

Using linguistic features longitudinally to predict clinical scores for Alzheimer’s disease and related dementias

Maria Yancheva¹, Kathleen Fraser¹, Frank Rudzicz^{1,2}

¹Department of Computer Science, University of Toronto, Toronto, Canada

²Toronto Rehabilitation Institute-UHN, Toronto, Canada

yancheva@cs.toronto.edu, kfraser@cs.toronto.edu, frank@cs.toronto.edu

Abstract

We use a set of 477 lexicosyntactic, acoustic, and semantic features extracted from 393 speech samples in DementiaBank to predict clinical MMSE scores, an indicator of the severity of cognitive decline associated with dementia. We use a bivariate dynamic Bayes net to represent the longitudinal progression of observed linguistic features and MMSE scores over time, and obtain a mean absolute error (MAE) of 3.83 in predicting MMSE, comparable to within-subject interrater standard deviation of 3.9 to 4.8 [1]. When focusing on individuals with more longitudinal samples, we improve MAE to 2.91, which suggests at the importance of longitudinal data collection.

Index Terms- Alzheimer’s disease, dementia, Mini-Mental State Examination (MMSE), dynamic Bayes network, feature selection

1. Introduction

Research into the early assessment, pathogenesis, and progression of dementia is becoming increasingly important, as the proportion of people it affects grows every year. Alzheimer’s disease (AD), the most common type of dementia, affects more than half of the population above 80 years of age and its impact on society is expected to grow as the “baby boomer” generation ages [2, 3, 4].

There is no single laboratory test that can identify dementia with absolute certainty. Typically, probable dementia is diagnosed using the Mini Mental State Examination (MMSE), which provides a score on a scale of 0 (greatest cognitive decline) to 30 (no cognitive decline), based on a series of questions in five areas: orientation, registration, attention, memory, and language [5]. While MMSE provides a unified scale for measuring the severity of the disease, it can be time-consuming and relatively costly, often requiring a trained neuropsychologist or physician to administer the test in a clinical setting.

Changes in cognitive ability due to neurodegeneration associated with AD lead to a progressive decline in memory and language quality. Patients experience deterioration in sensory, working, declarative, and non-declarative memory, which leads to a decrease in the grammatical complexity and lexical content of their speech [6]. Such changes differ from the pattern of decline expected in older adults [6], which suggests that temporal changes in linguistic features can aid in disambiguation of healthy older adults from those with dementia.

Some previous work used machine learning classifiers with linguistic features for two-class separation of patients with AD from controls (see section 1.1), but there appears to be no previous research that has used them to infer a clinical score for dementia — an indicator of the degree of cognitive decline. The

present work uses a set of automatically-extracted lexicosyntactic, acoustic, and semantic (LSAS) features for estimating continuous MMSE scores on a scale of 0 to 30, using a dynamic Bayes network for representing relationships between observed linguistic measures and underlying clinical scores.

Since dynamic changes in linguistic ability in patients with AD differ from those in typical healthy older adults [6], we hypothesize that considering speech samples over time would aid in estimating underlying cognitive status. Previous studies analyzing dynamic progression of language features in patients with AD did not employ machine learning techniques, and are characterized by a small number of subjects (between 3 and 6) and a limited set of features that do not include acoustics. The present work improves on these analyses by extracting LSAS features from a relatively large collection of longitudinal speech, in order to estimate MMSE scores.

1.1. Related Work

Previous work has explored the use of lexicosyntactic features for identifying individuals with AD from controls. Orimaye *et al.* [7] used DementiaBank¹, one of the largest existing datasets of pathological speech [8], to perform binary classification of 242 patients with dementia and 242 controls; a support vector machine classifier achieved their best F-measure of 0.74 [7]. Another experiment by Jarrold *et al.* collected spontaneous speech data from 9 controls, 9 patients with AD, and 30 patients with frontotemporal lobar degeneration (FTLD) [9]. A multi-layer perceptron model obtained classification accuracy of 88% on a two-class task (AD:controls, and FTLD:controls), and 80% on a three-class task (AD:FTLD:controls).

