Data Collection from Persons with Mild Forms of Cognitive Impairment and Healthy Controls - Infrastructure for Classification and Prediction of Dementia

Dimitrios Kokkinakis, Kristina Lundholm Fors, Eva Björkner
Department of Swedish
University of Gothenburg
Sweden
`first.last@svenska.gu.se`

Arto Nordlund
Department of Psychiatry and Neurochemistry
Sahlgrenska Academy, University of Gothenburg
Sweden
`first.last@neuro.gu.se`

Abstract

Cognitive and mental deterioration, such as difficulties with memory and language, are some of the typical phenotypes for most neurodegenerative diseases including Alzheimer's disease and other dementia forms. This paper describes the first phases of a project that aims at collecting various types of cognitive data, acquired from human subjects in order to study relationships among linguistic and extra-linguistic observations. The project's aim is to identify, extract, process, correlate, evaluate, and disseminate various linguistic phenotypes and measurements and thus contribute with complementary knowledge in early diagnosis, monitor progression, or predict individuals at risk. In the near future, automatic analysis of these data will be used to extract various types of features for training, testing and evaluating automatic classifiers that could be used to differentiate individuals with mild symptoms of cognitive impairment from healthy, age-matched controls and identify possible indicators for the early detection of mild forms of cognitive impairment. Features will be extracted from audio recordings (speech signal), the transcription of the audio signals (text) and the raw eye-tracking data.

1 Introduction

Aiding the detection of very early cognitive impairment in Alzheimer's disease (AD) and assessing the disease progression are essential foundations for effective psychological assessment, diagnosis and planning; enabling patients to participate in new drug therapy research and design of new clinical trials; evaluating potential disease-modifying agents in suitable populations etc. Efficient tools for routine dementia screening in primary health care, and particularly non-invasive and cost-effective methods in routine dementia screening for the identification of subjects who could be administered for further cognitive evaluation and dementia diagnostics, could provide specialist centres the opportunity to engage in more demanding, advanced investigations, care and treatment. New paths of research for acquiring knowledge about Alzheimer's disease (AD) and its subtypes using Computational Linguistic/ Natural Language Processing (CL/NLP) techniques and tools based on the exploration of several complementary modalities, parameters and features, such as speech analysis and/or eye tracking could be integrated into established neuropsychological, memory and cognitive test batteries in order to explore potential (new) biomarkers for AD. This paper describes current efforts to acquire data from people with subjective (SCI) and mild cognitive impairment (MCI) and healthy, age-matched controls in order to analyse and evaluate potential useful linguistic and extra-linguistic features and build classifiers that could differentiate between benign and malignant forms of cognitive impairment. Non-invasive and cost-effective methods that could identify individuals at an early preclinical dementia stage remains a priority and a challenge for health care providers while language deficits have been reported early in the development of AD (Taler & Phillips, 2008). Moreover, NLP methods are applied more and more in various biomedical and clinical settings, while patient language samples and large datasets are used routinely in NLP research such as the Dementia Bank corpus (a part of the TalkBank project; MacWhinney et al., 2011) and the Cambridge Cookie-Theft Corpus (Williams et al., 2010). The SCI, the MCI, and the

Alzheimer's disease (AD) are on a spectrum of disease progression. Subjective cognitive impairment (SCI) is a common diagnosis in elderly people, sometimes suggested to be associated with e.g. depression, stress or anxiety, but also a risk factor for dementia (Jessen et al., 2010). On the other hand, mild cognitive impairment (MCI) is a prodromal state of dementia (Ritchie & Touchon, 2010), in which a human subject has minor problems with cognition (e.g., problems with memory or thinking) but these are not severe enough to warrant a diagnosis of dementia or interfere significantly with daily life, but still difficulties which are worse than would normally be expected for a healthy person of their age.

This paper describes some efforts underway to acquire, assess, analyze and evaluate linguistic and extra-linguistic data from people with subjective (SCI) and mild cognitive impairment (MCI) and healthy, age-matched controls, with focus on infrastructure (e.g., resource collection, envisaged analysis and feature acquisition and modeling).

2   Background

New findings aim to provide a comprehensive picture of cognitive status and some promising results have recently thrown more light on the importance of language and language (dis)abilities as an essential factor that can have a strong impact on specific measurable characteristics that can be extracted by automatic linguistic analysis of speech and text (Ferguson et al., 2013; Szatloczki et al., 2015). The work by Snowdon et al. (2000), "The Nun Study", was one of the earliest studies which showed a strong correlation between low linguistic ability early in life and cognitive impairment in later life by analyzing autobiographies of American nuns. Snowdon et al. could predict who could develop Alzheimer's by studying the degradation of the idea density (that is, the average number of ideas expressed in 10 words; Chand et al., 2010) and syntactic complexity on the nuns' autobiographical writings. Since then, the body of research and interest in CL/NLP in the area of processing data from subjects with mental, cognitive, neuropsychiatric, or neurode-generative impairments has grown rapidly. Automatic spoken language analysis and eye movement measurements are two of the newer complementary diagnostic tools with great potential for dementia diagnostics (Laske et al.,

2014). Furthermore, the identification of important linguistic and extra-linguistic features such as lexical and syntactic complexity, are becoming an established way to train and test supervised machine learning classifiers that can be used to differentiate between individuals with various forms of dementia and healthy controls or between individuals with different types of dementia (Lagun et al., 2011; Roark et al., 2011; Olubolu Orimaye et al., 2014; Rentoumi et al., 2014).

