NLP-Oriented Contrastive Study of Linguistic Productions of Alzheimer's And Control People

Maïté Boyé¹, Thi Mai Tran^{1,2}, Natalia Grabar²

¹ Institut d'Orthophonie, Université Lille 2, Lille, France ² CNRS UMR8163 STL, Université Lille 3, France maite.boye@gmail.com, thimai.tran@univ-lille2.fr, natalia.grabar@univ-lille3.fr

POLTAL 2014, Warsaw, Poland, September 2014

Contrastive Study of Linguistic Productions of Alzheimer's

M Boyé, TM Tran, N Grabar

イロト イポト イラト イラト 一日

Plan

- Context
- 2 Material
- Methods
- esults and Discussion
- Onclusion and Future work

э

イロト イポト イヨト イヨト

Introduction

- Alzheimer's disease (AD):
 - related to the aging of the population
 - first cause of neurodegenerative dementia
 - 26 million people concerned in 2005
 - over 100 million people estimated in worldwide by 2050
- Progressive appearance of cognitive, emotional and behavior troubles
 - $\rightarrow\,$ loss of autonomy and dependency (demential stage of the disease)
- First clinical cognitive signs:
 - loss of language and memory, executive function troubles
- Evolution:
 - communication troubles become irreversible
 - severely impact daily life
- One of the main objectives of the AD therapy:
 - preservation of communication ability as long as possible

3

Existing work

Language and aging

- Studies on healthy eldery people:
 - language performance depends on education and age
 - but compared to the other cognitive functions (i.e. memory, executive functions), language is less affected by aging
 - difficulty in finding more frequent words
 - ambiguous references and redundancies in the discourse
 - simplification of syntax
 - frequent references to the past, little attention to the interlocutors
 - topics respected but frequent digressions

M Boyé, TM Tran, N Grabar

- (同) (回) (回) - 回

Existing work

Language and AD

- At early stages of the AD disease:
 - memory disorders with relative preservation of communication
 - first language deficits:
 - lexico-semantic disorders affecting first the production: word finding difficulties, semantic paraphasia
 - preservation of phonological and morpho-syntactic skills
- With the evolution of AD:
 - increase of cognitive disorders
 - language deficit:
 - affect language production and understanding
 - affect all the levels of language (phonology, syntax...)
 - decrease in language competence and in cognitive functions
 - lead to muteness and progressive social isolation
- Major challenge: maintain the communication

イロト 不得 トイヨト イヨト 二日

Our study

- Our objectives:
 - contrastive study of verbal conversation of the AD and healthy patients
 - additional diagnostic and therapeutic elements:
 - define specific criteria of the conversational AD language
 - distinguish deficient and preserved elements
- Our data (transcribed spoken corpora):
 - coupled AD (5, early and moderate stage) and healthy participants (5)
 - over 80 years old, average = 90 years
 - collected in ecological and non-artificial context
 - known interlocutor, non-supervised conversation
 - average conversation = 20 min
 - over 100 hours for the transcription
- Our tools:
 - The corpus is processed with Natural Language Processing (NLP) tools and through manual analysis

Contrastive Study of Linguistic Productions of Alzheimer's

M Boyé, TM Tran, N Grabar

Two aspects analyzed

- Verbal interaction:
 - turns of speech
 - time of speech and overlapping
- Onversation content:
 - Spoken features : breath groups, empty pauses, non-empty pauses, primes, repetitions and stutters, self-corrections, interruptions, verbal output
 - Lexical features : number of words, informativity, yes/no enunciations, lexical diversity, ratio lemmas/total words, morphological complexity, lexical frequency
 - Syntactic features : average length of enunciations, interpolated clauses and reported speech, personal pronouns, verbs, distribution within syntactic categories

イロト 不得 とくほ とくほ とうほう

M Boyé, TM Tran, N Grabar

- Solution Transcriber software (Barras et al., 1998)
- In the second second
 - Treetagger: POS tagging, lemmatization
 - Dérif: morphological analysis

Contrastive Study of Linguistic Productions of Alzheimer's

Results

Features	AD	Control	%
Verbal interaction			
Turns of speech (avg.)	194.80	181.80	-7
Time of speech (min.)	7'56"	11'08"	-46
Time of speech (max.)	13'11"	17'24''	-31
Time of speech (avg.)	11'16"	14'12"	-26
Overlapping (avg.)	0.59	1.01	-71
Spoken features			
Breath groups (avg. words)	3.90	4.31	-10
Empty pauses (avg.)	65.00	31.00	52
Non-empty pauses (avg.)	13.00	34.60	-166
Disfluencies (avg.)	30.00	37.00	-24
Primes of words (avg.)	2.8	8.4	-200
Interrupted sentences (avg.)	18.8	19.40	-3
Speech output (max. words/min.)	184.23	> 190	-3

