Detecting semantic changes in Alzheimer's disease with vector space models

Kathleen C. Fraser, Graeme Hirst
Department of Computer Science
University of Toronto, Toronto, Canada
kfraser@cs.toronto.edu, gh@cs.toronto.edu

Abstract

Numerous studies have shown that language impairments, particularly semantic deficits, are evident in the narrative speech of people with Alzheimer's disease from the earliest stages of the disease. Here, we present a novel technique for capturing those changes, by comparing distributed word representations constructed from healthy controls and Alzheimer's patients. We investigate examples of words with different representations in the two spaces, and link the semantic and contextual differences to findings from the Alzheimer's disease literature.

Keywords: distributional semantics, Alzheimer's disease, narrative speech

1. Introduction

Vector space models of semantics have become an increasingly popular area of research in computational linguistics, with notable successes on tasks such as query expansion for information retrieval (Manning et al., 2008), synonym identification (Bullinaria and Levy, 2012), sentiment analysis (Socher et al., 2012), machine translation (Zou et al., 2013), and many others. Here we present a preliminary study on how we can use vector space models to detect semantic changes that may occur with Alzheimer's disease (AD).

The general idea is simple: if we construct two semantic spaces from two different corpora, we expect the differences between the spaces to be related to the differences between the corpora. If we generate word vectors from a corpus of text about cars, and another set of word vectors from a corpus about wildlife, we expect that the word *jaguar* will have two very different representations in these two spaces. If the dimensions are the same, we can measure the distance between the two vectors for *jaguar*, and we would expect to find that it is non-zero.

In this study, we fix the topic of the two corpora to be the same: each document in the two corpora is a description of the "Cookie Theft" picture shown in Figure 1. Rather, the difference is that the documents in one corpus were produced by people with AD, and the documents in the other corpus were produced by healthy, older controls. We suggest that differences in the vector representations trained on the two corpora will be due, at least in part, to the semantic impairment that often occurs with AD.

In the following sections, we first present a brief summary of the literature on language in AD as well as related computational work. We then describe our data and procedure, examine the differences between the two corpora using a simple vector representation, and present three methods to help interpret these differences, with specific examples from the narratives. We also discuss the limitations of this study and suggest ways to build on these preliminary results.

2. Background

A great deal of work has been undertaken studying the degradation of semantic processing in AD, of which we will only begin to scratch the surface in this discussion.

Semantic memory deficits have been widely reported, with AD patients having difficulty on naming tasks and often substituting high-frequency hypernyms or semantic neighbours for target words which cannot be accessed (Kempler, 1995; Giffard et al., 2001; Kirshner, 2012; Reilly et al., 2011). Numerous studies have reported a greater impairment in category naming fluency (e.g., naming animals or tools) relative to letter naming fluency (e.g., naming words that start with the letter R) (Salmon et al., 1999; Monsch et al., 1992; Adlam et al., 2006). As a result of word-finding difficulties and a reduction in working vocabulary, the language of AD patients can seem "empty" (Ahmed et al., 2013) and lacking coherence (Appell et al., 1982). In the famous "Nun Study" (Snowdon et al., 1996), it was shown that decreased idea density in writing produced in early life was associated with developing Alzheimer's disease decades later.

Specifically with regards to the "Cookie Theft" picture description task that we consider here, AD patients tend to show a reduction in the amount of information that is conveyed (Giles et al., 1996; Croisile et al., 1996; Lira et al., 2014). That is, they do not mention all the expected facts or inferences about the picture. Furthermore, these impairments are noticeable from a very early stage in the disease (Forbes-McKay and Venneri, 2005). Nicholas et al. (1985) found that AD patients mentioned roughly half of the expected information units, and produced a large number of deictic terms and indefinite terms (e.g. pronouns without antecedents). Ahmed et al. (2013) found that AD patients made fewer references to the people and their actions depicted in the picture than controls.

Recently, there has been some progress on automatically determining the information content of picture description narratives using computational techniques. Pakhomov et al. (2010) generated a list of expected information units and some of their lexical and morphological variants, then searched for matches. Hakkani-Tür et al. (2010) scored picture descriptions using information retrieval techniques

to match the narratives with a list of 35 key concepts.
In previous work, we used a combination of keyword-spotting and dependency parsing to identify relevant information units in "Cookie Theft" narratives (Fraser et al., 2015). However, accurately identifying atypical speech patterns will require accounting for not just *what* words are used, but *how* they are used. A better understanding of the semantic space and the different senses in which words are used will be a first step towards better models for detecting AD from speech.

