Domain Adaptation for Detecting Mild Cognitive Impairment

Vaden Masrani[1], Gabriel Murray[2], Thalia Shoshana Field[3], and Giuseppe Carenini[1]

[1] Department of Computer Science, University of British Columbia, Vancouver, Canada
[2] Department of Computer Information Systems, University of the Fraser Valley, Abbotsford, Canada
[3] Department of Neurology, University of British Columbia, Vancouver, Canada
vadmas@cs.ubc.ca, gabriel.murray@ufv.ca, thalia.field@ubc.ca, carenini@cs.ubc.ca

Abstract. Lexical and acoustic markers in spoken language can be used to detect mild cognitive impairment (MCI), a condition which is often a precursor to dementia and frequently causes some degree of dysphasia. Research to develop such a diagnostic tool for clinicians has been hindered by the scarcity of available data. This work uses *domain adaptation* to adapt Alzheimer's data to improve classification accuracy of MCI. We evaluate two simple domain adaptation algorithms, AUGMENT and CORAL, and show that AUGMENT improves upon all baselines. Additionally we investigate the use of previously unconsidered *discourse features* and show they are not useful in distinguishing MCI from healthy controls.

Keywords: Domain Adaptation, Mild Cognitive Impairment, Dementia, Alzheimer's

1 Introduction

Mild cognitive impairment (MCI) is a non-specific diagnosis characterized by cognitive decline that is less severe than dementia and does not significantly interfere with activities of daily living [1]. Population-based studies estimate its prevalence to be between 12-18% in people over the age of 60 [2]. While a proportion of patients will revert to normal cognition or stay mildly impaired, 8-15% annually will progress to dementia [2]. MCI can be due to neurodegenerative (most commonly, but not exclusively, Alzheimer's dementia(AD)) or reversible causes, including psychiatric illness or metabolic disturbances including thyroid disease or vitamin B12 deficiency [3].

As MCI can involve a number of potential underlying causes, there are no specific treatments. However, early diagnosis can affect testing for potentially treatable causes, allow for optimization of vascular risk factors that may accelerate onset of dementia, prompt further diagnostic testing, and better allow for planning for social supports and closer medical follow-up.

With recent improvements in natural language processing (NLP) and machine learning, there has been a push to use speech processing to develop a tool to assist clinicians in diagnosis of "Alzheimer's disease and other dementias" (ADOD)[4]. Clinical research has shown that dysphasia is common among ADOD. Weiner et al. found a significant association between cognitive test scores and multiple language measures, including language fluency, animal naming and repetition [4–6]. However, building a similar diagnostic tool for MCI presents a challenge due to the limitations involved in data collection. MCI data is difficult to acquire due to limited time resources available for detailed assessment in primary care settings [7], insufficient sensitivity to MCI of screening tools such as the Mini Mental Status Exam [8], and limited access to primary care.

A technique that can be used to address this challenge is *domain adaptation* (also known as "transfer learning"). The situation often arises when one has a limited amount of data related to a problem of interest (the "target domain"), but a large amount of data from a separate but related problem (the "source domain"). Domain adaptation is the task of leveraging ("adapting") data from the source domain so that it can be used in the target domain. More specifically, we wish to use data collected from persons with AD to improve our classification accuracy for persons with MCI.

The main contribution of this work is to demonstrate the efficacy of domain adaptation in using AD data to improve our ability to diagnose MCI. We compare two domain adaptation algorithms, AUGMENT and CORAL, both of which are simple to implement and have been successfully applied on a range of datasets, and show AUGMENT improving upon all baselines. These algorithms are discussed in detail in section 2.3. A secondary contribution of this work is in testing the usefulness of a new set of "discourse" features, as discussed in section 2.2, which have not been used in previous work in the area. We show discourse features surprisingly do not substantially improve the final classification performance.

