'estimation automatique de la difficulté de textes de FLE. *Actes de RECITAL 2009*.
- THORSON, E., SNYDER, R. (1984). Viewer recall of television commercials: structure of commercial scripts. *Psychological Review* 85, 363-394.

Annexe IV : Exemples de corpus

❖ FHb1 : sujet sain, femme, niveau 4, 76 ans

Constance : alors est-ce que vous pourriez me raconter le plus beau jour de votre vie

FHb1 : oh c'est peut-être le jour de la naissance de ma première fille oui y'en a eu beaucoup des beaux jours mais le plus peut-être bouleversant c'est bah on m'avait endormie pour l'accouchement donc quand je me suis éveillée je l'ai découverte à côté de moi dans le berceau et oui je crois que ça ça m'a vraiment remuée oui oui j'étais toute seule mon mari était encore étudiant à [ville de grande importance] donc il est arrivé après par le train mes parents à ce moment-là étaient repartis chez eux le temps que je sorte de mon anesthésie j'étais toute seule face à ce bébé que je découvrais et j'avais trouvé ça bah extraordinaire oui c'est certainement le plus beau jour et ma mère m'avait dit surtout tu n'embrasseras pas ton bébé parce que tu peux lui donner des microbes ouais eh bah c'était une ancienne institutrice ça ressortait hein et première chose que j'ai faite dès que j'ai été un peu capable de me redresser j'ai pris mon bébé et je l'ai embrassé voilà

Constance : est-ce que maintenant vous pourriez me raconter une bagarre ou une dispute qui vous est arrivée et qui vous a marquée

FHb1: je ressors toujours une bagarre je devais avoir 13 ans 14 ans j'étais un garçon manqué faut croire euh notre prof de math était absent et était remplacé par un pion de [lycée de ville de moyenne importance] je me souviens et ce pauvre homme il était chahuté par tout le monde il nous pose une interrogation une composition comme on disait à l'époque une composition de maths et euh y a eu tout un groupe de filles qui évidemment ne travaillaient pas particulièrement bien qui ont tout copié et moi élevée comme je l'étais par mes parents j'ai mis mon point d'honneur à faire ce que je pouvais et lorsqu'il a rendu les notes j'avais une moins bonne note évidemment que ces filles-là et à la récréation y en a une qui est venue se moquer de moi j'avais un frère donc j'étais un peu habituée à me battre avec un garçon et je lui ai flanqué une volée toute fille que j'étais voilà ça ça me revient souvent cette bagarre

Constance : alors maintenant est-ce que vous pourriez me raconter votre plus mauvais souvenir

FHb1: oui la mort de mon père bien-sûr deuil la mort de mon mari j'aimais beaucoup mon père il comptait beaucoup beaucoup pour moi et je suis arrivée il venait de mourir quelques minutes avant alors évidemment c'était bouleversant mais la mort de mon mari bah on était en train de parler tous les deux il était déjà bien bien malade, allongé dans le lit on parlait de mon dernier fils qui prenait l'avion à ce moment-là il était à [ville de moyenne importance] il prenait l'avion pour venir il avait obtenu une permission pour venir voir son père qui était très mal et ce que mon mari m'a dit il m'a fait venir à côté de lui il m'a dit je me sens pas bien viens t'asseoir à côté de moi et il me dit tiens [Prénom] va prendre l'avion j'ai dit non tu vois ça y est il vient de le prendre j'ai regardé l'heure j'y pense va je sais plus ce que je lui ai dit encore tiens cette nuit je vais t'arranger ça et ça à côté de toi et puis je le regarde et paf il s'est renversé en arrière et il était déjà mort alors ça c'est évidemment on n'y croit pas hein j'ai essayé de le réanimer puis le temps de bondir pour appeler un médecin pour appeler le SAMU ça ça revient presque tous les

jours voilà la mort est quelque chose de on n'y est pas habitué heureusement mais ça bouleverse toute une vie hein et des années après

Constance : alors maintenant on a fini les questions on passe à quelque chose de totalement différent donc la description d'image y a pas de piège faut juste me la décrire le plus précisément possible