3. Data

The narrative speech data were obtained from the Pitt corpus in the DementiaBank database[1] (MacWhinney, 2007). These data were collected between 1983 and 1988 as part of the Alzheimer Research Program at the University of Pittsburgh. Detailed information about the study cohort is available from Becker et al. (1994), and demographic information is given in Table 1. Unfortunately, the patient and control groups are not matched for age and education; the AD patients tend to be both older ($p < 0.01$) and less educated ($p < 0.01$), which is one limitation of this data set. There is no significant difference on sex ($p = 0.8$).
The language samples were elicited using the "Cookie Theft" picture description task from the Boston Diagnostic Aphasia Examination (BDAE) (Goodglass and Kaplan, 1983), in which participants are asked to describe everything they see going on in a picture. The stimulus picture is shown in Figure 1. The data were manually transcribed following the CHAT transcription protocol (MacWhinney, 2000).
Patients in the Pittsburgh study were diagnosed on the basis of their clinical history and their performance on neuropsychological testing, and the diagnoses were updated in 1992, taking into account any relevant information from the intervening years. Autopsies were performed on 50 patients, and in 43 cases the AD diagnosis was confirmed (86.0%) (Becker et al., 1994). A more recent study of clinical diagnostic accuracy in AD found that of 526 cases diagnosed as probable AD, 438 were confirmed as neuropathological AD post-mortem (83.3%) (Beach et al., 2012), suggesting that the DB diagnoses are generally as reliable as diagnoses made using present-day criteria.
We include 240 narratives from 167 participants diagnosed with possible or probable AD (average number of narratives per participant is 1.44, median is 1.0), and 233 narratives from 97 healthy, elderly controls (average 2.40, median 2.0). As shorthand throughout the paper, we refer to the set of narratives from participants with AD as the "AD corpus", and the set of narratives from healthy controls as the "CT corpus." In total, the AD corpus contains 31,906 words, and the CT corpus contains 27,620 words.

4. Differences in word representations between AD patients and controls

To compare the vector spaces directly, we require that the dimensions be identical (in interpretation as well as number). For this reason, we do not consider popular neural network models such as skip-gram and CBOW (Mikolov et al., 2013), whose resulting dimensions are not easily interpretable. Instead we consider a simple word-word co-occurrence model, in which the rows and columns represent words from the vocabulary, and the value of (r_i, c_i) is the number of times context word c_i appears near the given word r_i. We use a window size of three words on each side of the target word, with the exception of words at the beginning and end of narrative samples (i.e. the window is not permitted to overlap with the end of one sample and the beginning of the next). To reduce data sparsity, we consider only words that occur a minimum of 10 times in both the CT and AD corpora, and we lemmatize the words using NLTK's WordNet lemmatizer, after first tagging the words to increase lemmatization accuracy (Bird et al., 2009). After examining the frequency distribution of words in the two corpora, we decided not to remove any stop-words, as many of the highest frequency words are actually content words (e.g. *cookie*). Furthermore, we predict that common words such as prepositions and pronouns might show some variation in usage between the groups.
We stated above that the most likely reason for differences between the vector representations would be differences between the language of people with AD and healthy controls. Of course, another reason for differences could simply be random variation in word choice and speaking style between individuals, which may be a factor here given the relatively small size of the data set. To mitigate this effect, we adopt the following procedure:

1. First, split the CT corpus in half, create two co-occurrence matrices, and measure the cosine distance

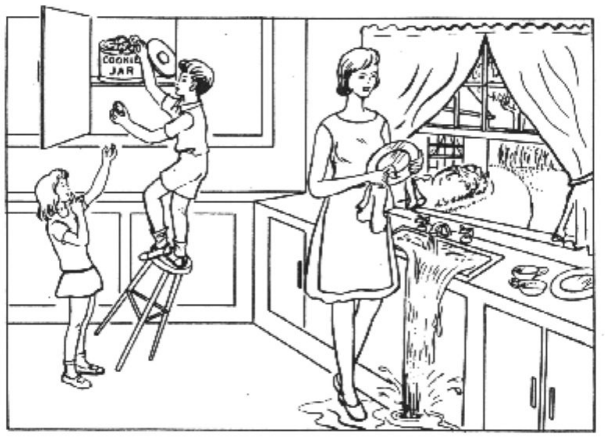


Figure 1: The Cookie Theft Picture (Goodglass and Kaplan, 1983).

	AD	Controls
	$n = 240$	$n = 233$
Age	71.8 (8.5)	65.2 (7.8)
Education	12.5 (2.9)	14.1 (2.4)
Sex (M/F)	82/158	82/151
MMSE	18.5 (5.1)	29.1 (1.1)

Table 1: Demographic information (mean and standard deviation).