While these studies have obtained promising results in classifying patients with dementia based on linguistic features, there is limited work modelling the progression of such features over time. Le *et al.* [10] examined the longitudinal changes in a small set of hand-selected lexicosyntactic measures, such as vocabulary size, repetition, word class deficit, and syntactic complexity, in 57 novels of three British authors written over a period of several decades. They found statistically significant lexical deterioration in Agatha Christie’s work evidenced by vocabulary impoverishment and a pronounced increase in word repetitions [10], but the measures for syntactic complexity did not yield conclusive results. A similar analysis performed by Sundermann examined the progression of a small set of lexicosyntactic features, such as length, frequency, and vocabulary measures in 6 patients with AD or mild cognitive impairment (MCI), with a minimum of 3 longitudinal samples in DementiaBank [11]. Analysis of the features over time did not reveal con-

¹<http://talkbank.org/DementiaBank/>

clusive patterns; Sundermann suggested that the limited sample size and feature set selection may be the cause. Neither study involved acoustics or machine learning techniques.

2. Methodology

2.1. Data

We use data from DementiaBank, a large dataset of speech produced by people with dementia (including probable AD, possible AD, vascular dementia, and MCI) and healthy older adults, recorded longitudinally at the University of Pittsburgh’s Alzheimer’s Disease Research Center [8]. Annual visits with each subject consist of a recording of speech data, its textual transcription, and an MMSE score. Subjects have a variable number of longitudinal samples ($min = 1, max = 5, M = 1.54, SD = 0.79$). Each speech sample consists of a verbal description of the Boston Cookie Theft picture, which typically lasts about a minute. We partition subjects between controls (CT) and those with probable AD, possible AD or MCI (or, collectively, “AD”²). Considering only subjects with associated MMSE scores, the working set consists of 393 speech samples from 255 subjects (165 AD, 90 CT).

2.2. Features

Three major types of features are extracted from the speech samples and their transcriptions: (1) *lexicosyntactic measures*, extracted from syntactic parse trees constructed with the Brown parser and POS-tagged transcriptions of the narratives [12, 13, 14, 15, 16]; (2) *acoustic measures*, including the standard Mel-frequency cepstral coefficients (MFCCs), formant features, and measures of disruptions in vocal fold vibration regularity [17]; and (3) *semantic measures*, pertaining to the ability to describe concepts and objects in the Cookie Theft picture. The full list of features, along with their major type and subtype, is shown in Table 1.

2.3. Feature Analysis

Two feature selection methods are used to identify the most informative features for disambiguating AD from CT. Since the MMSE score is a measure of the progression of cognitive impairment and is used to distinguish AD from CT generally, we hypothesize that highly discriminating features of the two groups would also be good predictors of MMSE. This is quantified by Spearman’s rank-order correlation between the most informative features and the MMSE score, ρ_{MMSE} , shown in Table 2.

The first feature ranking method is a two-sample t -test ($\alpha = 0.001$, two-tailed) which quantifies the significance of the difference in each feature value between the two classes; the features are ordered by increasing p -value. Table 2 shows the type and p -value of the top 10 features, along with their correlation with MMSE. Control subjects use longer utterances, more gerund + prepositional phrase constructions (VP \rightarrow VBG PP, e.g., *standing on the chair*), more content words such as noun phrases (NP) and verbs, and are more likely to talk about what they see through the window (info_window), which is in the background of the scene (e.g., *it seems to be summer out*). On the other hand, subjects with AD use more words not found in the dictionary (NID), and more function words such as pronouns (PRP). Honoré’s statistic measures lexical richness, ex-

tending type-token ratio, which is decreased in AD. These findings are consistent with expectations.

Table 2: The top 10 features selected by a two-sample t -test ($\alpha = 0.001$, two-tailed) as the most informative discriminators of AD versus CT. ρ_{MMSE} is Spearman’s rank-order correlation coefficient between the given feature and the MMSE score. The features in **bold** are among the top 10 selected by mRMR.