FHb1 : donc je suppose que c'est une cuisine une maîtresse de maison elle essuie bien la vaisselle mais elle laisse déborder son évier et elle patauge dedans ce qui est un peu surprenant faudrait vraiment être aveugle pour ça des enfants qui en profitent pour aller chipper je pense des biscuits là-haut euh le gamin évidemment qui est en train de se casser la figure qui tombe de son escabeau et la petite sœur qui le pousse à en attraper un peu plus c'est une maison très simple parce que l'évier après y a plus rien sur le plan de travail on aperçoit dehors un jardin une maison voisine oui des rideaux au-dessus de l'évier moi j'aurais pas mis des rideaux comme ça hein au-dessus d'un évier surtout avec la tringle heu qu'est-ce qu'on peut dire encore on dirait qu'elle a des talons pour faire sa vaisselle oh bon si ça lui plaît euh oui y a des choses un peu discordantes et euh oui oui je crois que c'est tout ce que j'ai à dire oui

❖ **MBb1 : sujet sain, homme, niveau 2, 77 ans**

Elsa : voilà alors est-ce que vous pouvez me raconter le plus beau jour de votre vie

MBb1: le plus beau jour de ma vie ah alors là c'est bah c'est quand je m'suis marié le plus beau jour de ma vie oui c'est le plus beau jour de ma vie pour le moment euh j'en ai pas eu d'autres c'est un peu c'est un jour vraiment exceptionnel hein euh on s'en souvient tout le temps hein ça c'est sûr que mais enfin bon euh non ben je vois rien à vous dire qu'est-ce que vous voulez que je vous dise d'autre

Elsa: ben me raconter votre mariage comment

MBb1: mon mariage oui ben mon mariage bah y c'était j'ai connu mon épouse euh toute jeune et je travaillais chez son père donc euh toute jeune et bon je la connais elle était elle avait 4 ans attendez voir 32 35 3 ans de plus que moi elle allait encore à l'école et moi je travaillais j'avais 17 ans je la voyais aller à l'école encore mais jamais je pensais un jour me marier avec elle hein c'est pour ça et puis ça s'est fait comme ça on a vieilli elle a été elle a été bien malade elle avait pris la polyo puis le temps a passé et puis finalement euh petit à petit on s'est rapprochés et ses parents m'invitaient souvent chez eux et puis petit à petit bah on se on s'est aperçu qu'on qu'on s'aimait quoi c'est ça le le principe et puis bon on était jeune encore on a laissé passer du temps comme elle avait été bien malade elle euh il a fallu qu'elle se remette elle avait pris la polyo à 14 ans alors on a attendu elle s'est remis petit à petit et puis elle est arrivée à remarcher toute seule alors on s'est mariés en 57 je l'ai connue en 49 et on s'est mariés en 57 donc 8 ans après et puis voilà bon ça a été jamais j'aurais pensé que ça arrive quoi comme ça mais enfin bon c'est le destin c'était le destin c'est que ça devait se passer comme ça ouais et donc je suis devenu le j'ai épousé la fille du patron ça arrive encore oui enfin je sais pas si ça arrive encore hein mais remarque moi c'est arrivé pour moi alors c'était pas mal voilà

Elsa: alors maintenant est-ce que vous pouvez me raconter une dispute ou une bagarre qui vous est arrivée et qui vous a marqué

MBb1: une dispute

Elsa: ou une bagarre qui vous est arrivée que ce soit enfant maintenant

MBb1: hof non j'ai pas l'habitude de me dispu disputer euh ni de me bagarrer j'aimais pas ça quand j'étais petit heu même à l'école j'évitais les bagarres j'ai c'est pas qu'j'avais peur mais j'aimais pas j'aimais pas les bagarres même plus grand euh il est arrivé dans des sorties dans des bals euh des trucs comme ça y en a dès qu'y avait une bagarre ou des trucs comme ça ils se précipitaient et ils y allaient aussi moi au contraire je m'éclipsais du ça me regarde pas du moment que j'étais pas incombé je voulais pas me je voulais pas me mêler de ça non non j'aimais pas les bagarres et je me suis jamais bien disputé euh non je me suis jamais non j'aimais pas alors ni les disputes ni les bagarres ou alors si y a une dispute avec quelqu'un bon ça peut mais ça dure pas moi j'suis pas j'suis pas rancunier et si ça m'est arrivé de me brouiller avec quelqu'un moi j'ai toujours refait essayé de refaire le premier pas pour se réconcilier mais en principe j'ai toujours évité aussi les bagarres que les disputes