2 Related Work

2.1 Diagnosing Dementia from Speech

There has been a recent interest in using lexical and acoustic features derived from speech to diagnose ADOD. In 2013 Ahmed et al. [9] determined features that could be used to identify dementia from speech, using data collected in the Oxford Project to Investigate Memory and Ageing (OPTIMA) study. They used a British cohort of 30 participants, 15 with Alzheimers disease at either mild cognitive impairment or mild stages, and 15 whose age and education matched healthy controls. Ahmed et al. found that language progressively deteriorates as Alzheimer's disease (AD) progresses and suggested using semantic, lexical content, and syntactic complexity features to identify cases. Rentoumi et al. [10] then used a Naive Bayes Gaussian Classifier with lexical and syntactic features

[4] http://www.alz.org/greaterdallas/documents/AlzOtherDementias.pdf

to distinguish between AD with and without additional vascular pathology. They achieved a classification accuracy of 75% on 36 transcripts from the OPTIMA dataset.

In 2014 Fraser et al. [11] compared different feature sets that could be used in discriminating between three different types of primary progressive aphasia, a form of Frontotemporal Dementia, which is a rarer cause of dementia than AD with a distinct disease course. They concluded that a smaller relevant subset of features achieves better classification accuracy than using all features and highlighted the importance of a feature selection step. They also showed how psycholinguistic features, such as frequency and familiarity, were useful in detecting aphasic dementia. In later work Fraser [12] achieved state-of-the-art of 81.92% in distinguishing individuals with AD from those without using logistic regression. Fraser used DementiaBank, an American cohort of 204 persons with dementia and 102 controls, and performed factor analysis on a set of 370 lexical and acoustic features, finding optimal performance when 35-50 features are used[13].

Roark et al. [14] did the largest study to date classifying MCI from speech, using transcripts and audio recordings of patients undergoing the Wechsler Logical Memory I/II test. This test involves a patient twice retelling a short story, once immediately after hearing it and again after a 30 minute delay. Roark extracted two broad set of features; "linguistic complexity" features which measure the complexity of a narrative, and "speech duration" features including number of pauses, pause length and pause-to-speech ratio. Using SVM's, they achieved a maximum AUC of 0.74 and concluded that NLP techniques could be used to automatically derive measures to discriminate between healthy and MCI subjects.

Our work differs from the previous work done in this area in a number of ways. Unlike Roark, we are using MCI data collected from DementiaBank, described in section 4.1, where patients undergo a picture description task rather than a narrative retelling task. We are also using a larger feature set proposed by Fraser et al., to which we are adding the "discourse features" described in sections 2.2 and 4.2. Most significantly, we are applying two domain adaptation algorithms which aim to use data collected from AD patients to improve the diagnosis of MCI. Our goal is not to improve upon the accuracy of previous work but to demonstrate the viability of domain adaption in this setting.

2.2 Discourse Analysis

One measure of coherence which has been neglected in the aforementioned work comes from *discourse analysis.* In a coherent passage, a reader can clearly discern how one sentence relates to the next. A given sentence may *explain* or *elaborate* upon a previous sentence (as this one is doing), or act as *background* for a future sentence. Such relations can be formed on an intra-sentential level as well, with *elementary discourse units* (EDU's) being clause-like units of text which can be related to one another by *discourse relations.* Discourse parsing is the task of segmenting a piece of text into its EDU's and then forming a *discourse tree* with edges corresponding to discourse relations, as seen in Figure 1. Features related to the discourse structure of a passage can then be extracted from the discourse tree, as discussed in section 4.2.

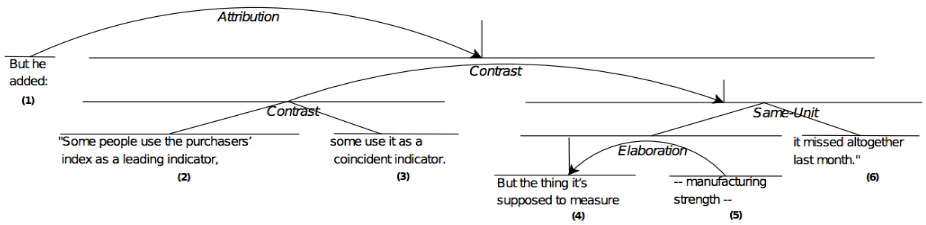

Fig. 1. Discourse tree for two sentences. Each sentence contains three EDUs. EDUs correspond to leaves of the tree and discourse relations correspond to edges. (Figure adapted from [15])