Although language is not the only diagnostic factor for cognitive impairment, several recent studies (Yancheva et al., 2015) have demonstrated that automatic linguistic analysis, primarily of connected speech samples, produced by people with mild or moderate cognitive impairment compared to healthy individuals can identify with good accuracy objective evidence and measurable (progressive) language disorders. Garrard & Elvevåg (2014) comment that computer-assisted analysis of large language datasets could contribute to the understanding of brain disorders. Although, none of the studies presented in the special issue of Cortex vol. 55 moved "beyond the representation of language as text" and therefore finding reliable ways of incorporating features, such as prosody and emotional connotation, into data representation remains a future challenge, the editors acknowledged that current research indicates that "the challenges of applying computational linguistics to the cognitive neuroscience field, as well as the power of these techniques to frame questions of theoretical interest and define clinical groups are of practical importance". Nevertheless, studies have shown that a steady change in the linguistic nature of the symptoms and the degree in speech and writing are early and could be identified by using language technology analysis (Mortimer et al., 2005; Le et al., 2011). New findings also show a great potential to increase our understanding of dementia and its impact on linguistic degradation such as loss of vocabulary, syntactic simplification, poor speech content and semantic generalization. Analysis of eye movement is also a relevant research technology to apply, and text reading by people with and without mild cognitive impairment may give a clear ruling on how reading strategies differ between these groups, an area that has so far not been researched to any significant extent in this particular domain (Fernández et al., 2013, 2014;

Molitor et al., 2015). With the help of eye-tracking technology the eye movements of participants are recorded while suitable stimuli is presented (e.g., a short text).

3 Ethical Issues and Patient Recruitment

The ongoing Gothenburg mild cognitive impairment study (Nordlund et al., 2005; Wallin et al., 2016) is an attempt to conduct longitudinal in-depth phenotyping of patients with different forms and degrees of cognitive impairment using neuropsychological, neuroimaging, and neurochemical tools. The study is clinically based and aims at identifying neurodegenerative, vascular and stress related disorders prior to the development of dementia. All patients in the study undergo baseline investigations, such as neurological examination, psychiatric evaluation, cognitive screening (e.g., memory and visuospatial disturbance, poverty of language and apraxia), magnetic resonance imaging of the brain and cerebrospinal fluid collection. At biannual follow-ups, most of these investigations are repeated. The overall Gothenburg MCI-study is approved by the local ethical committee review board (reference number: L091–99, 1999; T479-11, 2011); while the currently described study by the local ethical committee decision 206-16, 2016). The project aims at gathering a rather homogeneous group of participants with respect to age and education level (50 with SCI/MCI and 50 controls). All subjects have participated into a comprehensive battery of e.g. memory, language and other tests which have been described in Wallin et al. (2016). Recruitment of patients in the project was consecutive and took place over the course of several months, from July 2016 to January 2017. All participants gave informed written consent and were advised that no notations were made that could be related to their identity, while the exclusion and inclusion criteria are specified according to the following protocol.

Inclusion criteria:

- Age range 50-79 years
- Swedish as a first language and not speaking languages other than Swedish before the age of 5
- Comparable education length of the participants

- No apparent organic cause of symptoms (such as e.g. stroke or brain tumor)
- Research subjects have read information about the research project and approved voice recording and eye movement measurements
- Participants have conducted *recent* neuropsychological tests – participants should not have deteriorated significantly since the last testing

Exclusion criteria:

- Participants have comorbid conditions affecting reading (e.g. dyslexia or other reading difficulties)
- Participants have deep depression
- Participants have an ongoing abuse of any kind
- Participants suffer from serious psychiatric or neurological diseases such as Parkinson's, Amyotrophic lateral sclerosis or have/had a brain tumor
- Participants do not understand the question or the context in the selection process
- Participants have poor vision (that cannot be corrected by glasses or lenses), cataract, nystagmus, or cannot see and read on the computer screen
- Participants decline participation during telephone call or later at the recording site
- Participants decline signing the paper of informed consent
- Recordings or eye movement measurements are technically unusable.

4 Material and Experimental Design

4.1 Spoken signal/audio

For the acquisition of the audio signal we use the Cookie-theft picture [1] (see Figure 1) from the *Boston Diagnostic Aphasia Examination* (BDAE;

---

[1] Cookie-theft is a picture that has been a source of knowledge for various clinical and experimental research worldwide which enables even future cross-linguistic comparisons.

Goodglass & Kaplan, 1983) which is often used to elicit speech from people with various mental and cognitive impairments. During the presentation of the Cookie-theft stimuli (which illustrates an event taking place in a kitchen) the subjects are asked to tell a story about the picture and describe everything that can be observed while the story is recorded. For the task the original label of the cookie jar is translated and substituted from the English "COOKIE JAR" to the Swedish label "KAKBURK". The samples are recorded in an isolated environment and the whole task is designed to evoke a monologue by the participant. The instruction given to the subject was: "Tell me everything you see going on in this picture, describe objects and events. You can go on as long as you prefer and you will be not interrupted until you indicate that you do not have more to say".
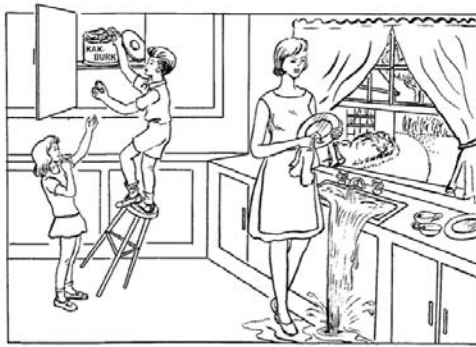

Figure 1: The Cookie-theft picture.