Elsa : et maintenant pouvez-vous me raconter votre plus mauvais souvenir le plus mauvais souvenir

MBb1 : ah le plus mauvais souvenir ben c'est quand mon épouse elle est tombée malade et euh (pleurs)

Elsa : on n'est pas obligé d'en parler

MBb1 : le plus mauvais souvenir si j'ai eu un mauvais souvenir avant hein je me suis cassé la jambe ça c'est un mauvais souvenir en 1967 bon ça c'est un mauvais souvenir c'est jamais agréable mais enfin ça se guérit ça mais le plus mauvais c'est vraiment quand mon épouse est tombée malade elle avait pris euh quelque chose au cervelet et heu quand le chirurgien l'a vue il avait dit au départ qu'elle était pas opérable y avait rien à faire qu'elle allait qu'elle allait certainement mourir alors ce jour là c'était vraiment un mauvais souvenir

Elsa : mais aujourd'hui elle est là elle est bien allez on passe à autre chose

MBb1 : voyez vous me faites

Elsa : on arrête si vous voulez

MBb1 : non non ça va maintenant c'est passé

Elsa : on passe totalement à autre chose on arrête les questions c'est fini je voudrais juste que vous me décriviez cette image comme si je la voyais pas

MBb1 : là c'est une femme qui fait la vaisselle et puis qui est un peu étourdie parce qu'elle laisse déborder son évier et puis les enfants qui s'amuse et puis qui sont prêts à tomber parce que aussi y en a un qu'est en déséquilibre voilà c'est une scène d'étourdissement d'étourdi quoi et dangereuse parce que ça peut devenir dangereux

Elsa : ok super c'est terminé

❖ **PFH3 : patiente atteinte de la MA, niveau 3, 87 ans, score MMS 19**

Etudiante : je vais vous demander de me raconter le plus beau jour de votre vie

PFH3 : il doit y en avoir plusieurs des des beaux souvenirs ouff écoutez sincèrement là alors hein vous oui oui il y a il y a en a beaucoup de bons souvenirs et puis je je risque d'en d'oublier le meilleur hein

Etudiante : vous m'avez parlé de l'Afrique tout à l'heure est-ce qu'il y a des bons souvenirs en Afrique vous voulez m'en raconter un

PFH3 : ah oui oui ben ma fois écoutez j'en ai j'ai l'Afrique noire et l'Afrique du Nord mais j'ai vécu en Afrique du Nord alors en l'Afrique noire c'est c'est mes enfants qui y étaient et je le le bon souvenir ça a été vraiment d'aller les rejoindre et de rester là-bas quelques temps

Etudiante : où est-ce qu'ils étaient vos enfants

PFH3 : [ville d'Afrique] [pays d'Afrique] voilà

Etudiante : et vous êtes allez les rejoindre vous partiez d'où vous d' [pays d'Afrique] vous

PFH3 : hé oui non je partais de [pays européen] là j'étais chez moi j'étais en [pays européen] et puis et puis puis voilà je suis allée rejoindre je suis restée mais euh à titre de vacances quoi si on peut dire oui alors ben j'ai tellement de bons souvenirs

Etudiante : et alors ce bon souvenir là-bas avec eux et vous êtes restée longtemps là-bas avec eux

PFH3 : 1 mois je suis restée 1 mois 1 mois 1 mois et demi peut-être oui voilà euh et puis puis puis je j'ai fait la mon gendre euh bon le la responsabilité de l'organisation de de comment pourrais-je vous dire il fallait euh qu'il s'occupe euh de comment ça de je ne trouve pas le mot là enfin de faire faire des ri des plantations des XXX des voilà il était il était in il est ingénieur ingénieur agricole oui

Etudiante : et vous vous avez passé un mois là bas avec eux

PFH3 : non non toute seule oui j'étais toute seule je n'avais pas mes enfants mes enfants étaient mariés déjà hé hein voilà oh j'ai des petits-enfants j'ai même j'ai euh arrières-petits-enfants à [ville de grande importance] alors euh ceux-là sont

Etudiante : et moi maintenant si je vous demandais de me raconter votre pire souvenir