[1] https://talkbank.org/DementiaBank/

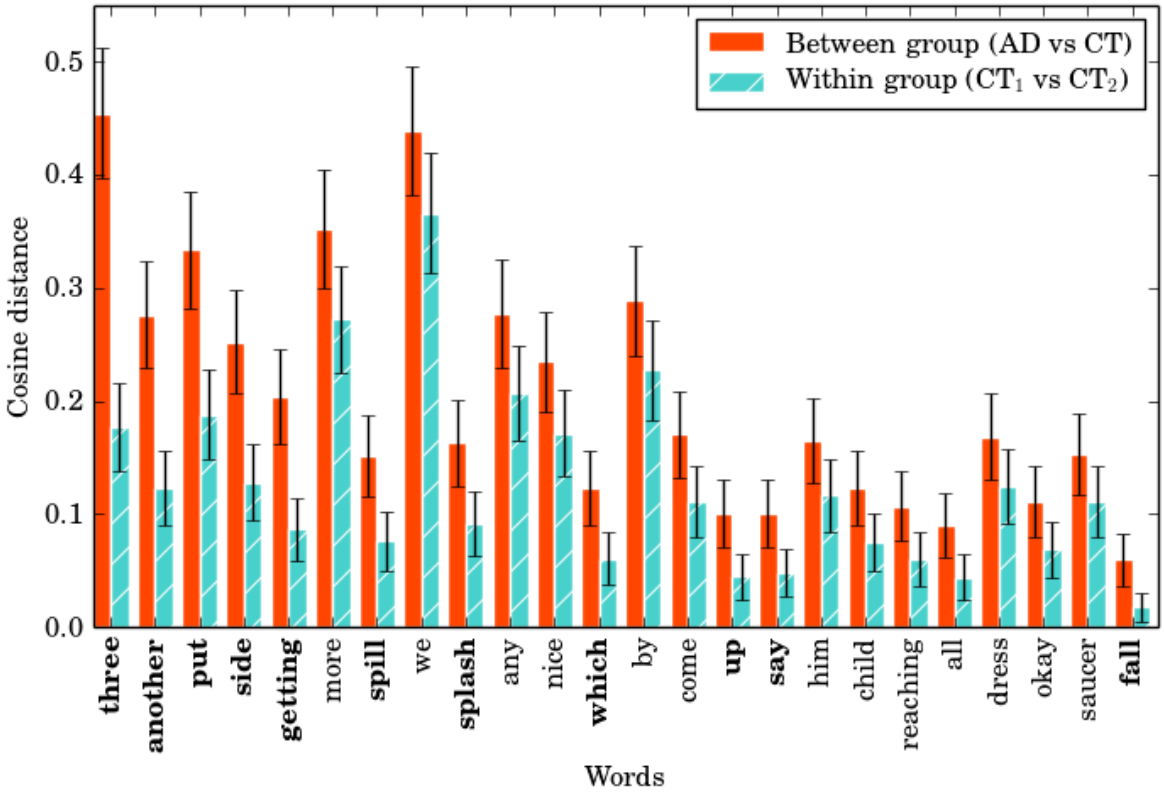


Figure 2: Cosine distances within the control group, and between the control and AD groups. Words marked in bold on the horizontal axis were selected for analysis in the next section (the difference between groups was greater than the error).

between the vector for word *w* in the first and the vector for word *w* in the second. This gives us an idea of the expected variation that occurs for each word.

2. Then, measure the cosine distance between the representation of word *w* trained on the CT corpus and trained on the AD corpus. This represents the variation of the word across the two groups.

3. Finally, only select a word for analysis if the distance *across* groups is greater than the distance *within* the control group; that is, if the variation between AD patients and controls is greater than the normal variation within healthy speakers. (Note that we do not consider the variation within the AD group in this calculation, as we want to measure whether the across-group variation is greater than *typical* variation.)

One difficulty that we encountered in performing this calculation was choosing an appropriate metric for measuring statistical significance in cosine differences. We experimented with partitioning the data into several folds, obtaining observations from each fold, and then testing for significance (as in Bullinaria and Levy (2012)), but concluded that our data set is simply too small for this method to be feasible. For lack of a better option, and given the close relationship between cosine similarity and correlation (Van Dongen and Enright, 2012), we instead computed the standard error for each cosine distance (treating each word as an observation). Figure 2 shows the cosine distances and errors for a subset of vectors, including all of those which were selected for analysis in the following section.