We chose to use the Cookie Theft picture since it provides a standardized test that has been used in various studies in the past, and therefore comparisons can be made based on previous results, e.g. with research on the DementiaBank database or other collections (MacWhinney, 2007; Williams et al., 2010; Fraser & Hirst, 2016) and even Swedish studies (Tyche, 2001). The picture is considered an "ecologically valid approximation" to spontaneous discourse (Gilles et al., 1996). Moreover, in order to allow the construction of a comprehensive speech profile for each research participant, the speech task also includes reading aloud a short text from the *International Reading Speed Texts* collection (IReST; Trauzettel-Klosinski et al., 2012) presented on a computer screen. As a matter of fact, two texts are used from the IReST collection, in connection to the eye tracking experiment, but only one of those texts is read aloud and thus combined with eye-tracking recording; cf. Meilán et al., 2012 and 2014 for

similar "reading out" text passage experiments. IReST is a multilingual standardized text collection used to assess reading performance, for multiple equivalent texts for repeated measurements. Specifically in our project we use the Swedish IReST translations, namely texts "one" and "seven" (Öqvist Seimyr, 2010). For the audio capture of both we use a H2n Handy recorder while the audio files are saved and stored as uncompressed audio in .wav 44.1 kHz with 16-bit resolution. This recording is carried out in the same isolated environment in order to avoid noise.

4.2 Verbatim transcriptions

The textual part of the infrastructure consists of manually produced transcriptions of the two audio recordings previously described. The digitized speech waveform is semi-automatically aligned with the transcribed text. During speech transcription, special attention is also paid to non-speech acoustic events including speech dysfluencies consisting of filled pauses a.k.a. hesitation ("um"), false-starts, repetitions as well as other features, particularly non-verbal vocalizations such as laughing, sniffing and coughing. A basic transcription manual, with the various conventions to be used, is produced which helps the human transcribers accomplish a homogeneous transcription. For instance, all numerals should be written out as complete words, while symbols, such as square brackets, are used for the encoding of pauses or transcriber's comments. Furthermore, for the transcription the PRAAT application (Boersma & Weenink, 2013) is utilized; using a 2-tier text grid configuration, one for orthographic transcription (standardized spelling) and one with maintained spoken language phenomena, such as partial words, see Figure 2.
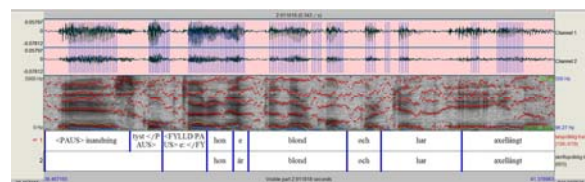

Figure 2: Transcription of the collected data.

4.3 Eye-tracking

The investigation of eye movement functions in SCI/MCI, and any differences or changes in eye movements that could be potentially detected for those patients is of great importance to clinical

AD research. However, until now, eye tracking has not been used to investigate reading for MCI-persons on a larger scale, possibly due to the number of procedural difficulties related to this kind of research. On the other hand, the technology has been applied in a growing body of various experiments related to other impairments such as autism (Yaneva et al., 2016; Au-Yeung et al., 2015), dyslexia (Rello & Ballesteros, 2015) and schizophrenia (Levy et al., 2010); for a thorough review see Anderson & MacAskill (2013). For the experiments we use EyeLink 1000 Desktop Mount with monocular eye tracking with head stabilization and a real-time sample access of 1000Hz. Head stabilization provides an increased eye tracking range performance. The participants are seated in front of the monitor at a distance of 60-70 cm. While reading, the eye movements of the participants are recorded with the eye-tracking device while interest areas around each word in the text are defined by taking advantage of the fact that there are spaces between each word in the text. The eye-tracking measurements are used for the detection and calculation of fixations, saccades and backtracks. Fixation analyses is conducted within predefined Areas of Interest (AOI); in our case each word is an AOI (see Figure 3).
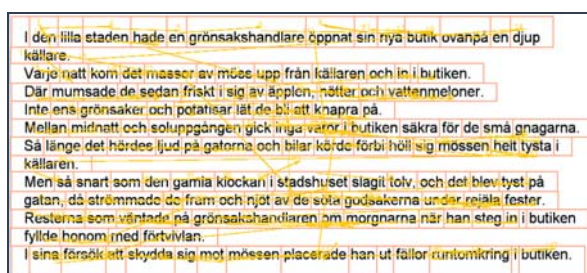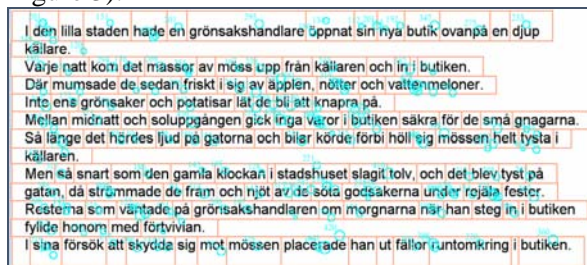
Figure 3: A text from the IReST collection with marked fixations (top) and saccades (bottom).