PFH3 : oh mon pire souvenir ma vie a é n'a pas été bah le pire souvenir ça a été euh le peur une peur té oui mais ça c'est quelque chose quand même bon j'ai eu une fille qui a eu une méningite cé cérébro-spinale et qui était petite hein toute petite hein la peur imaginez-vous quand un médecin vous dit je ne sais pas parce que j'étais en [pays d'Afrique] à ce moment-là et et on avait pas tout ce qu'il fallait loin de là pour soigner un enfant alors j j là j'ai eu de la chance parce que mon mari étant militaire il a joint mon mari était militaire

en [pays européen] il était à cette époque-là et il a to pu joindre le colonel qui qui lui a passé mon mari et et il a dit voilà il me faut tel tel médi médicaments si vous sauvez la fille du commandant [Nom de famille] euh je m'appelais [Nom de famille] et dans les 24 heures j'ai mis les médicaments et je crois que le médecin était aussi euh surpris et content et et estomaqué et j'ai eu une fille qui qui est professeur de latin grec français vous voyez tout était bien tout c'était bien en f c'est un miracle

Etudiante : oui ça a été votre pire souvenir et votre plus beau souvenir à la fois

PFH3 : oui oui ça a été vraiment parce que j'y étais toute seule quoi en [pays d'Afrique] mon mari était militaire alors il était pas là ma famille naturellement était en [pays européen] aussi ça a été un un souvenir épouvantable parce que il faut entendre ce que c'est qu'un enfant de 7 mois qui hurle elle euh elle avait tellement mal dans sa tête que 7 mois ah oui ça j'ai gardé ça ça a été bon b et p elle est professeur de latin grec français donc voyez vous elle a fait son chemin mais ça été le pire des souvenirs parce que [Prénom] a été malade aussi ma fille aînée on a jamais cru qu'on la perdrait tandis que celle-là le médecin qui l'a soignée j'avais je ne suis pas particulièrement euh attachée à la religion mais ma maman m'avait envoyé une vierge de [ville de petite importance] bon alors inutile de vous dire que je l'avais sur euh la table de nuit la vierge de [ville de moyenne importance] et ce médecin était juif il s'est tourné vers cette vierge je lui dit écoutez oh je ff vous savez ça m'est resté ça si vous devez enlever cette enfant attendez qu'elle soit dans sa famille et ça ça m'a ça m'est resté et on me l'a pas enlevée euh professeur les professeurs de ci de là du reste de voilà mais j'ai passé un mauvais moment j'étais toute seule vous savez je me suis mariée à 18 ans alors voyez j'étais pas vieille hein mais voilà y a des miracles et je peux le dire qu'il y a eu un miracle

Etudiante : je vous montre une image je vous demande juste de me décrire ce que vous voyez

PFH3 : ce que je vois je vois un petit garçon pe perché sur un tabouret et qui heu qui est en train de prendre de fouguer dans un placard pour y prendre pour y prendre ce ce que qu'est ce qui est écrit dessus bateau non

Etudiante : des gâteaux

PFH3 : des gâteaux ha bah voilà oui je ne voyais pas oui bah pour y prendre des friandises la petite fille est en bas et attend que le garçon lui en passe quelques unes sans doute voilà

Etudiante : ensuite vous pouvez me dire quoi encore

PFH3 : euh toujours pareil ou de ce côté

Etudiante : toute l'image

PFH3 : ah toute l'image bah la maman ne voit pas ce qui se passe mais elle doit être consentante parce qu'elle entend bien que que les enfants sont là et et par contre elle a laissé le robinet de d'eau ouvert il me semble puisque là il y a eu débordement de l'évier et pourtant elle n'a pas l'air de s'en apercevoir oui oui bah elle essuie elle fait la vaisselle donc elle est essuie euh avec euh torchon elle a l'air de chanter elle a l'air de d'avoir euh

de chanter puis euh y a une fenêtre de avec des rideaux oui la fenêtre je ne vois pas ce qu'elle va faire là je sais pas ce que c'est

Etudiante : moi non plus

PFH3 : bon bah tant mieux comme ça écoutez bon bah elle regarde par la fenêtre en essuyant mon dieu mon dieu mon dieu je me ferais du souci moi si j'étais la fille voyez et elle le laisse elle a pas l'air de se préoccuper de l'évier qui déborde en tout cas puisqu'elle sourit bon il y a deux tasses vides sur le près de l'évier et une assiette également euh vide alors là je me méfierais hein le tabouret il est vraiment bien bien mal en coin j'ai l'impression qu'il va aller atterrir c c contre sa maman j'espère pour lui