Feature	Feature type	p	ρ_{MMSE}
avelength	lexicosyntactic	1.24E-13	0.3837
VP \rightarrow VBG PP	lexicosyntactic	1.90E-13	0.3757
NID	lexicosyntactic	3.23E-11	-0.3712
NP \rightarrow DT NN	lexicosyntactic	1.12E-10	0.3438
NP \rightarrow PRP	lexicosyntactic	2.14E-10	-0.3186
prp_ratio	lexicosyntactic	1.16E-09	-0.3089
honoré	lexicosyntactic	2.53E-09	0.3400
verbs	lexicosyntactic	4.81E-09	0.2604
frequency	lexicosyntactic	9.37E-09	-0.3725
info_window	semantic	1.27E-08	0.3420

Since the majority of the extracted acoustic features consist of MFCCs and measures related to aperiodicity of vocal fold vibration, the lack of significance of the acoustic features as discriminators between the two classes may be attributed to the fact that AD is not strongly associated with motor impairment of the articulators involved in speech production.

The second feature selection method is minimum-redundancy-maximum-relevance (mRMR), which minimizes the average mutual information between features and maximizes the mutual information between each feature and the class [18]; the features were ranked from most relevant to least. The results of this technique generally corroborate the selection made by the t -test, with no acoustic features among the top 10 selected. Here, mRMR selects a greater proportion of semantic features (e.g., mentions of the window and sink, and the number of occurrences of *curtain* and *stool*), placing more weight on the content of what the speaker is saying as a way of discriminating the two classes.

All of the features displayed in Table 2 have moderate statistically significant correlation with MMSE ($p < 0.001$). Since we are interested in the task of predicting clinical MMSE scores, the experiments described in Sec. 3 use correlation itself as a third feature selection method. The features are ranked by their correlation with MMSE, and the ones with the highest correlations are selected.

3. Experiments

3.1. Predicting MMSE score using LSAS features

To model the longitudinal progression of MMSE scores and LSAS features, we constructed a dynamic Bayes network (DBN) with continuous nodes, i.e., a Kalman filter with 2 variables, shown in Figure 1. Each time slice (Q_t, Y_t) represents one annual visit for a subject. Each conditioning node Q_t represents the underlying continuous MMSE score for that visit ($\mathbb{R}^{1 \times 1}$), while each node Y_t represents the vector of observed continuous LSAS features ($\mathbb{R}^{477 \times 1}$). A Kolmogorov-Smirnov test for normality was performed on the MMSE scores of all AD subjects, with the null hypothesis that they come from a normal distribution. The test did not reject this null hypothesis at the 5% confidence level, demonstrating that the data come from a

²Ongoing work distinguishes between AD and MCI.

Table 1: Summary of all extracted features (477 in total). The number of features in each type and subtype is shown in parentheses.

Type	Feature Subtype	Description and examples
Lexicosyntactic (182)	Production rule (121)	Number of times a production rule is used, divided by the total number of productions.
	Phrase type (9)	Phrase type proportion, rate and mean length.
	Syntactic complexity (4)	Depth of the syntactic parse tree.
	Subordination/coordination (3)	Proportion of subordinate and coordinate phrases to the total number of phrases, and ratio of subordinate to coordinate phrases.
	Word type (25)	Word type proportion; type-to-token ratio, Honoré’s statistic.
	Word quality (10)	Imageability; age of acquisition (AoA); familiarity; transitivity.
	Length measures (5)	Average length of utterance, T-unit and clause, and total words per transcript.
	Perseveration (5)	Cosine distance between pairs of utterances within a transcript.
Acoustic (210)	MFCCs (170)	The first 42 MFCC parameters, along with their means, kurtosis and skewness, and the kurtosis and skewness of the mean of means.
	Pauses and fillers (8)	Total and mean duration of pauses; long and short pause counts; pause to word ratio; fillers (<i>um</i> , <i>uh</i>).
	Pitch and Formants (8)	Mean and variance of F0, F1, F2, F3.
	Aperiodicity (13)	Jitter, shimmer, recurrence rate, recurrence period density entropy, determinism, length of diagonal structures, laminarity.
	Other speech measures (11)	Total duration of speech, zero-crossing rate, autocorrelation, linear prediction coefficients, transitivity.
Sem. (85)	Mention of a concept (21)	Presence of mentions of indicator lemmas, related to key concepts in the Cookie Theft picture.
	Word frequency (64)	Number of times a given lemmatized word, relating to the Cookie Theft picture, was mentioned

normal distribution with $M=18.52$, $SD=5.16$. There are three conditional probability densities: the MMSE prior probability $P(Q_1)$, the MMSE transition probability $P(Q_t|Q_{t-1})$, and the LSAS feature observation probability $P(Y_t|Q_t)$.