4.4 Comparison over a two year period

The previously outlined experiments/audio recordings will be repeated two years after the first recording which took place during the second half of 2016. This way we want to analyze whether there are any differences between the two audio and eye-tracking recordings, and at which level and magnitude. We want to compare and examine whether there any observable, greater, differences/decline on some features and which these could be and therefore the nature and eventually progression of speech impairment or eye movement alterations observed over a two year period. We are aware that more longitudinal data samples over longer time periods would be desirable but at this stage only a single repetition is practically feasible to perform. Longitudinal experiments, e.g. in investigating the nature and progression of the spontaneous writing, patterns of impairment were observed in patients with Alzheimer's disease over a 12-month period, these were dominated by semantic errors (Forbes-McKay et al., 2014). Ahmed et al. (2013) reported changes that took place in spoken discourse over the course of three clinical stages. Measures of language function mirrored global progression through the successive clinical stages of the disease. In an individual case analysis, results showed that there were significant but heterogeneous changes in connected speech for 2/3 of the studied MCI-group.

5 Analysis and Features

The envisaged analysis and exploration intends to extract, evaluate and combine a number of features from the three modalities selected to be investigated. These are speech-related features, text/transcription-related features and eye tracking-related features.

5.1 Speech-related analysis

A large number of acoustic and prosodic features has been proposed in the literature which pinpoints the importance of distinguishing between vocal changes that occur with "normal" aging and those that are associated with MCI (and AD). Finding reliable and robust acoustic features that might differentiate spoken language of SCI/MCI and healthy controls remains an ongoing challenge but the technology develops rapidly. Based on related literature, we would expect that our spoken samples might show different qualities depending on whether they are produced spontaneously (when talking about the Cookie-theft picture) or they consist of a read aloud task.

Prosodic features have been found to be useful in distinguishing between subjects with cognitive impairment and healthy controls, and between groups with varying degrees of cognitive impairment. Pause frequency has been identified as a feature differentiating spontaneous speech in patients with AD from control groups (Gayraud et al., 2011), and may also be used to distinguish between mild, moderate and severe AD (Hoffman et al., 2010). Subjects with AD also tend to make more pauses and non-syntactic boundaries (Lee et al., 2011). Speech tempo, which is defined as phonemes per second (including hesitations) differs significantly between subjects with AD and controls. Speech tempo is also positively correlated with Mini Mental State Examination (MMSE) results (Hoffman et al., 2010), which suggests that people with less cognitive deficits will produce speech at a faster rate.

Speech-related features have been used successfully in machine learning experiments where the aim has been to identify subjects with AD. Roark et al. (2011) used 21 features in supervised machine learning experiments (using Support Vector Machines, SVM) from 37 MCI subjects and equally many controls (37/37). Features from both the audio and the transcripts included: pause frequency, filled pauses, total pause duration and linguistic variables such as Frazier and Yngve scores and idea density, while best accuracy with various feature configurations were 86.1% for the area under the ROC curve. Pause frequency has been identified as a feature differentiating spontaneous speech in patients with AD from control groups (Gayraud et al., 2011), and may also be used to distinguish between mild, moderate and severe AD (Hoffman et al., 2010). Meilán et al. (2014) used AD subjects and spoken data (read loud and clear sentences on a screen). They used acoustic measures such as pitch, volume and spectral noise measures. Their method was based on linear discriminant analysis and their results could characterize people with AD with an accuracy of 84.8%. Yancheva et al. (2015) used spoken and transcriptions features provided from the DementiaBank (Cookie-theft descriptions) using 393 speech samples (165/90). They extracted and investigated 477 different features both lexicosyntactic ones (such as syntactic complexity; word types, quality and frequency) and acoustic ones (such as Melfrequency cepstral coefficients – MFCC, including mean, variance, skewness, and kurtosis; pauses and fillers; pitch

and formants and aperiodicity measures) and semantic ones (such as concept mention) in order to predict MMSE scores with a mean absolute error of 3.83 while with individuals with more longitudinal samples the mean absolute error was improved to 2.91, which suggested that the longitudinal data collection plays an important role. König et al. (2015) looked also at MCI and AD subjects (23/26) and examined vocal features (silence, voice, periodic and aperiodic segment length; mean of durations) using Support Vector Machines. Their classification accuracy of automatic audio analysis was 79% between healthy controls and those with MCI; 87% between healthy controls and those with AD; and between those with MCI and those with AD, 80%. Tóth et al. (2015) used also SVM and achieved 85.3% F-score (32 MCI subjects and 19 controls) by starting with eight acoustic features extracted by applying automatic speech recognition (such as speech tempo i.e. phones per second) and extending them to 83. Finally, Fraser et al. (2016) also looked at the DementiaBank and using 240 samples from AD subjects and 233 from healthy controls, extracted 370 features, such as linguistic variables from transcripts (e.g., part-of-speech frequencies; syntactic complexity and grammatical constituents), psycholinguistic measures (e.g., vocabulary richness) and acoustic variables from the audio files (e.g., MFCC). Using logistic regression, Fraser et al. could obtain a classification accuracy of 81% in distinguishing individuals with AD from those without based on short samples of their language on the Cookie-theft picture description task.

In our analysis, we plan to extract prosodic features such as pitch variation, pause length and frequency, hesitation rate and speech rate, and use these both in stand-alone machine learning experiments, and combined with features extracted from voice analysis, eye-tracking and the transcriptions.