Annexe V : Scores de densité des idées

❖ Population saine

Tranche d'âge	65-74 ans										75-85 ans								
	FEMININ					MASCULIN					FEMININ					MASCULIN			
Sexe	Age	Score grille Poitrenaud	Score de DI	Age	Score grille Poitrenaud	Score de DI	Age	Score grille Poitrenaud	Score de DI	Age	Score grille Poitrenaud	Score de DI	Age	Score grille Poitrenaud	Score de DI	Age	Score grille Poitrenaud	Score de DI	
Haut niveau socio culturel	FHa1	70	1	0.447	MHa1	68	4	0.439	FHb1	76	4	0.441	MHb1	83	4	0.423			
	FHa2	73	1	0.416	MHa2	74	4	0.414	FHb2	79	4	0.413	MHb2	79	4	0.450			
	FHa3	69	1	0.484	MHa3	65	4	0.378	FHb3	76	4	0.451	MHb3	78	4	0.401			
	FHa4	72	1	0.427	MHa4	73	4	0.491	FHb4	82	4	0.425	MHb4	83	4	0.451			
	FHa5	68	1	0.416	MHa5	73	4	0.469	FHb5	75	4	0.418	MHb5	83	4	0.448			
Bas niveau socio culturel	FBa1	73	2	0.404	MBa1	67	2	0.374	FBb1	78	1	0.423	MBb1	77	2	0.398			
	FBa2	73	2	0.340	MBa2	65	2	0.334	FBb2	76	2	0.409	MBb2	80	1	0.405			
	FBa3	74	2	0.393	MBa3	70	2	0.424	FBb3	82	2	0.406	MBb3	85	2	0.393			
	FBa4	72	2	0.411	MBa4	74	2	0.355	FBb4	80	1	0.361	MBb4	81	2	0.405			
	FBa5	66	2	0.413	MBa5	65	1	0.395	FBb5	81	2	0.415	MBb5	76	2	0.355			

❖ Patients atteints de la maladie d'Alzheimer et sujets sains appariés

Patients atteints MA	Age	Score grille Poitrenaud	Score MMS	Score de DI	Sains appariés	Score de DI
PFH1	68	4	26	0.351	FHa2	0.416
PFH2	67	4	25	0.409	FHa3	0.484
PFH3	87	3	19	0.299	FHb4	0.425
PFH4	88	4	21	0.420	MHb4	0.451
PFB1	82	2	24	0.385	FBb3	0.406
PFB2	77	2	22	0.402	FBb2	0.409
PFB3	81	1	20	0.342	FBb4	0.361
PFB4	72	2	20	0.373	FBa3	0.393
PMB1	79	2	20	0.289	MBb4	0.405
PMB2	71	2	23	0.375	MBa4	0.355
PMB3	89	2	18	0.408	MBb3	0.393

❖ Discours narratif versus discours descriptif

	Score de DI : Discours descriptif	Score de DI : Discours narratif
FHa4	0.344	0.425
FBa1	0.399	0.369
MHa5	0.390	0.483
Mba2	0.313	0.284
FHb1	0.412	0.440
FBb4	0.364	0.290
MHb1	0.338	0.437
MBb2	0.361	0.428

TABLE DES ILLUSTRATIONS

1. Liste des Tableaux

Tableau 1: Répartition de la population en fonction de l'âge, du sexe et du niveau socioculturel	30
Tableau 2: Composition des populations de sujets sains et des patients atteints de la maladie d'Alzheimer	31
Tableau 3 : Statistiques de groupe : moyenne des scores de densité des idées obtenus pour chaque groupe d'âge.....	41
Tableau 4: Rang moyen de chaque tranche d'âge.....	42
Tableau 5 : Résultats obtenus au test ^b pour les deux groupes d'âge.....	42
Tableau 6 : Statistiques de groupe : moyenne des scores de densité des idées obtenus pour chaque sexe	43
Tableau 7 : Rang moyen de chaque sexe	43
Tableau 8 : Résultats obtenus au test ^b pour les deux sexes	44
Tableau 9 : Statistiques de groupe : moyenne des scores de densité des idées obtenus pour chaque niveau socioculturel	44
Tableau 10 : Rang moyen de chaque niveau socioculturel	45
Tableau 11 : Résultats obtenus au test ^b pour les deux niveaux socioculturels.....	45
Tableau 12 : Statistiques descriptives: moyenne des scores de densité des idées obtenus pour le groupe « sain » et le groupe de personnes atteintes de la maladie d'Alzheimer	46
Tableau 13 : Rang moyen du groupe « sain » et du groupe de personnes atteintes de la maladie d'Alzheimer	47
Tableau 14: Résultats obtenus au test ^b pour le groupe « sain » et le groupe de personnes atteintes de la maladie d'Alzheimer	48
Tableau 15 : Statistiques descriptives : moyenne des scores de densité des idées obtenus pour les discours descriptif et narratif.....	49