Previous work has shown a disparity in the overall discourse ability of patients with ADOD compared to healthy controls [16–18]. Those with ADOD show a greater impairment in global coherence, have more disruptive topic shift, greater use of empty phrases, and produce fewer cohesive ties than controls.[19–22]. Discourse parsing has been useful in determining overall coherence in other domains such as essay scoring, and so we hypothesize that it will also be useful for MCI detection [23].

2.3 Domain Adaptation

Domain adaptation is a general term for a variety of techniques that aim to exploit resources in one domain (the *source* domain) in order to improve performance on some task in a second domain (the *target* domain). This is typically done when the target domain has little or no labelled data, while the source domain has a relatively large amount of labelled data, as well as existing models trained on that data. Typically the source data have been annotated for some phenomenon of interest, and the target data relate to another phenomenon that is highly similar in nature.

The issue of domain adaptation has received increasing attention in recent years. In work by Chelba and Acero [24], the source model is used to derive priors for the weights of the target model. They employ this technique with a maximum entropy model and apply it to the task of automatic capitalization of uniformly-cased data. They report that adaptation yields a relative improvement of 25-30% in the target domain.

Blitzer et al. [25] introduced Structural Correspondence Learning (SCL), in which relationships between features in the two domains are determined by finding correlations with so-called *pivot* features, which are features exhibiting similar behaviour in both domains. They used SCL to improve the performance of a parser applied to Biomedical data, but trained on Wall Street Journal data.

Daume [26] introduced an approach wherein each feature is copied so that there is a source version, a target version and a general version of the feature. More recently, Sun [27] proposed CORAL, a method which aligns the second-order statistics of the source and target domain. We have implemented these two approaches, and describe them in more detail in Section 3.

3 Domain Adaptation

3.1 Baselines

We describe two domain adaptation algorithms below, and compare against four baselines. *Majority class* predicts the majority class, *target only* trains the model only using target data, *source only* trains a model only using source data but evaluates on the target data. In the *relabeled source* model, we pool the target data and source data in the training folds and relabel AD to MCI.

3.2 Frustratingly Simple

Daume III's AUGMENT domain adaptation algorithm is simple ("frustratingly" so) and has been shown to be effective on a wide range of datasets [26]. It augments the feature space by making a "source-only", "target-only", and "common" copy of each feature, as seen below.

$$\underset{(n\times d)}{\begin{bmatrix} X_s \\ X_t \end{bmatrix}} \Rightarrow \underset{(n\times 3d)}{\begin{bmatrix} X_s & 0 & X_s \\ X_t & X_t & 0 \end{bmatrix}} \tag{1}$$

Here $X_s \in \mathbb{R}^{n_s \times d}$ and $X_t \in \mathbb{R}^{n_t \times d}$ are matrices of source and target data, where each of the n rows is an observation, each of the d column is a feature, $n = n_t + n_s$ and $n_t \ll n_s$. We create three copies of each column: a source-only column with zeros in target rows, a target-only column with zeros in source rows, and the original column with both target and source entries left untouched. This augmented dataset is then fed to a standard learning algorithm.

The motivation for this transformation is intuitive. If a column contains a feature (eg. mean word length) which correlates to a diagnosis in both the target and source data (eg. MCI and Alzheimer's), a learning algorithm will increase the weight in the common column and reduce the weight on target-only and source-only copies, thereby reducing their importance in the model. However, if a feature correlates to a diagnosis only with MCI data, a learning algorithm can increase the weight of the target-only column (which contains zeros for all the source data) and reduce weight of the original and source-only columns, thereby assuring the feature will be less relevant to the model when applied to Alzheimer's data. By expanding the feature space and padding with zeros, we allow a model to learn whether to apply a given feature on zero, one, or both datasets.