5.2   Voice acoustic-related analysis

Depression commonly occurs among patients diagnosed with MCI. Signs of depression are often expressed as an emotional feeling of sadness or "low mood". Johnson et al (2013) found that MCI participants with depression experienced greater deficits in cognitive functioning than their non-depressed counterparts, and "low mood" were shown by Caracciolo et al. (2011) to be particularly prominent in the very early stages of cognitive

decline and strongly associated with amnestic mild cognitive impairment (aMCI), i.e. the pre-dementia stage of Alzheimer's, than with global cognitive impairment. Different emotions are accompanied by various adaptive responses in the autonomic and somatic nervous systems (Johnstone & Scherer, 2000). These responses are known to lead to changes in the functioning parts of the speech production system, such as respiration, vocal fold vibration and articulation. The vibrating vocal folds produce the voiced sound (voice source) and the articulation determines the position of the formant frequencies (Hz), determining the vowel and the sound quality of the voice.

The most commonly used parameters in speech acoustic analysis are fundamental frequency (F0) and formant frequency analysis, perturbation measurement such as jitter and shimmer (cycle-to-cycle variations in frequency and amplitude, respectively), and harmonic-to-noise ratios (HNR/NHR). Studying the role of voice production in emotional speech, Patel et al (2011) found significant emotion main effects for 11 of 12 acoustic parameters for five emotions (joy, relief, hot anger, panic fear, sadness) where sadness was characterized by low energy and a hypo-functioning voice quality. Further, Meilán et al. (2014) found voice perturbation parameters to distinguish people with AD from healthy controls with an accuracy of 84.8%.

The aim of the acoustic analysis in the present study is to see if voice parameters can be used to distinguish also between healthy controls and SCI/MCI-patients. Several vowel /a/ and /i/-samples from the read aloud text and the spontaneously spoken Cookie-theft picture will be analyzed for each subject, using the Praat software (Boersma and Weenink). The acoustic data will be compared and correlated according to age, gender, length of education, depression score and other parameters gained from the neuropsychological assessment (Wallin et al., 2016).

5.3 Transcribed speech analysis

Many of the previous studies combine both acoustic features and features from the transcriptions; cf. the supplementary material in Fraser et al. (2016). Some of the most common features and measures from transcribed text follow the lexicon-syntax-semantics continuum. These measures include (i) *lexical distribution measures* (such as type-token ratio, mean word length, long word counts, hapax legomena,

hapax dislegomena, automated readability index and Coleman-Liau Index; also lexical and non-lexical fillers or disfluency markers, i.e. "um", "uh", "eh") and out-of-vocabulary rate (Pakhomov et al., 2010). (ii) *syntactic complexity markers* (such as frequency of occurrence of the most frequent words and deictic markers; [context free] production rules, i.e. the number of times a production rule is used divided by the total number of productions; dependency distance, i.e. the length of a dependency link between a dependent token and its head, calculated as the difference between their positions in a sentence; parse tree height, i.e. is the mean number of nodes from the root to the most distant leaf; depth of a syntactic tree, i.e. the proportion of subordinate and coordinate phrases to the total number of phrases and ratio of subordinate to coordinate phrases; noun phrase average length and noun phrase density, i.e. the number of noun phrases per sentence or clause; words per clause); and (iii) *semantic measures* (such as the idea or propositional density, i.e. the operationalization of conciseness – the average number of ideas expressed per words used; the number of expressed propositions divided by the number of words; a measure of the extent to which the speaker is making assertions, or asking questions, rather than just referring to entities etc.). Since some of the features to be extracted (e.g. part-of-speech and syntactic labels from the speech transcriptions) are language-dependent it requires the use of a language-specific infrastructure (in our case Swedish), for that reason we plan to use available resources; cf. Ahlberg et al. (2013); therefore testing and modifications to the transcribed language are also envisaged. Two wide-coverage parser systems will be used for parsing the speech transcripts. The Malt parser for Swedish (Nivre et al., 2006), that outputs grammatical dependency relations, and a constituent parser for the same language (Kokkinakis, 2001) that utilises a semi-automatically developed grammar. Although the transcribed corpus is describing spoken language and contains various spoken language phenomena, such as filled pauses, we chose to keep the verbatim transcriptions intact. Such phenomena are usually deleted prior to parsing for better performance (Lease & Johnson 2006; Geertzen, 2009). Moreover, since we apply a 2-tier text grid configuration during the transcription, we can easily experiment with both

the orthographic transcription (standardized spelling) and the verbatim one.

5.4 Eye-tracking analysis

Eye tracking data has been used in machine learning methods in the near past. By taking advantage of biomarkers extracted from eye dynamics (Lagun et al; 2011) there is an indication that these could aid the automatic detection of cognitive impairment (i.e., distinguish healthy controls from MCI-patients). Several studies provide evidence and suggest that eye movements can be used to detect memory impairment and serve as a possible biomarker for MCI and, in turn, AD (Fernández et al., 2013). Basic features we intend to investigate in this study are *fixations* (that is the state the eye remains still over a period of time); *saccades* (that is the rapid motion of the eye from one fixation to another) and *backtracks* (that is the relationship between two subsequent saccades where the second goes in the opposite direction than the first); for a thorough description of possible eye-tracking related features cf. Holmqvist et al. (2015:262). Saccades are of particular interest because they are much related to attention and thus, they are likely to be disturbed by cognitive impairments associated with neurodegenerative disorders (Anderson & MacAskill, 2013). Note that there are many assumptions behind the use of eye tracking technology for experiments designed for people with MCI. For instance, the longer the eye gaze fixation is on a certain word, the more difficult the word could be for cognitive processing, therefore the durations of gaze fixations could be used as a proxy for measuring cognitive load (Just & Carpenter, 1980). Molitor et al. (2015) provide a recent review on the growing body of literature that investigates changes in eye movements as a result of AD and the alterations to oculomotor function and viewing behaviour.