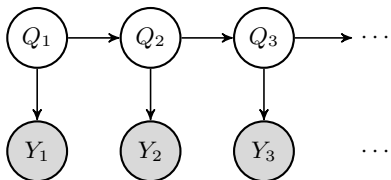


Figure 1: Temporal Bayes network (TBN) with continuous hidden (Q_t) and observed (Y_t) nodes. Hidden nodes represent MMSE score, and observed vectors represent LSAS features extracted from speech.

The feature set described in Sec. 2.2 is preprocessed to (i) remove features with zero variance across all samples, and (ii) normalize feature values to zero-mean and unit-variance, as is standard practice. Since the number of features (477) is large compared to the number of samples (393), the three feature selection methods described in Sec. 2.3 (i.e., a paired two-tailed t -test, mRMR, and correlation with MMSE score) are used to avoid overfitting, by varying the number of features selected by each method in order to determine the optimal feature set size.

The parameters of the three probability distributions in our model are trained using maximum likelihood estimation (MLE) since all training data are fully observed. During testing, the observed features for each test case are provided and junction

tree inference on the trained model computes the marginal distribution of the now hidden (MMSE) nodes. Performance is measured as the mean absolute error (MAE) between actual and predicted MMSE scores. Since not all subjects have the same number of longitudinal samples, MAE is evaluated at the first and last hidden node, and averaged. Experiments are performed with leave-one-out cross-validation, where data from each subject, in turn, are used for testing and all other data for training, over all 255 subjects.

The results, with varying feature set sizes and feature selection methods, are shown in Table 3. The lowest MAE of 3.83 ($\sigma = 0.49$) is achieved when correlation is used to select the top 40 features. A two-factor repeated measures ANOVA performed on the mean MAE shows that both main effects are statistically significant, i.e., feature set size ($F_{7,24} = 8.67$, $p < 0.001$) and the feature selection method ($F_{2,24} = 4.07$, $p < 0.05$). The interaction effect is not significant ($F_{14,24} = 0.16$, ns), as expected given that the factors are independent.

To illustrate the longitudinal changes in cognitive and linguistic ability, Fig. 2 shows the pattern of decline of MMSE and the top 5 most correlated features for the subset of subjects with AD. This demonstrates the MMSE score declining non-monotonically over four annual visits (the maximum number of visits for AD subjects in DementiaBank), along with similar patterns across the indicated LSAS features.

3.2. Effect of longitudinal data on predicted MMSE score

To test the hypothesis that using longitudinal speech data aids in identifying underlying cognitive status (i.e., improving MMSE estimation), the Kalman filter experiment described in 3.1 is repeated for subsets of the dataset consisting of different amounts

Table 3: MAE in predicting MMSE scores using three feature selection methods and different feature set sizes. The lowest error for each feature selection method is highlighted in **bold**.

$N_{features}$	t -test	mRMR	ρ_{MMSE}
1	5.9788	5.3034	5.6396
5	5.6575	4.4440	5.0758
10	5.5148	4.3403	4.2098
20	5.2264	4.0426	4.1518
30	4.9066	4.1420	3.8628
40	4.8073	4.0648	3.8333
50	4.8520	3.8551	3.9180
all	7.3106	7.3106	7.3106