3.3 CORAL

CORAL (CORrelation ALignment) is another recently proposed "frustratingly easy" domain adaptation algorithm which works by aligning the covariances of the source and target features [27]. The algorithm first normalizes the source data to zero mean and unit variance, and then a whitening transform[5] is performed on the source data to remove the correlation between the source features. Finally, the source matrix is "recoloured" with the correlations from the target data. These three steps are shown in Figure 2. A model is then trained on the recoloured source data and used to classify the target data.

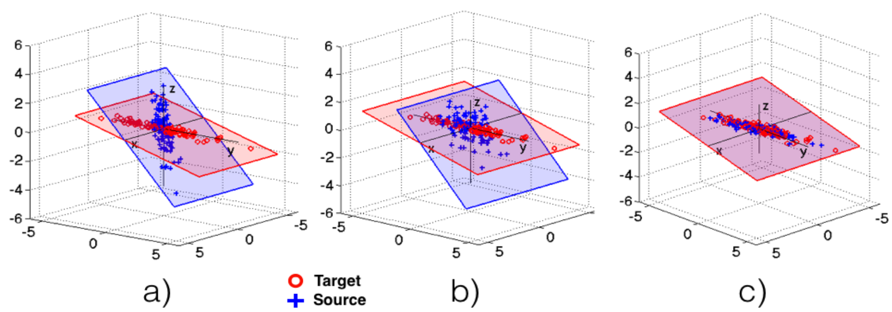

Fig. 2. The CORAL algorithm is shown in three steps. The target and source dataset consist of three features; x, y, z. In a) the source data and target data are normalized to unit variance and zero mean, but have difference covariances distributions. b) The source data is whitened to remove the correlations between features. c) The source data is recolored with the target domain's correlations and the two datasets are aligned. A classifier is then trained on the re-aligned source data. (Figure adapted from [27])

4 Experimental Design

An overview of our experimental design is as follows. We chose logistic regression with l2 regularization as a model, which has been used successfully in previous work on detecting Alzheimer's [12]. We trained the model using a 10-fold cross validation procedure. Within each fold, we first separated 10% of the data to be used as a test set, assuring that if a patient has multiple interviews, those interviews appeared either in the training set or the test set but not both. Then, before training the model, we ran a feature selection step where we selected for inclusion into the model only those features which have highest correlation (positive or negative) with the labels in the training set. We were interested to

[5] We used ZCA whitening which is discussed in greater detail here: http://ufldl.stanford.edu/wiki/index.php/Whitening

see how model accuracy varied as a function of k, the number of features fed to the model, so we trained a model for each value of k up to the total number of features. This entire procedure was repeated for each of the 10 folds, and we report the highest average F-Measure across all k.

With the AUGMENT, CORAL, and *relabeled* approaches, each fold of the training set contains a combination of MCI+AD data and the test set contains only MCI data. Our goal was to verify whether the accuracy achieved by using these domain adaptation methods outperforms the accuracy achieved by using MCI data alone. A secondary goal was to evaluate the effect of "discourse features" (described below), which have not previously been applied in dementia classification.

4.1 Corpora

We used the DementiaBank dataset, a publicly available dataset which consists of transcripts and recordings of English-speaking participants describing the "Cookie Theft Picture", a component of the Boston Diagnostic Aphasia Examination [28]. A patient is asked to describe a cartoon image and their answer is manually transcribed, including false starts, pauses, and paraphasia, and segmented into utterances, where an utterance is defined as a unit of speech bounded by silence.

DementiaBank consists of 309 samples from 208 persons with dementia and 242 samples from 102 normal elderly controls (age 45-90). Of the 309 interviews with dementia patients, 43 were classified as MCI and 256 as possible/probable AD. The remaining interviews were not used in this study. We split the DementiaBank dataset into target (MCI) and source (AD) data, where the target data contains 86 rows (43 MCI, 41 control) and the source data contains 458 rows (236 probable AD, 21 possible AD, 201 control). Interviews from a single control were contained in either the target or the source datasets, but not both.

