



HAL
open science

Utilisation de l'analyse textuelle automatique dans la recherche sur la maladie d'Alzheimer

Hyeran Lee, Philippe Gambette, Melissa Barkat-Defradas

► **To cite this version:**

Hyeran Lee, Philippe Gambette, Melissa Barkat-Defradas. Utilisation de l'analyse textuelle automatique dans la recherche sur la maladie d'Alzheimer. Colloque international des jeunes chercheurs en Didactique des Langues et en Linguistique, Jun 2010, Grenoble, France. halshs-00778744

HAL Id: halshs-00778744

<https://shs.hal.science/halshs-00778744v1>

Submitted on 11 Apr 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

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

Utilisation de l'analyse textuelle automatique dans la recherche sur la maladie d'Alzheimer

Hyeran Lee¹, Philippe Gambette², Melissa Barkat-Defradas¹

¹ Praxiling - Université Montpellier 3 - CNRS / ² LIRMM - Université Montpellier 2 – CNRS

Introduction

Avec le vieillissement de la population, l'incidence de la maladie la maladie d'Alzheimer (MA) augmente de façon exponentielle. En l'absence de diagnostic fiable et de traitements curatifs efficaces, la recherche visant à combattre cette pathologie devient un véritable enjeu de santé publique.

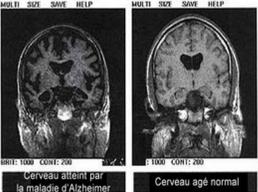


Figure 1. Comparaison des Images d'IRM cérébrale d'un sujet sain et d'un malade atteint d'Alzheimer
Illustration : NASA

Démence
=
Affaiblissement progressif & global des fonctions cognitives

L'altération de la capacité langagière refléterait les troubles cognitifs sous-jacents (e.g. le déficit de la mémoire, de l'attention, et des fonctions exécutives) et les zones de lésion.

$$DI = \frac{\text{nombre de propositions sémantiques (prédicat + arguments)}}{\text{nombre total de mots}} \times 10$$

- Permet d'évaluer l'aptitude d'un locuteur à exprimer efficacement les idées ou les informations
- Utilisée dans divers domaines de la psycholinguistique appliquée :

- ✓ La compréhension du texte (Kintsch et al., 1973)
- ✓ La mémoire (Thorson et al., 1984)
- ✓ Le vieillissement (Kemper et al., 2001)
- ✓ La schizophrénie (Covington et al., 2007)
- ✓ La qualité de prise de notes des étudiants (Takao et al., 2002)
- ✓ La maladie d'Alzheimer (Snowdon et al., 1996; Lee et al., 2010)

Objectifs

- Développer un outil automatique permettant de donner un résultat rapide & objectif de la DI pour une perspective de mise en pratique
- Comparer le score de DI du groupe de patients Alzheimer et du groupe de contrôles pour évaluer la pertinence de cette mesure comme outil de diagnostic différentiel entre vieillissement cognitif normal vs. pathologique
- Analyser l'impact des variables socioculturelles sur la performance langagière

Méthode

Sujets

- ✓ 20 sujets cliniquement diagnostiqués comme atteints de la maladie d'Alzheimer
- ✓ 20 sujets ne présentant pas de troubles cognitifs
- ✓ Appariés par âge, sexe, et niveau socioculturel (grille de Poitrenaud)

	Groupe MA	Groupe de contrôle
Sexe	11 F + 9 H	11 F + 9 H
Âge	M= 75.74 (± 6.1)	M= 76 (± 5.78)
NSC	M= 2.45 (± 1)	M= 2.45 (± 1)
MMSE	M= 21.2 (± 4.81)	M= 30

Tableau 1. Profil des sujets

Procédure

Recueil des données

- Entretien individuel semi-dirigé
- Évocation d'un souvenir personnel
- Enregistrement

Constitution des données

- Transcription
- 350 mots chacune (comparaison statistique)

Prétraitement manuel

- Marquage de certaines caractéristiques spécifiques à l'oral inexploitable d'un point de vue automatique

Analyse automatique

- Étiquetage morphologique : *TreeTagger*
- Calcul de DI : *Densidées**

* Densidées :

- Adaptation française du logiciel CPIDR (Brown et al., 2008)
- Idée principale = un prédicat correspond typiquement à un verbe (prédicateur), à un adjectif, à un adverbe (modificateurs), à une préposition, ou à une conjonction (connecteurs)
- Étiquetage morphosyntaxique = la base du calcul approximatif de la densité des idées
- Règles d'ajustement (e.g. gestion basique de certaines répétitions) pour le discours oral

« c'est le jour de mon mariage (voilà) il y a 52 ans bientôt (donc voilà) bien que ça a été un mariage tout à fait simple »