TABLE DES ILLUSTRATIONS

Tableau 16: Rangs négatifs et positifs moyens pour les discours narratif et descriptif 49

Tableau 17 : Résultats obtenus au test^b pour les discours narratif et descriptif 50

2. Liste des Figures

Figure 1 : Exemples de règles de Densidées 1.2 36

Figure 2: L'interface graphique de Densidées sous windows 39

Figure 3 : Score moyen de densité des idées en fonction de l'âge 42

Figure 4 : Score moyen de densité des idées en fonction du sexe 43

Figure 5 : Score moyen de densité des idées en fonction du niveau socioculturel 45

Figure 6 : Scores de densités des idées obtenus par les 11 sujets atteints de la maladie d'Alzheimer et leurs sujets sains appariés 47

Figure 7 : Score moyen de densité des idées en fonction du type de discours 49

Figure 8 : Score moyen de densité des idées en fonction du type de discours produit et du niveau socioculturel 50

Figure 9 : Score moyen de densité des idées en fonction de l'âge et du niveau socioculturel 51

Figure 10 : Score moyen de densité des idées chez les sujets sains et chez les sujets atteints de la maladie d'Alzheimer 51

TABLE DES MATIERES

ORGANIGRAMMES	2
1. <i>Université Claude Bernard Lyon I</i>	2
1.1. Secteur Santé :	2
1.2. Secteur Sciences et Technologies :	2
2. <i>Institut Sciences et Techniques de Réadaptation FORMATION ORTHOPHONIE</i>	3
REMERCIEMENTS	4
SOMMAIRE	5
INTRODUCTION	7
PARTIE THEORIQUE	9
II. CONSEQUENCES DU VIEILLISSEMENT COGNITIF SUR LE LANGAGE.....	10
1. <i>Le vieillissement cognitif normal</i>	10
1.1. Définition.....	10
1.2. Processus cognitifs fragilisés	10
2. <i>Retentissements du vieillissement cognitif sur le langage</i>	11
2.1. Au niveau lexical	11
2.2. Au niveau morphosyntaxique	11
2.3. Au niveau textuel	12
3. <i>Facteurs influençant l'évolution du langage</i>	12
4. <i>Prise en compte du niveau socioculturel dans l'étude de langage</i>	12
III. CONSEQUENCES DU VIEILLISSEMENT COGNITIF PATHOLOGIQUE SUR LE LANGAGE : LA MALADIE D'ALZHEIMER	13
1. <i>La maladie d'Alzheimer</i>	13
1.1. Définition.....	13
1.2. Physiopathologie.....	14
1.3. Facteurs de risque	14
1.4. Systèmes de mémoires altérés par la maladie d'Alzheimer	14
2. <i>Conséquences de la maladie d'Alzheimer sur le langage</i>	15
IV. LA DENSITE DES IDEES	16
1. <i>Définition</i>	16
2. <i>Paramètres pouvant influencer le score de densité des idées</i>	16
3. <i>Evaluation de la densité des idées dans le discours</i>	17
3.1. Le discours.....	17
3.2. Les mécanismes impliqués dans la production du discours	17
3.3. Le discours oral spontané.....	18
3.3.1. Intérêts du discours oral spontané pour l'étude de la densité des idées.....	18
3.3.2. Type de discours et production orale.....	19
a. Le discours narratif.....	19
b. Le discours descriptif.....	19
4. <i>Mesure de la densité des idées</i>	19
4.1. L'analyse prédicative.....	19
4.1.1. Les fondements de l'analyse prédicative	20
4.1.2. Exemples de modèles d'analyse prédicative.....	21
4.2. Application à la densité des idées	22
4.3. Les limites de l'analyse prédicative	22
4.3.1. Limites de ces modèles pour une application à l'oral	22
4.3.2. Limites liées à l'analyste	23
5. <i>Un exemple d'analyse automatique de l'anglais, Computerized Propositional Idea Density Rater : CPIDR 3.2 (Brown, Snodgrass, Kemper, Herman, & Covington, 2008)</i>	23
5.1. Présentation du logiciel.....	23
5.2. Limite de ce logiciel pour notre étude.....	24
PROBLEMATIQUE ET HYPOTHESES	25
I. PROBLEMATIQUE	26
II. HYPOTHESES.....	27
1. <i>Hypothèses générales</i>	27
2. <i>Hypothèses opérationnelles</i>	27