5.5 Correlation analysis

We intend to further perform correlation analysis with the features previously outlined and the results of the various measures/scores on tasks from language-related tests performed in the Gothenburg MCI-study, applied for assessing possible dementia. Typically, clinicians use tests such as the MMSE, linguistic memory tests and language tests. Language tests include the token test, subtest V, which is a test of syntax comprehension; the Boston naming test, the semantic similarity test; the letter/word fluency FAS test (the generation of words beginning by the letters F, A, and S) and the category or semantic fluency test (the generation of words that fall into a given semantic category, such as animals). This investigation intends to identify whether there are language-related features, acquired from the range of available tests, which could be (highly negative or highly positive) correlated with i.e. the MCI class, yet uncorrelated with each other i.e. the healthy controls or SCI. We want to further investigate which scores correlate with which variables derived from the picture description. It has been argued, Kavé & Goral (2016), that the picture naming task *could* be a better predictor of word retrieval in context than the semantic fluency task for several reasons, for instance the speech elicitation method most likely involves cognitive demands that are similar to the ones required for the picture naming task, e.g., specific labelling.

6 Conclusions and Future Work

In this paper we have introduced work in progress towards the design and infrastructural development of reliable multi-modal data resources and a set of measures (features) to be used both for experimentation with feature engineering and evaluation of classification algorithms to be used for differentiating between SCI/MCI and healthy adults, and also as benchmark data for future research in the area. Evaluation practices are a crucial step towards the development of resources and useful for enhancing progress in the field, therefore we intend to evaluate both the relevance of features, compare standard algorithms such as Support Vector Machines and Bayesian classifiers and perform correlation analysis with the results of established neuropsychological, memory and cognitive tests. We also intend to repeat the experiments two years after (2018) the current acquisition of data in order to assess possible changes at each level of analysis. We believe that combining data from three modalities (a form of data fusion; Mitchel, 2007) could be useful, but at this point we do not provide any clinical evidence underlying these assumption since the analysis and experimentation studies are currently under way (year 2 of the project, 2017). Therefore, at this stage, the paper only provides a high-level review of the current stage of the work.

References

Malin Ahlberg et al. 2013. *Korp and Karp – a bestiary of language resources: the research infrastructure of Språkbanken*. 19th Nordic Conf of Computational Linguistics (NODALIDA). Linköping Electronic Conference Proceedings #85.

Samrah Ahmed, Anne-Marie Haigh, Celeste de Jager and Peter Garrard. 2013. Connected speech as a marker of disease progression in autopsy-proven Alzheimer's disease. *Brain*. 136(Pt 12):3727-37.

Tim J. Anderson and Michael R. MacAskill. 2013. Eye movements in patients with neurodegenerative disorders. *Nat Rev Neurology* 9: 74-85. doi:10.1038/nrneurol.2012.273.

Eiji Aramaki, Shuko Shikata, Mai Miyabe and Ayae Kinoshita. 2016. Vocabulary Size in Speech May Be an Early Indicator of Cognitive Impairment. *PLoS One*. 11(5):e0155195.

Sheena K. Au-Yeung, Johanna Kaakinen, Simon Liversedge and Valerie Benson. 2015. Processing of Written Irony in Autism Spectrum Disorder: An Eye-Movement Study. *Autism Res.* 8(6):749-60. doi: 10.1002/aur.1490.

Paul Boersma and David Weenink. 2013. *Praat: doing phonetics by computer* [Computer program]. Version 6.0.19, retrieved in Aug. 2016 from <http://www.praat.org/>.

Barbara Caracciolo et al. 2011. The symptom of low mood in the prodromal stage of mild cognitive impairment and dementia: a cohort study of a community dwelling elderly population. *J Neurol Neurosurg Psychiatry*. 82:788-793.

Vineeta Chand, Kathleen Baynes, Lisa M. Bonnici and Sarah Tomaszewski Farias. 2012. A Rubric for Extracting Idea Density from Oral Language Samples Analysis of Idea Density (AID): A Manual. *Curr Protoc Neurosci*. Ch. Unit10.5. doi:10.1002/0471142301.ns1005s58.

James W Dodd 2015. Lung disease as a determinant of cognitive decline and dementia. *Alzh Res & Therapy*, 7:32.

Alison Ferguson, Elizabeth Spencer, Hugh Craig and Kim Colyvas. 2014. Propositional Idea Density in women's written language over the lifespan: Computerized analysis. *Cortex* 55. 107-121.

Gerardo Fernández et al. 2013. Eye Movement Alterations during Reading in Patients with Early Alzheimer Disease. *Investigative Ophthalmology & Visual Science*. Vol.54, 8345-8352. doi:10.1167/iovs.13-12877.

Katrina Forbes-McKay, Mike Shanks and Annalena Venneria. 2014. Charting the decline in spontaneous writing in Alzheimer's disease: a longitudinal study. *Acta Neuropsychiatrica*. Vol. 26:04, pp 246-252.