5. Interpretation of the differences

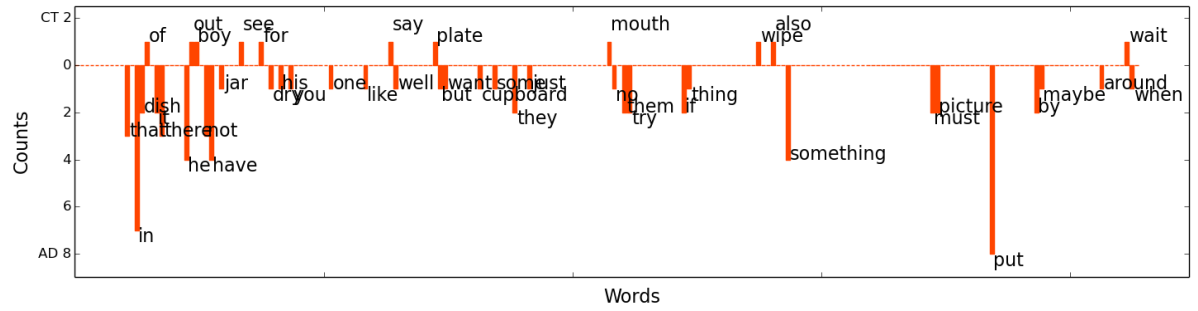
The method described above leaves us with a fairly small set of vectors that differ notably between the groups (i.e. only those 11 features shown in bold in Figure 2). The next question is: in what way are they different, and how does this relate to our knowledge of Alzheimer's disease? This proved to be a more difficult question to answer. In the following sections we present three different approaches, with illustrative examples for each.

5.1. Contextual differences

As a first step, we examined those dimensions which were non-zero in one group and zero in the other (i.e. context for a given word that appeared in one group but not the other). In the selected words, two different scenarios were observed: In the first case, the control participants used a number of context words not used by the AD participants. An example of this is shown in Figure 3a, for the word *another*. Two context words which occur fairly often with *another* in the control group are *he* and *window*. Some examples of these words in context (words inside the context window are italicized) are:

- And *he's getting another one out of* the cookie jar
- *He's handing another one to the* little girl
- *And there's another window and some* trees apparently

(a) another



(b) put

Figure 3: Differences in context for the words *another* and *put*. Counts above the horizontal axis indicate context words that occurred only in the control group; counts below the horizontal axis indicate words that occurred only the AD group.

- You can *see beyond that another window and um*

These examples demonstrate a certain attention to detail — to say that the boy is getting *another cookie*, one must first observe that he already has a cookie in his hand; the second window referenced in *another window* is a minor detail seen through the first, more prominent window. They also reflect an element of cohesion, in that *another window* makes reference to an earlier window mentioned by the speaker, and clarifies to the listener that this new reference to a window is distinct from the previous one. Prior work has shown that attention is often one of the first areas of cognition (after memory) to be affected in AD (Perry and Hodges, 1999), and that the narratives of people with AD tend to show a lack of cohesion (Chenery and Murdoch, 1994).

In the second case, the AD participants use a number of context words that do not occur in the CT corpus. An example of this is illustrated in Figure 3b, for the word *put*. One of the most frequent context words in the figure is *put* itself. Control participants rarely repeat the same word within the 6-word context window, but it is not uncommon in the AD group. Another interesting context word is *in*. The controls do not tend to describe any of the actions in the Cookie Theft picture as *putting* something *in* something else. They use *put* to describe the action of the girl (e.g. the little girl is putting her finger to her mouth). On the other hand, some examples from the AD corpus include:

- he's *trying to put put put food in* that in that crocker jar
- has a cookie jar up *there he's putting cookies in and* the thing's falling over

These errors are similar to the "implausible details" that Croisile et al. (1996) found to occur more frequently in AD narratives than controls. The underlying explanation is unclear, although it could represent a breakdown in logic and understanding. It also demonstrates a potential pitfall of the keyword-spotting approach to scoring — a participant may mention the boy and the cookie jar, but the action connecting the two is also fundamentally important.