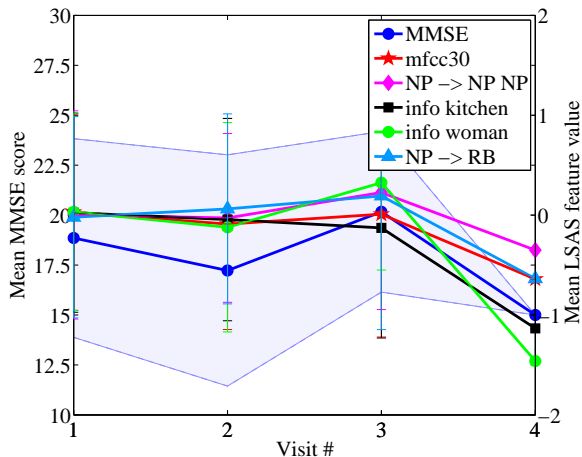


Figure 2: Pattern of decline of mean MMSE score and top 5 LSAS features most correlated with it, plotted versus annual visit number, for the subset of subjects with AD in DementiaBank. Standard deviation for MMSE is shown shaded in blue.

of longitudinal samples, T : (i) entire dataset (393 samples, 255 subjects, $1 \leq T \leq 5$), (ii) subset of subjects with 1 visit (154 samples, 154 subjects, $T = 1$), (iii) subset of subjects with at least two visits (239 samples, 101 subjects, $T \geq 2$), and (iv) subset of subjects with at least three visits (91 samples, 27 subjects, $T \geq 3$). The number of subjects with at least four visits is too low to conduct statistical experiments. The number of features used in the model is fixed to the optimal feature set size found in 3.1, and the feature selection method is varied (t -test, mRMR, correlation). Leave-one-out cross-validation is performed on each of the four datasets. The results are presented in Table 4. The lowest MAE for each feature selection method occurs on the dataset consisting of the highest number of longitudinal visits ($T \geq 3$). A two-factor repeated measures ANOVA performed on the mean MAE shows that the main effect of the data subset is statistically significant ($F_{3,9} = 5.43$, $p < 0.05$) while neither the second main effect ($F_{2,9} = 0.94$, ns) nor the interaction effect ($F_{6,9} = 0.54$, ns) is significant.

4. Discussion

Automatically extracted linguistic features can be used to effectively estimate underlying cognitive status, in terms of the most predominant clinical measure of dementia. The best result obtained with leave-one-out cross-validation on the entire dataset of 393 samples is an MAE of 3.83 ($\sigma = 0.49$), using

Table 4: MAE in predicting MMSE score using three feature selection methods and different subsets of subjects with varied number of longitudinal datapoints. The lowest error for each feature selection method is highlighted in **bold**.

Dataset	t -test	mRMR	ρ_{MMSE}
all	4.807311	4.064823	3.8332502
1 visit	5.030811	4.978016	4.4916474
≥ 2 visits	4.334934	3.534478	3.430414
≥ 3 visits	2.905163	3.063524	3.3577102

correlation to select the top 40 features. This corresponds to a mean absolute relative error (MARE) of 21.0% (obtained as the absolute difference between predicted and actual MMSE score, divided by the actual MMSE score, and averaged over all runs). Molloy and Standish [19] reported that different rating styles among clinicians administering the MMSE and variance in test-retest scoring can lead to a within-subject interrater standard deviation of 3.9 to 4.8 and within-subject intrarater standard deviation of 4.8, with higher variation in low-scoring subgroups of subjects [1, 19]. The MAE obtained through statistical speech analysis in our present work is comparable to such variability. Further, the results obtained with the Kalman filter model significantly outperform an initial baseline multilinear regressor ran with leave-one-out cross-validation on the same dataset ($t = 2.31$, $p < 0.05$). This is being explored further.

The fact that correlation outperforms the other two feature selection methods is expected, as it computes the relationship between the features and the MMSE score directly whereas the others use the presumed diagnosis to dichotomize the data into classes. The majority of features selected on each iteration of cross-validation are typically lexicosyntactic and semantic, with acoustic features typically not being among the most relevant. While this may suggest that anatomical irregularities in speech production are less meaningful, we note that the lexicosyntactic features depend, to a large extent, on the free expression of language through speech. Specifically, the working memory impairment associated with AD affects preferred syntactic constructions in speech, leading to shorter utterances, fewer complex noun and verb phrases, a higher number of pronouns, and lexical impoverishment indicated by Honoré’s statistic.