TABLE DES MATIERES

PARTIE EXPERIMENTALE	28
I. POPULATION	29
1. <i>Sujets sains</i>	29
1.1. Méthode de recrutement	29
1.2. Répartition de la population	29
1.2.1. Critères d'inclusion	29
1.2.2. Critères d'exclusion	30
2. <i>Sujets atteints de la maladie d'Alzheimer</i>	31
II. PROTOCOLE EXPERIMENTAL	31
1. <i>Objectif</i>	31
2. <i>Elaboration du protocole expérimental</i>	31
2.1. Situations de communication sélectionnées	31
2.2. Choix des thèmes de production	32
2.2.1. Choix des questions	32
2.2.2. La description d'image	33
3. <i>Protocole de passation</i>	33
4. <i>Matériel</i>	34
III. CONDITIONS D'EXPERIMENTATION	34
1. <i>Passation</i>	34
1.1. Sujets sains	34
1.2. Sujets atteints de la maladie d'Alzheimer	34
2. <i>Lieu d'expérimentation</i>	35
3. <i>Durée de la passation</i>	35
IV. ANALYSE DES DONNEES	35
1. <i>Création du logiciel d'analyse automatique de la densité des idées : Densidées</i>	35
1.1. Extraction des règles d'identification des prédicats	36
1.2. Création de deux règles particulières à l'analyse de l'oral	36
1.3. Fiabilité et validité du logiciel	37
2. <i>Procédure d'analyse de nos corpus</i>	38
PRESENTATION DES RESULTATS	40
I. ETUDE DE LA POPULATION GENERALE	41
1. <i>Méthodologie statistique</i>	41
2. <i>Effet de l'âge</i>	41
3. <i>Effet du sexe</i>	43
4. <i>Effet du niveau socioculturel</i>	44
II. ETUDE DE L'ECHANTILLON DE PATIENTS ATTEINTS DE LA MALADIE D'ALZHEIMER	46
1. <i>Méthodologie</i>	46
2. <i>Effet de la pathologie</i>	46
III. ETUDE DE L'EFFET DU TYPE DE DISCOURS SUR LA DENSITE DES IDEES	48
1. <i>Méthodologie</i>	48
2. <i>Impact du type de discours sur la DI</i>	48
3. <i>Impact du niveau socioculturel sur les scores de densité des idées obtenus pour chaque type de discours</i>	50
IV. SYNTHESE DES RESULTATS	51
DISCUSSION DES RESULTATS	53
I. VALIDATION DES HYPOTHESES DE TRAVAIL	54
1. <i>Effet de l'âge</i>	54
2. <i>Effet du sexe</i>	56
3. <i>Effet du niveau socioculturel</i>	56
4. <i>Effet de la pathologie</i>	57
5. <i>Effet du type de discours</i>	58
5.1. Impact du type de discours sur la DI	58
5.2. Effet du niveau socioculturel sur les scores de densité des idées obtenus pour chaque type de discours	59
II. LIMITES ET INTERETS DE CETTE ETUDE	60
1. <i>Limites</i>	60
1.1. Entretien avec les patients atteints de la maladie d'Alzheimer	60
1.2. Sélection de la population saine	60
1.3. Protocole de passation	61
1.4. Analyse des données	62

