VCU School of Engineering

VIRGINIA COMMONWEALTH UNIVERSITY

Automated Linguistic Analysis of Patients with Dementia and Mild Cognitive Impairment

Korcovelos, E.A.[1], Pakhomov, S.V.[2], and McInnes, B.T.[1],

[1]Virginia Commonwealth University, Department of Computer Science and [2]University of Minnesota, College of Pharmacy

Introduction

Dementia is a neurodegenerative disease that results in a gradual decline in cognition and memory. Recent reports on dementia suggest 23 in 100 people aged 60-84 currently possess the disease, whereby the prevalence is expected to double every five years. Dementia is known to affect cognition and behavior in the elderly and is comprised of both behavioral and language impairments. These language impairments heavily impact communication, with many individuals experiencing declines in expressive and receptive language abilities. However, it is important to note that individuals with dementia often have difficulties with the transmission of words in written and spoken communication rather than a deficit in word knowledge and word meanings. For individuals with Mild Cognitive Impairment (MCI), a syndrome that lies between normal cognitive aging and dementia, there is a high risk of developing dementia in the following months and years. Language variants between those with MCI and dementia have not been heavily studied computationally. For this reason, there is a critical need to develop algorithms and tools that are able to assess speech in ways that would lead to earlier diagnoses and treatment. **The present study aims to elucidate the psycholinguistic properties of dementia and MCI patients, with particular focus on speech complexity through natural language processing techniques.**

Methods

DementiaBank's Cookie Theft Test to determine lexical variance

Metric of focus: *type to token ratio* (unigram to 5-gram models)
Hypothesis: Individuals with dementia or MCI will demonstrate more repetitive speech (little lexical variation) through smaller type:token ratios.

Metrics of focus: *information content*
Hypothesis: Individuals with dementia or MCI will use more indefinite nouns and verbs when describing the Cookie Theft Test.
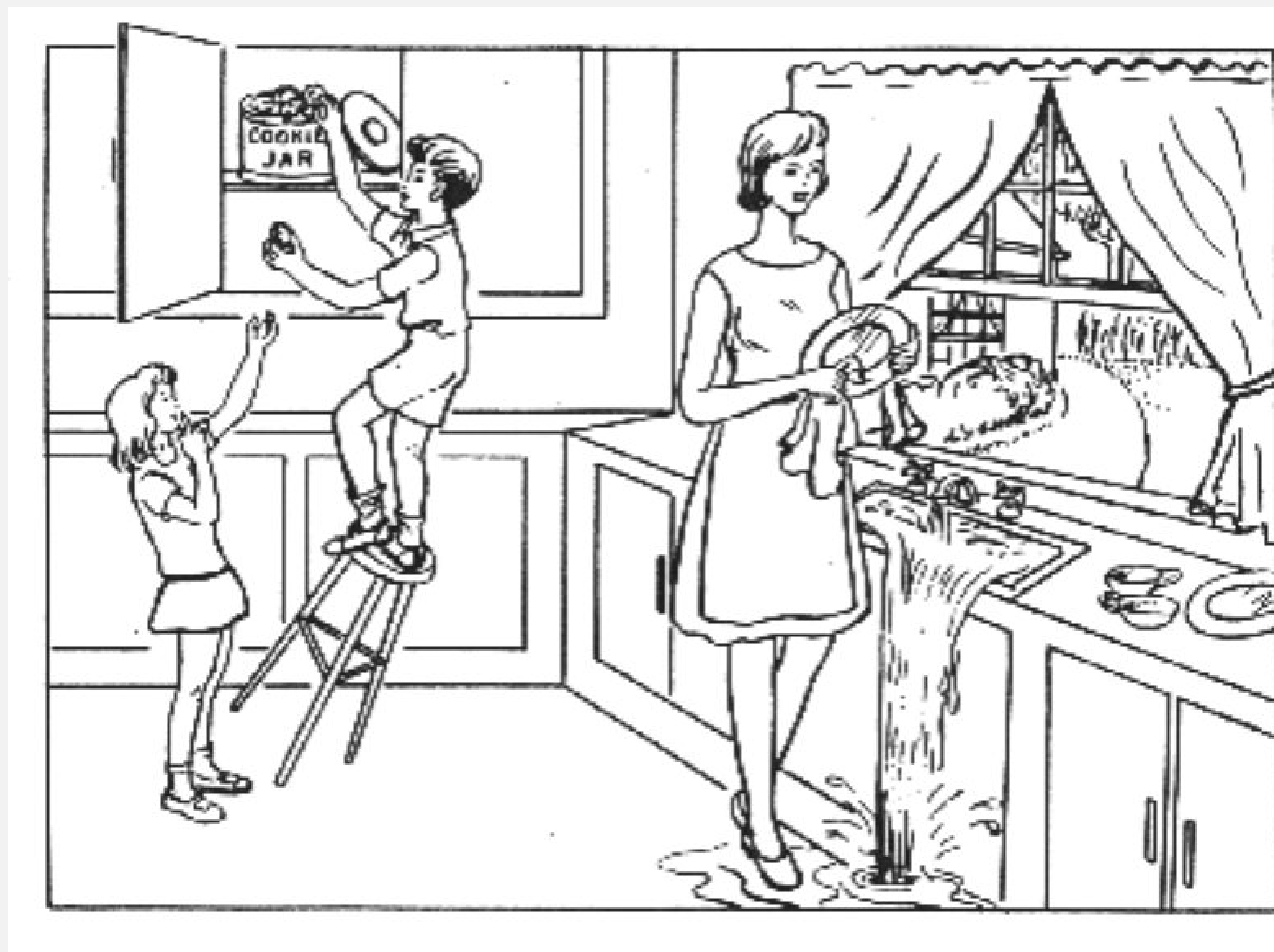
Metric of focus: *filled pauses*
Hypothesis: Individuals with dementia or MCI will use filled pauses (i.e. "uh", "uhm", laughter) more frequently than the control individuals.