Kathleen C. Fraser and Graeme Hirst. 2016. *Detecting semantic changes in Alzheimer's disease with vector space models*. LREC Workshop: Resources and ProcessIng of linguistic and extra-linguistic Data from people with various forms of cognitive/psychiatric impairments (RaPID). Pp. 1-8. Portorož Slovenia.

Peter Garrard and Brita Elvevåg. 2014. Special issue: Lang., computers and cognitive neuroscience. *Cortex* 55; 1-4.

Frederique Gayraud, Hye-Ran Lee and Melissa Barkat-Defradas. 2011. Syntactic & lexical context of pauses and hesitations in the discourse of Alzheimer patients and healthy elderly subjects. *Clin Ling&Phon*. 25(3):198-209.

Jeroen Geertzen. 2009. *Wide-coverage parsing of speech transcripts*. 11th Pars. Tech (IWPT). Pp 218–221. France.

Elaine Gilles, Karalyn Patterson and John Hodges. 1996. Performance on the Boston Cookie Theft picture description task in patients with early dementia of the Alzheimer's type: missing information. *Aphasiology*. 10:4:395-408.

Harald Goodglass and Edith Kaplan. 1983. *The Assessment of Aphasia and Related Disorders*. Lea&Febiger. USA.

Ildikó Hoffmann et al. 2010. Temporal parameters of spontaneous speech in Alzheimer's disease. *J of Speech-Language Pathology*, *12*(1), 29–34.

Kenneth Holmqvist, Richard Dewhurst, Marcus Nyström, Joost van de Weijer, Halszka Jarodzka and Richard Andersson. 2015. *Eye Tracking - A comprehensive guide to methods & measures*. OUP.

Frank Jessen et al. 2010. Prediction of dementia by subjective memory impairment: effects of severity and temporal association with cognitive impairment. *Arch. Gen. Psychiatry*, 67(4). Pp. 414–422.

Leigh A Johnson et al. 2013. Cognitive differences among depressed and non-depressed MCI participants. *J Geriatr Psychiatry*. 28(4):377-82.

Tom Johnstone and Klaus R. Scherer. 2000. Vocal communication of emotion. *The Handbook of Emotion*. Lewis & Haviland (eds). NY Guildford.

Marcel A. Just and Patricia A. Carpenter. 1980. A theory of reading: from eye fixations to

comprehension. *Psychological review*, 87(4):329-354.

Gitit Kavé & Mira Goral. 2016. Word retrieval in picture descriptions produced by individuals with Alzheimer's disease. *J Clin Exp Neuropsychol.* 38(9):958-66.

Dimitrios Kokkinakis. 2001. *More than Surface-Based Parsing; Higher Level Evaluation of Cass-SWE*. 13th Nordic Computational Linguistics Conference (NODALIDA). Uppsala, Sweden.

Alexandra König et al. 2015. Automatic speech analysis for the assessment of patients with predementia and Alzheimer's disease. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring.* 1:112–124. Elsevier.

Dmitry Lagun et al. 2011. Detecting cognitive impairment by eye movement analysis using automatic classification algorithms. *J Neurosci Methods.* 201(1): 196–203. doi:10.1016/j.jneumeth.2011.06.027.

Christoph Laske et al. 2014. Innovative diagnostic tools early detection of Alzheimer's disease. *Alzheimer's & Dementia.* 1-18.

Xuan Le, Ian Lancashire, Graeme Hirst, and Regina Jokel. 2011. Longitudinal Detection of Dementia through Lexical and Syntactic Changes in Writing: A Case Study of Three British Novelists. *JLLC* 26 (4): 435-461.

Matthew Lease and Mark Johnson. 2006. Early deletion of fillers in processing conversational speech. Proceedings of the Human Language Technology Conference of the North American Chapter of the ACL, pages 73–76.

Hyeran Lee, Frederique Gayraud, Fabrice Hirsh and Melissa Barkat-Defradas. 2011. Speech dysfluencies in normal and pathological aging: A comparison between Alzheimer patients and healthy elderly subjects. *ICPhS: proceedings of the 17th International Congress of Phonetic Sciences*. Pp. 1174–1177. Hong Kong.

Deborah L. Levy, Anne B. Sereno, Diane C. Gooding,and Gilllian A. O'Driscoll. 2010. Eye Tracking Dysfunction in Schizophrenia: Characterization and Pathophysiology. *Curr Top Behav Neurosci.* 4: 311–347.

Juan JG. Meilán, Francisco Martínez-Sánchez, Juan Carro, José A. Sánchez and Enrique Pérez. 2012. Acoustic Markers Associated with Impairment in Language Processing in AD. Spanish *J of Psych*. Vol. 15:2, 487-494.

Juan JG. Meilán et al. 2014. Speech in Alzheimer's Disease: Can Temporal and Acoustic Parameters Discriminate Dementia? *Dement Geriatr Cogn Disord* 2014;37:327–334. doi: 10.1159/000356726.

H. B. Mitchell. (2007). *Multi-Sensor Data Fusion: An Introduction.* Springer.

Robert J. Molitor, Philip C. Ko and Brandon A. Ally. 2015. Eye Movements in Alzheimer's Disease. *J of Alzheimer's Disease* 44, 1–12. IOS Press.

James A. Mortimer, Amy R. Borenstein, Karen M. Gosche and David A. Snowdon. 2005. Very Early Detection of Alzheimer Neuropathology and the Role of Brain Reserve in Modifying Its Clinical Expression. *J Geriatr Psychiatry Neurol.* 18(4): 218–223.