4.2 Classification Features

In addition to the age of the patient, which is a known predictor of dementia [29], We used a total of 353 lexical and acoustic features which can be divided into nine groups. The first eight have been used in previous work [12].

- **Parts-of-speech:** We use the Stanford Tagger[6] to capture the frequency of various parts of speech tags (nouns, verbs, adjectives, adverbs, pronouns, determiners, etc). Frequency counts are normalized by the number of words in the transcript. We also count disfluencies ("um", "er", "ah"), not-in-dictionary words of three or more letters, and word-type ratios (noun to verb, pronoun to noun, etc).
- **Context-free-grammar rules:** Features which count how often a phrase structure rule occurs in an utterance, including NP→VP PP, NP→DT NP, etc. Parse trees come from the Stanford parser.

[6] Available at: http://nlp.stanford.edu/software/tagger.shtml

– **Syntactic Complexity:** Features which measure the complexity of an utterance through metrics such as the depth of the parse tree, mean length of word, sentences, T-Units and clauses and clauses per sentence.
– **Vocabulary Richness:** We calculated various metrics which capture the range of vocabulary in a text, include type-token ratio, Brunet's index, Honore's statistic, and the moving-average type-token ratio (MATTR) [30].
– **Psycholinguistic:** Psycholinguistic features are linguistic properties of words that effect word processing and learnability [31]. We used five psycholinguisic features, *Familiarity*, *Concreteness*, *Imagability*, *Age of acquisition* and *SUBTL*, which measures the frequency with which a word is used in daily life [32].
– **Content words:** Croisile et al. [33] compiled a list of 23 items which can be discerned in the Cookie Theft Picture. These "information units" can be either actions or nouns and examples include "jar", "cookie", "boy", "kitchen", "boy taking" and "woman drying". For each information unit we extracted two features; a binary feature indicating whether the subject has mentioned the item (or one of its synonyms in WordNet), and a frequency count of how many times an item has been mentioned.
– **Repetitiveness:** We vectorized the utterances using TF-IDF and measure the cosine similarity between utterances. We then recorded the mean cosine distance, the average cosine distance, and proportion of distances below three thresholds (0, 0.3, 0.5).
– **Acoustic:** We calculated the mean, variance, skewness, and kurtosis of the first 14 mel-frequency cepstral coefficients (MFCCs), representing spectral information from the speech signal.

In addition to the features considered in previous work, we also perform a *discourse analysis* on the transcripts as described in Section 2.2.

– **Discourse:** We use CODRA to segment the speech EDU's and identify the relations between them [15]. We count the number of occurrences of each of the 17 discourse relations, the depth of the discourse tree, the average number of EDU's per utterance, the ratio of each discourse relation to the total number of discourse relations, and the discourse relation type-to-token ratio.

5 Results

We use the *F-measure* as our evaluation metric, which is the weighted harmonic mean of precision and recall. The F-measures for all systems are shown in Figure 3. The main positive result is that domain adaptation does help with the task of detecting MCI. The best overall approach is the AUGMENT adaptation system without discourse features (F-Measure of 0.712, and 90% CI=0.633-0.791). The confidence intervals with this AUGMENT system are also tighter than the other approaches. Somewhat surprisingly, the source-only method (F-Measure of 0.681, and 90% CI=0.576-0.786) outperforms target-only (F-Measure of 0.640,