We also show that focussing on subsets of subjects with a higher number of longitudinal samples improves the accuracy of inference in the Kalman filter model, lowering MAE to 2.91 ($\sigma = 0.31$) or equivalently lowering MARE to 12.5%, using a t -test for selecting the top 40 features. Since DementiaBank contains a variable number of samples for each subject, the number of subjects and the proportion of subjects with AD in each subgroup explored in Sec. 3.2 is not balanced. We therefore suggest that future data collection of pathological speech should involve more longitudinal samples across participants.

While MMSE is one of the most widely used clinical tests for cognitive ability, it is somewhat coarse, lacking sensitivity to subtle changes in cognition in the early stages of dementia, as well as having a high false-negative rate in addition to inter-annotator disagreement and test-retest variability [20, 1, 19]. While automated prediction of the MMSE score may aid the screening process for AD by reducing the cost and time involved, and improving reliability, future work will explore more precise measures of cognitive decline. The Montreal Cognitive Assessment (MoCA) and the Repeatable Battery for the Assessment of Neuropsychological Status (RBANS) [21] are screening tests which have been shown to have higher sensitivity than

MMSE to subtle changes in cognitive decline in populations with MCI and mild dementia [22]; future studies are needed to assess the validity of automatic scoring of such tests as a more fine-grained measure of the progression of cognitive decline.

5. Acknowledgements

This work is funded by an NSERC Discovery grant (RGPIN 435874) and by a Young Investigator award by the Alzheimer Society of Canada.