Joakim Nivre et al. 2007. MaltParser: A language-independent system for data-driven dependency parsing. *Natural Language Engineering.* 13(2):95-135.

Arto Nordlund, S. Rolstad, P. Hellström, M. Sjögren, S. Hansen and Anders Wallin. 2005. The Goteborg MCI study: mild cognitive impairment is a heterogeneous condition. *J Neurol Neurosurg Psychiatry.* 76(11):1485-90.

Sylvester Olubolu Orimaye, Jojo Sze-Meng Wong and Kren J. Golden. 2014. *Learning Predictive Linguistic Features for Alzheimer's Disease and related Dementias using Verbal Utterances*. Workshop on Computational Ling. & Clinical Psychology: From Linguistic Signal to Clinical Reality. 78–87. Maryland, USA.

Serguei VS Pakhomov et al. 2010. A co-mputerized technique to assess language use patterns in patients with frontotemporal dementia. *J Neuroling.* 23(2):127–144.

Sona Patel, Klaus R. Scherer, Eva Björkner, Johan Sundberg. 2011. Mapping emotions into acoustic space: The role of voice production. *Biological Psychology* 87. 93–98.

Sajidkhan S. Pathan et al. 2011. Association of lung function with cognitive decline and dementia: the Atherosclerosis Risk in Communities (ARIC) Study. *Eur J Neurol.* 18(6):888-9.

Luz Rello and Miguel Ballesteros. 2015. *Detecting Readers with Dyslexia Using Machine Learning with Eye Tracking Measures*. Proceedings of the 12th Web for All Conference W4A. Florence, Italy.

Vassiliki Rentoumi, Ladan Raoufian, Samrah Ahmed and Peter Garrard. 2014. Features and Machine Learning Classification of Connected Speech Samples from Patients with Autopsy Proven Alzheimer's Disease with and without Additional Vascular Pathology. *J of Alzheimer's Disease* 42. IOS Press. S3–S17.

Karen Ritchie and Jacques Touchon. 2010. Mild cognitive impairment: conceptual basis and current nosological status. *The Lancet.* Vol. 355:9199. Pp. 225–228. Doi:10.1016/S0140-6736(99)06155-3.

Brian Roark, Margaret Mitchell, John-Paul Hosom, Kristy Hollingshead, and Jeffrey Kaye. 2011. *Spoken Language Derived Measures for Detecting Mild Cognitive Impairment*. IEEE Trans Audio Speech Lang Processing. 19(7): 2081–2090.

David A. Snowdon, Lydia Greiner and William R. Markesbery. 2000. Linguistic ability in early life and the neuropathology of Alzheimer's disease and cerebrovascular disease. Findings from the Nun Study. *Annals of the NY Academy of Sciences*. 903:34-8.

Greta Szatloczki, Ildiko Hoffmann, Veronika Vincze, Janos Kalman and Magdolna Pakaski. 2015. Speaking in Alzheimer's disease, is that an early sign? Importance of changes in language abilities in Alzheimer's disease. *Frontiers in Aging Neuroscience*. Vol 7, article 195. doi: 10.3389/fnagi.2015.00195.

Vanessa Taler and Natalie Phillips. 2008. Language performance in Alzheimer's disease and mild cognitive impairment: A comparative review. *J Clin Exp Neuropsychol.* 30(5):501-56. doi: 10.1080/13803390701550128.

Susanne Trauzettel-Klosinski, Klaus Dietz & the IReST Study Group. 2012. Standardized Assessment of Reading Performance: The New International Reading Speed Texts IReST. *Investigative Ophthalmol&Visual Sc.* 53:9.

Laszló Tóth et al. 2015. *Automatic Detection of MCI from Spontaneous Speech using ASR*. Interspeech. Germany.

Anders Wallin et al. 2016. The Gothenburg MCI study: Design and distribution of Alzheimers disease and subcortical vascular disease diagnoses from baseline to 6-year follow-up. *J Cer Blood Flow Metab.* 36(1):114-31.

Olof Tyche. 2001. *Subtila språkstörningar hos patienter med diagnosen MCI*. Master's thesis. Karolinska institute, Sweden (In Swedish).

Brian MacWhinney, Davida Fromm, Margaret Forbes and Audrey Holland. 2011. AphasiaBank: Methods for studying discourse. *Aphasiology*. 25 (11), 1286-1307.

Caroline Williams et al. 2010. *The Cambridge Cookie-Theft Corpus: A Corpus of Directed and Spontaneous Speech of Brain-Damaged Patients and Healthy Individuals*. 7th Language Resources and Evaluation (LREC). Pp. 2824-2830. Malta.

Maria Yancheva, Kathleen Fraser and Frank Rudzicz. 2015. *Using linguistic features longitudinally to predict clinical scores for Alzheimer's disease and related dementias*. 6th SLPAT. Pp. 134–139, Dresden, Germany.

Victoria Yaneva, Irina Temnikova and Ruslan Mitkov. 2016. *Corpus of Text Data and Gaze Fixations from Autistic and Non-autistic Adults*. 10th Language Resources and Evaluation (LREC). Pp. 480-487. Slovenia.

Gustaf Öqvist Seimyr. 2010. *Swedish IReST translation*. The Bernadotte Laboratory, Karolinska institute, Sweden.