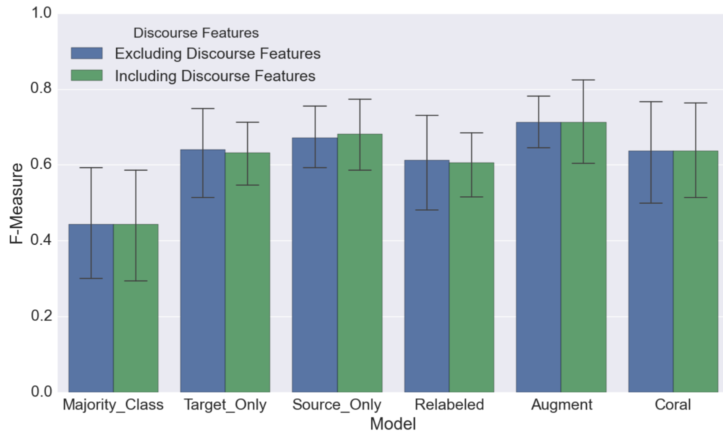

Fig. 3. Comparison of domain adaptation methods. We show the mean F-measure and 90% confidence intervals across a 10-fold CV. Only target data appears in the test fold.

and 90% CI=0.495-0.785), presumably because the source dataset is much larger. The AUGMENT system also selects a smaller percentage of the total features than the target-only, source-only and relabeled baselines. The effect of discourse features is a mixed result. The best performing baseline model does include the discourse features, but it is a very slight improvement. Furthermore, the AUGMENT, relabeled, and CORAL approaches all perform the same or slightly worse when discourse features are added. We suspect that because the speech elicited by the cookie theft test is both brief and highly specific, there will be few differences in the discourse structure between control and dementia groups. Discourse analysis may be more useful in longer and less structured narratives, where there is an opportunity for a speaker to use a larger set of discourse relations to connect one statement to the next. The main negative result is the performance of the CORAL domain adaptation method (F-Measure of 0.637, and 90% CI=0.487-0.786), which is nearly identical to the target-only method, i.e. equivalent to not doing domain adaptation at all. It has previously been found that CORAL does not always work well with boolean features such as bag-of-words features [27]. Info-units, which have been shown in previous work to be strong predictors of dementia, are largely boolean [12].

We also include a learning curve analysis showing AUGMENT's F-Measure score as a function of sample size, as seen in figure 4. We keep the ratio between target and source data constant and run the analysis using 25%, 50%, 75% and 100% of the data. We ran 15 trials as described[7] in section 4 on random subsets of the data and plot the average and 90% CI across all trials. Figure 4 shows a trend that increasing the size of the dataset improves the F-Measure and tightens the confidence intervals as we approach 100% but then the curve levels off. This

[7] With one small modification: We ran a 7-fold cross validation instead of 10-fold because there was not enough target data in the 25% trial to divide into 10 folds.

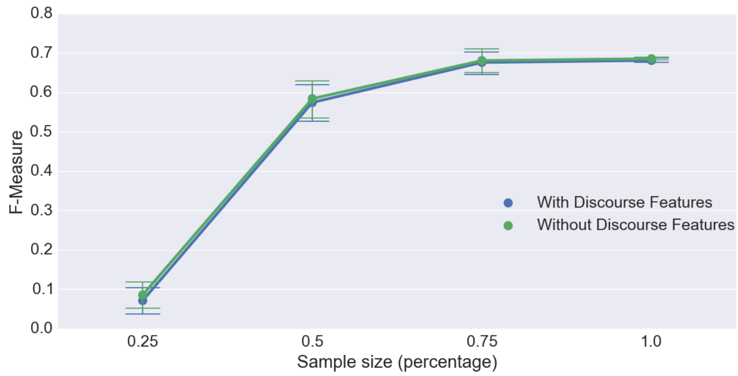

Fig. 4. Learning curve showing F-Measure with AUGMENT as a function of training data size. We keep the ratio of target and source data constant and average over 15 randomized trials.

suggests we may be nearing the limit of accuracy that can be achieved with a source-to-target ratio of approx 5:1, but we will investigate this more fully in future work.

6 Conclusion

Lack of data is a major obstacle facing researchers who wish to develop a tool to diagnose mild cognitive impairment from speech. In this work we evaluated two domain adaptation algorithms, AUGMENT and CORAL, which attempt to improve classification accuracy by using data collected from patients with Alzheimer's. Our main positive result is that the AUGUMENT domain adaptation algorithm outperformed all baseline algorithms and improved the F-measure by more than 7% over models trained on MCI data alone.