- P1. COPULE (X, a1) a1= jour
- P2. DE (a1, a2) a2= mariage
- P3. MON (a2)
- P4. IL Y A (a3) a3= ans
- P5. 52 (a3)
- P6. BIENTÔT (P4)
- P7. COPULE (P3, a2)
- P8. SIMPLE (a2)
- P9. TOUT A FAIT (P8)
- P10. BIEN QUE (P1, P8)

$$DI = 10 / 27 \times 10 = 3.7$$

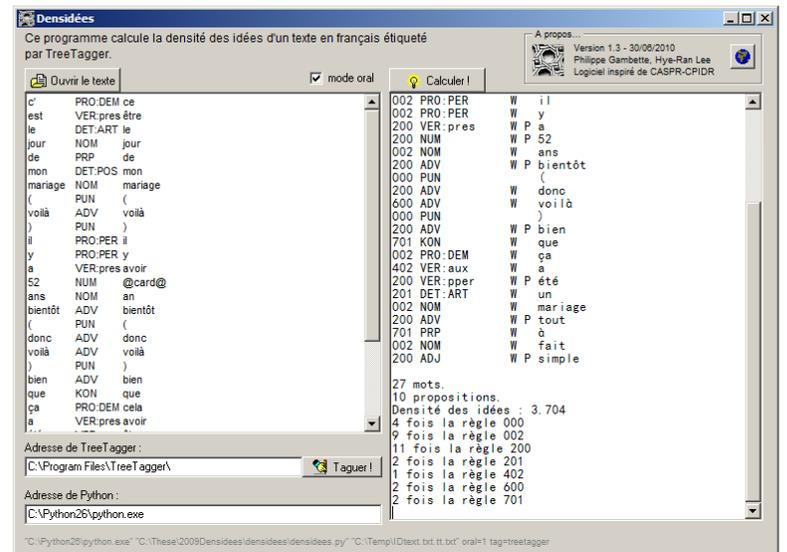


Figure 2. Interface graphique de Densidées sous Windows

Résultats

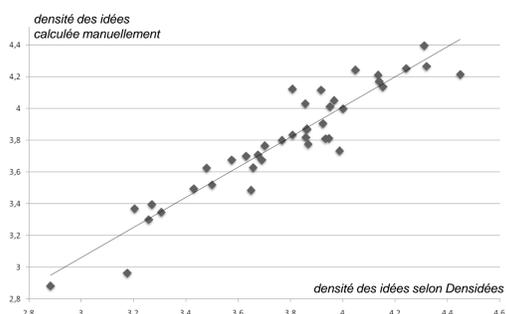


Figure 3. Représentation de densité des idées calculée manuellement en fonction de la densité des idées calculée automatiquement pour les 40 textes du corpus

- Coefficient de corrélation = 0.941 : nette corrélation entre analyse manuelle et analyse automatique

	Densidées	Manuelle
moyenne rangs sains	16,55	16,3
moyenne rangs malades	24,45	24,7
W de Wilcoxon	121	116
p-value	0,03264	0,02272

Figure 5. Résultat récapitulatif du test de Wilcoxon avec R

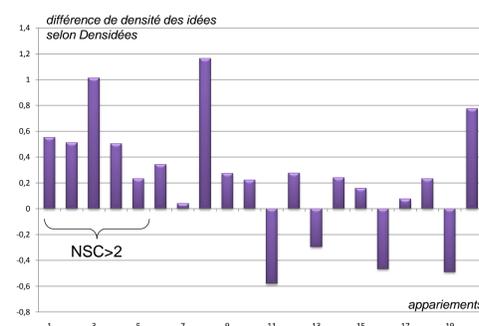


Figure 4. Différence de densité des idées, par appariement, entre les sujets contrôles et les sujets MA

- Forte tendance à la baisse du score de DI chez les patients MA
- Tendance non systématique = préservation de la DI chez certains patients

- Les deux méthodes montrent un effet significatif de la maladie d'Alzheimer sur le score de DI : déclin de la DI
- La p-value est légèrement meilleure avec l'analyse manuelle qu'avec Densidées

Conclusion

- Densidées permet une mesure fiable de la densité des idées.
- La baisse de la DI est associée à l'apparition de la maladie d'Alzheimer.
- Les différentes variables (âge, sexe, niveau socioculturel) n'ont pas d'influence significative sur le score de DI.
- Les résultats montrent l'utilité potentielle de la mesure automatique de la DI pour le diagnostic différentiel de la Maladie d'Alzheimer.

Références

- 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.
- 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.
- 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., 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.
- Lee, H., Barkat-Defradas, M. (2010). La densité des idées : une mesure pertinente de la dégradation linguistique chez les patients Alzheimer. *28^{ème} Journées d'étude sur la parole*, Mons, du 25 au 28 mai 2010.
- 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.
- Thorson E., Snyder R. (1984). Viewer recall of television commercials: prediction from the propositional structure of commercial scripts. *Journal of Marketing Research*, 21(2), 127-136.