Acknowledgements

I would like to thank Dr. Graeme Hirst for his support and guidance throughout the project.

Example

1. Transcription: Participants were asked to describe the Cookie Theft picture; utterances were transcribed into Dementia Bank's Transcript Database

*PAR: [\\] He's &uh in bad shape &uh the thing is falling over.

2. Pre-parsing: Transcriptions were parsed and removed of extraneous data, including part of speech tagging of words, emotion tagging ("laughs", "coughs", etc.), and investigator questions and responses.

He's [uh] in bad shape [uh] the thing is falling over.

He's in he's in bad shape the thing is falling over.

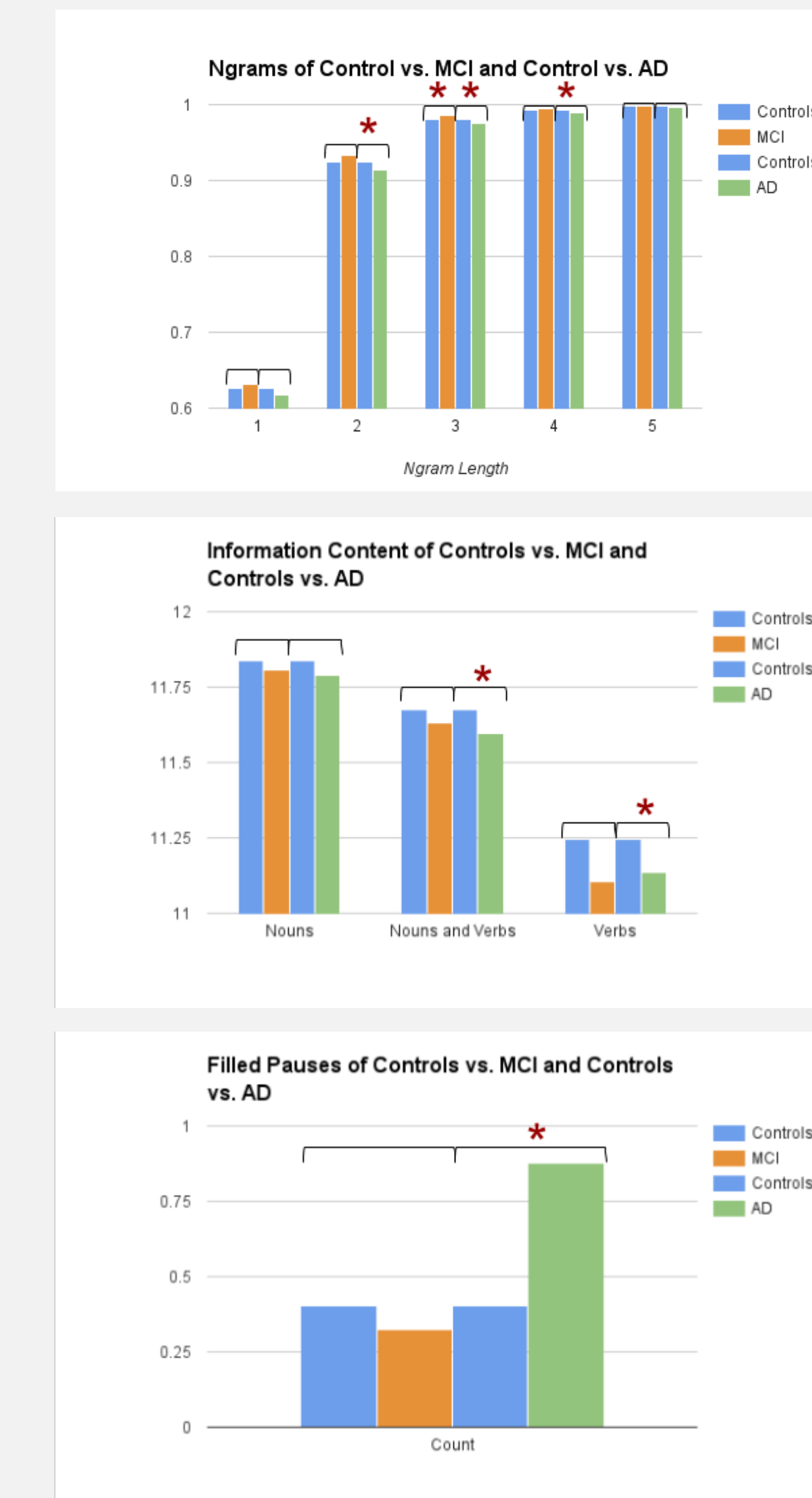He's in bad shape the thing is falling over.