6. References

- [1] D. W. Molloy, M. B. E. Alemaeyehu, and R. Roberts, "Reliability of a Standardized Mini-Mental State Examination compared with the traditional Mini-Mental State Examination," *The American Journal of Psychiatry*, vol. 148, no. 1, pp. 102–105, 1991.
- [2] C. Ballard, S. Gauthier, A. Corbett, C. Brayne, D. Aarsland, and E. Jones, "Alzheimer's disease," *The Lancet*, vol. 377, no. 9770, pp. 1019–1031, 2011.
- [3] R. M. Li, A. C. Iadarola, and C. C. Maisano. (2007) Why population aging matters: A global perspective. [Online]. Available: <http://www.nia.nih.gov/research/publication/why-population-aging-matters-global-perspective>
- [4] R. Sperling, P. Aisen, L. Beckett, D. Bennett, S. Craft, A. Fagan, T. Iwatsubo, C. R. J. Jr., J. Kaye, T. Montine, D. Park, E. Reiman, C. C. Rowe, E. Siemers, Y. Stern, K. Yaffe, M. C. Carrillo, B. Thies, M. Morrison-Bogorad, M. V. Wagster, and C. H. Phelps, "Toward defining the preclinical stages of Alzheimer's disease: recommendations from the National Institute on Aging—Alzheimer's Association workgroups on diagnostic guidelines," *Alzheimer's and Dementia*, vol. 7, no. 3, pp. 280–292, 2011.
- [5] M. Folstein, S. E. Folstein, and P. R. McHugh, "Mini-Mental State: a practical method for grading the cognitive state of patients for the clinician," *Journal of Psychiatric Research*, vol. 12, no. 3, pp. 189–198, 1975.
- [6] S. Kemper, M. Thomas, and J. Marquis, "Longitudinal change in language production: Effects of aging and dementia on grammatical complexity and propositional content," *Psychology and Aging*, vol. 16, no. 4, pp. 600–614, 2001.
- [7] S. Orimaye, J. S.-M. Wang, and K. J. Golden, "Learning predictive linguistic features for Alzheimer's disease and related dementias using verbal utterances," in *Proc. of the ACL 2014 Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality*, Baltimore, USA, Jun. 2014, pp. 78–87.
- [8] H. Goodglass and E. Kaplan, *The Assessment of Aphasia and Related Disorders*. Philadelphia, PA: Lea and Febiger, 1983.
- [9] W. Jarrold, B. Peintner, D. Wilkins, D. Vergryi, C. Richey, M. L. Gorno-Tempini, and J. Ogar, "Aided diagnosis of dementia type through computer-based analysis of spontaneous speech," in *Proc. of the ACL 2014 Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality*, Baltimore, USA, 2014, pp. 27–37.
- [10] X. Le, I. Lancashire, G. Hirst, and R. Jokel, "Longitudinal detection of dementia through lexical and syntactic changes in writing: a case study of three British novelists," *Literary and Linguistic Computing*, vol. 26, no. 4, pp. 435–461, 2011.
- [11] M. Sundermann, "Longitudinal effects of Alzheimer's disease," Master's thesis, Department of Speech and Hearing Science, Ohio State University, 2012.
- [12] H. Bird, M. A. L. Ralph, K. Patterson, and J. R. Hodges, "The rise and fall of frequency and imageability: Noun and verb production in semantic dementia," *Brain and Language*, vol. 73, pp. 17–49, 2000.
- [13] K. Gilhooly and R. Logie, "Age-of-acquisition, imagery, concreteness, familiarity, and ambiguity measures for 1,944 words," *Behavior Research Methods*, vol. 12, pp. 395–428, 1980.
- [14] D. Klein and C. D. Manning, "Accurate unlexicalized parsing," in *Proc. of the 41st Meeting of the Association for Computational Linguistics*, 2003, pp. 423–430.
- [15] J. O. de Lira, K. Z. Ortiz, A. C. Campanha, P. H. F. Bertolucci, and T. S. C. Minett, "Microlinguistic aspects of the oral narrative in patients with Alzheimer's disease," *International Psychogeriatrics*, vol. 23, no. 3, pp. 404–412, 2011.
- [16] H. Stadthagen-Gonzalez and C. J. Davis, "The Bristol norms for age of acquisition, imageability, and familiarity," *Behavior Research Methods*, vol. 38, no. 4, pp. 598–605, 2006.
- [17] A. Tsanas, M. A. Little, P. E. McSharry, J. Spielman, and L. O. Ramig, "Novel speech signal processing algorithms for high-accuracy classification of Parkinson's disease," *IEEE Trans Biomed Eng.*, vol. 59, no. 5, pp. 1264–1271, 2011.
- [18] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy," *Pattern Analysis and Machine Intelligence, IEEE Transactions*, vol. 27, no. 8, pp. 1226–1238, 2005.
- [19] D. W. Molloy and T. I. M. Standish, "A guide to the Standardized Mini-Mental State Examination," *International Psychogeriatrics*, vol. 9, no. 1, pp. 87–94, 1997.
- [20] A. Nelson, B. Fogel, and D. Faust, "Bedside cognitive screening instruments — a critical assessment," *The Journal of Nervous and Mental Disease*, vol. 174, no. 2, pp. 73–83, 1986.
- [21] C. Randolph, *Repeatable Battery for the Assessment of Neuropsychological Status Update*. San Antonio, TX: The Psychological Corporation, 2012.
- [22] C. Zadikoff, S. Fox, D. Tang-Wai, T. Thomsen, R. de Bie, P. Wadia, J. Miyasaki, S. Duff-Canning, A. Lang, and C. Marras, "A comparison of the Mini Mental State Exam to the Montreal Cognitive Assessment in identifying cognitive deficits in Parkinson's disease," *Movement Disorders*, vol. 23, no. 2, pp. 297–299, 2008.
- [23] D. Bone, T. Chaspari, K. Audkhasi, J. Gibson, A. Tsiartas, M. V. Segbroeck, M. Li, S. Lee, and S. Narayanan, "Classifying language-related developmental disorders from speech cues: the promise and the potential confounds," in *Proc. of the 14th Annual Conference of the International Speech Communication Association*, 2013, pp. 182–186.

- [24] L. Meteyard and K. Patterson, "The relation between content and structure in language production: an analysis of speech errors in semantic dementia," *Brain and Language*, vol. 110, no. 3, pp. 121–134, 2009.
- [25] B. Roark, M. Mitchell, J.-P. Hosom, K. Hollingshead, and J. Kaye, "Spoken language derived measures for detecting mild cognitive impairment," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 19, no. 7, pp. 2081–2090, 2011.
- [26] D. G. Silva, L. C. Oliveira, and M. Andrea, "Jitter estimation algorithms for detection of pathological voices," *EURASIP Journal on Advances in Speech Processing*, vol. 2009, pp. 1–9, 2009.