TABLE DES MATIERES

1.4.1. Choix de l'extrait étudié	62
1.4.2. Aspect qualitatif.....	62
1.4.3. Limites inhérentes à l'étude de la densité des idées.....	63
2. <i>Intérêts</i>	63
2.1. Intérêts du protocole expérimental.....	63
2.2. Intérêts de l'étude.....	63
III. APPORT CLINIQUE.....	64
IV. PERSPECTIVES ET FAISABILITE.....	65
1. <i>Perspectives cliniques</i>	65
2. <i>Autres perspectives</i>	66
CONCLUSION.....	67
BIBLIOGRAPHIE.....	69
ANNEXES	75
ANNEXE I : GRILLE DE POITRENAUD (KALAFAT, HUGONOT-DIENER, & POITRENAUD, 2005)	76
ANNEXE II : FICHE DE RENSEIGNEMENTS	77
ANNEXE III : DENSIDEES - CALCUL AUTOMATIQUE DE LA DENSITE DES IDEES DANS UN CORPUS ORAL	78
ANNEXE IV : EXEMPLES DE CORPUS	88
❖ FHb1 : sujet sain, femme, niveau 4, 76 ans	88
❖ MBb1 : sujet sain, homme, niveau 2, 77 ans	89
❖ PFH3 : patiente atteinte de la MA, niveau 3, 87 ans, score MMS 19	91
ANNEXE V : SCORES DE DENSITE DES IDEES	94
❖ Population saine	94
❖ Patients atteints de la maladie d'Alzheimer et sujets sains appariés.....	95
❖ Discours narratif versus discours descriptif.....	96
TABLE DES ILLUSTRATIONS.....	97
1. <i>Liste des Tableaux</i>	97
2. <i>Liste des Figures</i>	98
TABLE DES MATIERES	99

Elsa Maillé

Constance Thuillier

MESURE DE LA DENSITÉ DES IDÉES DANS LE VIEILLISSEMENT COGNITIF NORMAL EN FONCTION DU NIVEAU SOCIOCULTUREL ET DE L'ÂGE : Incidences sur l'analyse des productions orales de patients atteints de la maladie d'Alzheimer.

101 Pages

Mémoire d'orthophonie -UCBL-ISTR- Lyon 2010

RESUME

Des études récentes ont recensé un certain nombre d'indices linguistiques précocement altérés dans le cadre de la maladie d'Alzheimer. Parmi ces critères figure la densité des idées. Elle permet de mesurer la qualité informative du langage. Une densité des idées élevée peut refléter l'aptitude d'un locuteur à exprimer efficacement ses idées. Une densité des idées faible peut, quant à elle, révéler un discours peu efficace par l'utilisation d'un plus grand nombre de mots pour exprimer les idées essentielles. La présente étude a pour but d'évaluer la pertinence de ce critère pour distinguer des sujets présentant un vieillissement cognitif normal de ceux présentant une évolution pathologique liée à la maladie d'Alzheimer. Pour ce faire une normalisation a été réalisée. Des mesures de densité des idées de quarante sujets ont été réalisées en fonction de l'âge et du niveau socioculturel. L'effet de la maladie d'Alzheimer sur la densité des idées a ensuite été mesuré chez onze patients. L'expérimentation consiste à recueillir des discours oraux auxquels on attribue un score de densité des idées. Ce score est calculé à l'aide d'un logiciel d'analyse automatique : Densidées. Il se base sur le modèle formel de l'analyse prédictive. Les résultats obtenus attestent que la densité des idées n'est pas affectée par le vieillissement cognitif normal. Cependant, elle est corrélée au niveau socioculturel : les personnes de haut niveau présentent une densité des idées plus élevée que celles de bas niveau. De plus, la densité des idées des patients atteints de la maladie d'Alzheimer est altérée. Elle apparaît dès lors comme un indicateur pertinent de la maladie d'Alzheimer. Elle révèle une altération des capacités cognitives. La densité des idées est donc un champ d'investigation prometteur pour aider au diagnostic et pour mesurer l'avancée de la maladie. Par ailleurs, il serait intéressant de l'étudier pour d'autres pathologies neurodégénératives.

MOTS-CLES

Vieillessement cognitif, maladie d'Alzheimer, analyse du discours, Densité des idées, Niveau socioculturel, Traitement automatique du langage.

MEMBRES DU JURY

Emmanuelle Aujogues – Frédérique Gayraud – Anne Peillon

MAITRE DE MEMOIRE

Mélissa Barkat-Defradas

DATE DE SOUTENANCE

1 juillet 2010