He's uh in bad shape uh the thing is falling over.

He's in he's in bad shape the thing is falling over.

THING → FURNITURE → CHAIR → STOOL

3. (left) Type:token ratio, (middle) information content, and (right) filled pause word counts were calculated using newly designed algorithms for lexical analysis.

Results

Ngrams of Control vs. MCI and Control vs. AD

Information Content of Controls vs. MCI and Controls vs. AD

Filled Pauses of Controls vs. MCI and Controls vs. AD

Our models confirm that significant differences exist between repeated phrase usage (for 2-4grams), information content (for verbs and nouns+verbs), and filled pauses among individuals of the dementia and control groups.

Moreover, the results demonstrate a significant difference between the MCI and control groups for repeated phrases (trigrams only).

* indicates significance (p-value < .05)

Future Work

For future assessment, the following features will be incorporated into our algorithm to form a more robust analysis of language variation in dementia patients:
Indefinite word usage – combining information content with depth analysis
N-gram analysis of parts of speech – type:token extension to POS tagging
Coherence – using cosine similarity to measure target sentiment

References

[1] Fratiglioni, L., De Ronchi, D., Aguero-Torres, H. (1999). Worldwide Prevalence and Incidence of Dementia. Drugs and Aging, 15(5), 365-375.
[2] Potkins, D., Myint, P., Bannister, C., Tadros, G., Chithramohan, R., Swann, A., O'Brien, J., Fossey, J., George, E., Ballard, C., & Margallo-Lana, M. (2003). International Journal of Geriatric Psychiatry, 18, 1002-1006.
[3] Burke, D., & Shafto, M. (2004). Aging and Language Production, Current Directions in Psychological Science, 13(1), 21-24.
[4] Baiyewu, O., Unverzagt, F., Ogunniyi, A., Smith-Gamble, V., Gureje, O., Lane, K., Gao, S., Hall, K., & Hendrie, H. (2012). Behavioral Symptoms in Community-Dwelling Elderly Nigerians with Dementia, Cognitive Mild Impairment, and Normal Cognition. Geriatric Psychiatry, 27(9), 931-939.
[5] Unverzagt, F., Gao, S., Baiyewu, O., Ogunniyi, A., Gureje, O., Perkins, A., Emsley, C., Dickens, J., Evans, R., Musick, B., Hall, K., Hui, S., & Hendrie, H. (2001). Prevalence of Cognitive Impairment: Data from the Indianapolis Study of Health and Aging. Journal of the American Heart Association, 57(9), 1655-1662.
[6] Lancashire, I. And Hirst, G. (2009). Vocabulary Changes in Agatha Christie's Mysteries as an Indication of Dementia: A Case Study. Cognitive Aging: Research and Practice, 8-10 March 2009, Toronto, Canada.