A second objective of this paper was to evaluate the efficacy of discourse features, which had not been used in previous work in this area. We speculated that features extracted from a discourse tree of patient transcripts might capture the loss of coherency which is characteristic of MCI, but unfortunately the discourse features failed to consistently improve the results.

In future work, we will modify CORAL to improve its performance in this setting. One possibility we will investigate is to align only the non-boolean features of the source domain rather than the entire feature space. We will also try merging both AUGMENT and CORAL into a single algorithm by adding a "CORAL aligned" copy of the feature to the AUGMENT feature space. A parallel path of future work involves expanding our system so it can accommodate data from multiple source domains simultaneously. In this way we will be able to use speech samples collected from patients with Vascular Dementia, Dementia with Lewy bodies, and other Non-Alzhiemers dementias. Finally, we wish to expand our system to leverage data collected from diagnostics test other than the Cookie-Theft test, such as the Narrative Retelling task from the Wechsler Logical Memory I/II test.

References

1. Petersen, R.C.: Mild cognitive impairment as a diagnostic entity. Journal of internal medicine **256**(3) (2004) 183–194
2. Petersen, R.C.: Mild cognitive impairment. CONTINUUM: Lifelong Learning in Neurology **22**(2, Dementia) (2016) 404–418
3. Hsiung, G.Y.R., Donald, A., Grand, J., Black, S.E., Bouchard, R.W., Gauthier, S.G., Loy-English, I., Hogan, D.B., Kertesz, A., Rockwood, K., et al.: Outcomes of cognitively impaired not demented at 2 years in the canadian cohort study of cognitive impairment and related dementias. Dementia and geriatric cognitive disorders **22**(5-6) (2006) 413–420
4. Faber-Langendoen, K., Morris, J.C., Knesevich, J.W., LaBarge, E., Miller, J.P., Berg, L.: Aphasia in senile dementia of the alzheimer type. Annals of neurology **23**(4) (1988) 365–370
5. McKeith, I.G., Galasko, D., Kosaka, K., Perry, E., Dickson, D.W., Hansen, L.a., Salmon, D., Lowe, J., Mirra, S., Byrne, E., et al.: Consensus guidelines for the clinical and pathologic diagnosis of dementia with lewy bodies (dlb) report of the consortium on dlb international workshop. Neurology **47**(5) (1996) 1113–1124
6. Weiner, M.F., Neubecker, K.E., Bret, M.E., Hynan, L.S.: Language in alzheimer's disease. The Journal of clinical psychiatry **69**(8) (2008) 1223
7. Boise, L., Neal, M.B., Kaye, J.: Dementia assessment in primary care: results from a study in three managed care systems. The Journals of Gerontology Series A: Biological Sciences and Medical Sciences **59**(6) (2004) M621–M626
8. Mitchell, A.J.: A meta-analysis then accuracy of the mini-mental state examination in the detection of dementia and mild cognitive impairment. Journal of psychiatric research **43**(4) (2009) 411–431
9. Ahmed, S., Haigh, A.M.F., de Jager, C.A., Garrard, P.: Connected speech as a marker of disease progression in autopsy-proven alzheimers disease. Brain **136**(12) (2013) 3727–3737
10. Rentoumi, V., Raoufian, L., Ahmed, S., de Jager, C.A., Garrard, P.: Features and machine learning classification of connected speech samples from patients with autopsy proven alzheimer's disease with and without additional vascular pathology. Journal of Alzheimer's Disease **42**(s3) (2014)
11. Fraser, K.C., Hirst, G., Graham, N.L., Meltzer, J.A., Black, S.E., Rochon, E.: Comparison of different feature sets for identification of variants in progressive aphasia. ACL 2014 (2014) 17
12. Fraser, K.C., Meltzer, J.A., Rudzicz, F.: Linguistic features identify alzheimers disease in narrative speech. Journal of Alzheimer's Disease **49**(2) (2015) 407–422
13. Becker, J.T., Boiler, F., Lopez, O.L., Saxton, J., McGonigle, K.L.: The natural history of alzheimer's disease: description of study cohort and accuracy of diagnosis. Archives of Neurology **51**(6) (1994) 585–594
14. Roark, B., Mitchell, M., Hosom, J.P., Hollingshead, K., Kaye, J.: Spoken language derived measures for detecting mild cognitive impairment. IEEE Transactions on Audio, Speech, and Language Processing **19**(7) (2011) 2081–2090
15. Joty, S., Carenini, G., Ng, R.T.: Codra: A novel discriminative framework for rhetorical analysis. Computational Linguistics (2015)