Contrastive Study of Linguistic Productions of Alzheimer's

M Boyé, TM Tran, N Grabar

æ

・ロト ・回ト ・モト ・モト

Results

Features	AD	Control	%
Lexical features			
Number of words (avg.)	1,778.4	2,407.4	-35
Informativity (avg.)	9	8	11
Yes/No utterances (avg.)	88.80	52.40	40
Lexical diversity (avg.)	168.60	302.40	-79
Lemmas/total words ratio (avg. %)	10	14	-32
Morphological complexity	1	1.07	-3
Lexical frequency in corpora (avg.)	11	8	25
Lexical frequency on the web (avg.)	304 M	227 M	25

Contrastive Study of Linguistic Productions of Alzheimer's

M Boyé, TM Tran, N Grabar

æ

・ロト ・回ト ・モト ・モト

Results

Features	AD	Control	%
Syntactic features			
Length of utterances (avg.)	4.79	7.26	-51
Interpolated clauses (avg.)	1.20	4.80	-300
Reported speech (avg.)	1.60	14.20	-787
Personal pronouns (avg. %)	36	23	37
Pronoun/noun ratio	1	0.65	6
Verbs (avg.)	377	553	-46
Ratio of verbs	32	32	0
Ratio of nouns	51	55	-5
Ratio of adjectives	16	14	14

Contrastive Study of Linguistic Productions of Alzheimer's

M Boyé, TM Tran, N Grabar

◆□ > ◆□ > ◆臣 > ◆臣 > ─ 臣

Specificities of AD group

AD patients:

- produce less words for the same speaking time
- show higher number of turns of speech
- speak slowly, have a poorer lexicon and produce shorter sentences
- have little number of stutters, self-corrections and incomplete sentences
- often use personal pronouns like *je* (1) and produce important number of *yes* and *no* statements
- produce more empty pauses and less non-empty pauses
- seldom use reported speech and interpolated clauses

M Boyé, TM Tran, N Grabar

◆□ ▶ ◆□ ▶ ◆ □ ▶ ◆ □ ● ● ● ● ● ●

Comparison with healthy participants

- No notable difference between the 2 groups:
 - number of turns of speech and of breath groups
 - average speech output
 - distribution of disfluencies
 - distribution of lemmas within syntactic categories
 - ratio between pronouns and nouns
 - informativity and morphological complexity of words
- Nevertheless:
 - AD discourse is less fluent and rich than the discourse of healthy people
 - Several aspects that mark the spontaneous and natural speech of the healthy discourse are missing:
 - disfluencies, reported speech and interpolated clauses

イロト 不得 トイヨト イヨト 二日

Discussion

- At early stage of the AD:
 - some aspects of conversation language are preserved
 - remain similar to those observed with healthy people
- AD patients do not have the same linguistic capacity than healthy people:
 - in the context of free conversation
 - people are communicating
 - decrease of the enunciation length, reduced quantity and diversity of vocabulary, and of speech output
- AD patients find it difficult to communicate in natural situations:
 - they can hardly take the initiative in verbal exchanges, produce complex enunciations, and maintain the conversation
 - often, AD patients are helped for this by the existing language automatisms elaborated lifelong and preserved
 - despite the loss of informativity, spontaneity and dynamics of language exchange

Limitations of the study

- Participants: the results are to be taken with precaution:
 - small number of participants:
 - remains difficult the collect the data
 - variability within each group:
 - personal difference of participants
 - their habits and lifestyles
 - ightarrow must be completed with additional data
- Features exploited:
 - More features
 - Combination of features
- NLP tools:
 - AD discourse allows the use of NLP tools:
 - fluency, preserved phonological and syntactic levels

- 4 同 2 4 日 2 4 日 2 - 日

M Boyé, TM Tran, N Grabar

- improvement of the NLP tools
 - morphological and syntactic analysis
- Contrastive measure:
 - statistical significance
 - machine learning

Contrastive Study of Linguistic Productions of Alzheimer's

Conclusion

- Comparative analysis between AD and healthy participants
- Spoken transcribed conversations
- Coupled participants
- Natural conditions for the collection of linguistic data
- AD participants (early stage of the disease):
 - some conversation aspects remain intact
 - other conversation aspects begin to decrease
- Exploitation of these features for:
 - diagnosis of the disease
 - maintenance of the conversation capacity

M Boyé, TM Tran, N Grabar

・ 同 ト ・ ヨ ト ・ ヨ ト

Future work

- Study larger corpora collected with additional participants within similar ecological context
- Adaptation of some NLP tools to the processing of the data we use
 - enrich morphological analysis with other morphological rules
- Use of additional tools (eg, syntactic analysis)
- Complementary study of verbal and non-verbal (gesture) communication
- Similar work in other languages
- Automatic categorization of speakers as AD-suffering or not
- Automatic diagnosis of this disease at early stages
- Use of these data for the rehabilitation

◆□ ▶ ◆□ ▶ ◆ □ ▶ ◆ □ ● ● ● ● ● ●