16. Chapman, S.B., Ulatowska, H.K., King, K., Johnson, J.K., McIntire, D.D.: Discourse in early alzheimer's disease versus normal advanced aging. American Journal of Speech-Language Pathology **4**(4) (1995) 124–129
17. Blonder, L.X., Kort, E.D., Schmitt, F.A.: Conversational discourse in patients with alzheimer's disease. Journal of Linguistic Anthropology **4**(1) (1994) 50–71
18. Ellis, D.G.: Coherence patterns in alzheimer's discourse. Communication Research **23**(4) (1996) 472–495
19. Dijkstra, K., Bourgeois, M.S., Allen, R.S., Burgio, L.D.: Conversational coherence: Discourse analysis of older adults with and without dementia. Journal of Neurolinguistics **17**(4) (2004) 263–283
20. Ellis, C., Henderson, A., Wright, H.H., Rogalski, Y.: Global coherence during discourse production in adults: a review of the literature. International Journal of Language & Communication Disorders (2016)
21. Laine, M., Laakso, M., Vuorinen, E., Rinne, J.: Coherence and informativeness of discourse in two dementia types. Journal of Neurolinguistics **11**(1) (1998) 79–87
22. Davis, B.H.: So, you had two sisters, right? functions for discourse markers in alzheimers talk. In: Alzheimer Talk, Text and Context. Springer (2005) 128–145
23. Feng, V.W.: RST-style discourse parsing and its applications in discourse analysis. PhD thesis, University of Toronto (2015)
24. Chelba, C., Acero, A.: Adaptation of maximum entropy capitalizer: Little data can help a lot. Computer Speech & Language **20**(4) (2006) 382–399
25. Blitzer, J., McDonald, R., Pereira, F.: Domain adaptation with structural correspondence learning. In: Proc. EMNLP 2006, Sydney, Australia. (July 2006) 120–128
26. Daume, H.: Frustratingly easy domain adaptation. In: Proc. ACL 2007. (2007)
27. Sun, B., Feng, J., Saenko, K.: Return of frustratingly easy domain adaptation. arXiv preprint arXiv:1511.05547 (2015)
28. Giles, E., Patterson, K., Hodges, J.R.: Performance on the boston cookie theft picture description task in patients with early dementia of the alzheimer's type: missing information. Aphasiology **10**(4) (1996) 395–408
29. Gao, S., Hendrie, H.C., Hall, K.S., Hui, S.: The relationships between age, sex, and the incidence of dementia and alzheimer disease: a meta-analysis. Archives of general psychiatry **55**(9) (1998) 809–815
30. Covington, M.A., McFall, J.D.: Cutting the gordian knot: The moving-average type–token ratio (mattr). Journal of Quantitative Linguistics **17**(2) (2010) 94–100
31. Salsbury, T., Crossley, S.A., McNamara, D.S.: Psycholinguistic word information in second language oral discourse. Second Language Research **27**(3) (2011) 343–360
32. Brysbaert, M., New, B.: Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for american english. Behavior Research Methods **41** (2009) 977–990
33. Croisile, B., Ska, B., Brabant, M.J., Duchene, A., Lepage, Y., Aimard, G., Trillet, M.: Comparative study of oral and written picture description in patients with alzheimer's disease. Brain and language **53**(1) (1996) 1–19