5.2. Vectors shifting in space

Another way of looking at these differences is to see how the words in question have moved in the vector space. To visualize the space in two dimensions, we use the method of t-distributed stochastic neighbor embedding (t-SNE) (Van der Maaten and Hinton, 2008). The t-SNE method was proposed as a solution to the problem of visualizing high-dimensional data in two or three dimensions. It is capable of producing visualizations that reveal structure at both the local and global level, although the resulting dimensions are not generally interpretable (and therefore not labelled in the figures). The example word we consider here is *getting*. (Note that verbs ending in *-ing* are subject to a consistent issue in the tagging and lemmatizing pipeline which results in them not being reduced to the base form.) Figure 4 shows part of the two-dimensional representation of the word vector space. In many cases, the word representations in the AD and CT corpora lie very close to each other. However, in the case of *getting*, the vectors lie much further apart. Examining the surrounding vectors, it appears that *getting* is closer to *running*, *overflowing*, and

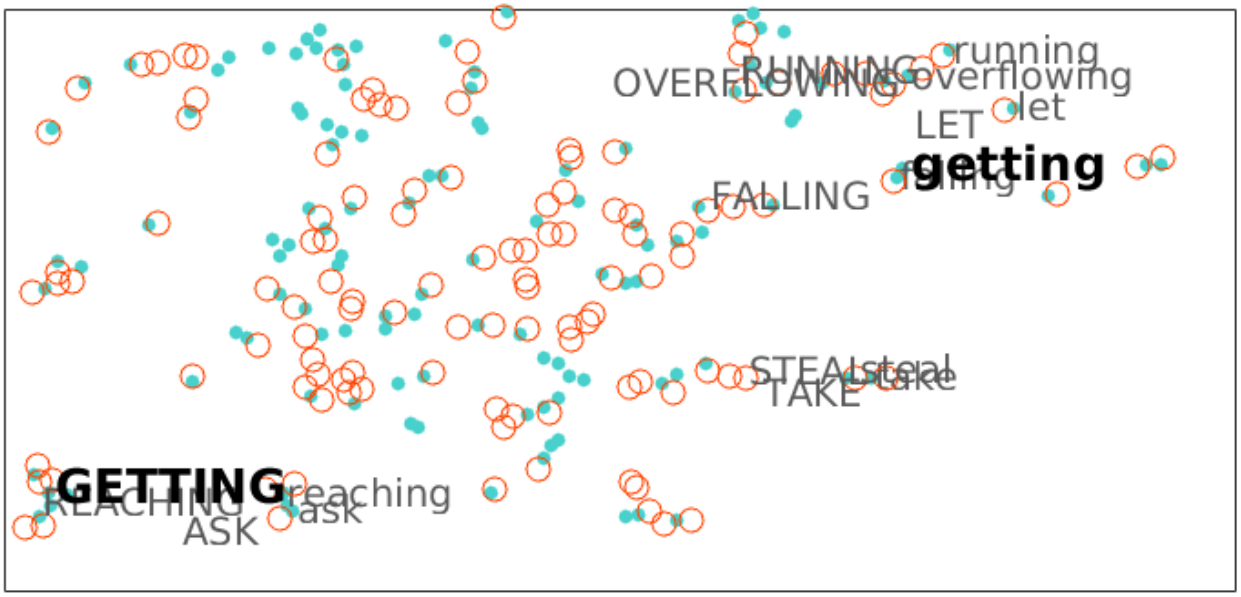


Figure 4: Two-dimensional visualization of the vector space using t-SNE. Word representations from the AD corpus are labelled with filled, green circles and lowercase labels; words from the CT corpus are labelled with open, orange circles and uppercase labels.

falling in the AD corpus, and closer to words like *reaching* and *ask* in the CT corpus. This is confirmed by comparing the cosine distances (Table 2).

The nearest neighbours of the vectors suggest that in the AD corpus, *getting* is used more in the context of the sink overflowing, while in the CT corpus *getting* is used in the context of the cookie theft. This is borne out in the data itself, as there is only one example in the control group of using *getting* in the context of the sink (*her foot is getting wet*) and the rest refer to the act of stealing the cookie. In the AD group there are a number of references to the sink context (e.g. *the floor is getting wet*, *mom is getting her foot wet*, *the water is getting over the sink*), as well as referring to stealing the cookie.

One explanation for this phenomenon could lie in the fact that *get* is a "light" verb, in that it does not convey very much semantic information about the action it describes (Breedin et al., 1998). Kim and Thompson (2004) showed that people with Alzheimer's disease produced more light verbs and fewer heavy verbs in a story-telling task, and were more impaired on retrieving heavy verbs than light verbs in a sentence completion task. Both AD and control speakers use *getting* in the sense of *getting a cookie*, which is the primary sense of the word *get*[2], meaning "to obtain or procure". However, we expect that AD speakers may also substitute light, easy-to-access verbs for more semantically appropriate verbs. This phenomenon is observed when, for example, an AD participant says *the water is getting over the sink* rather than *the water is flowing/overflowing/running over the sink*.

5.3. Cluster analysis

The example in the previous section illustrates how a single word can have multiple senses, and these senses can be distinguished by the different contexts in which they appear. This idea is the basis of the Distributional Hypothesis, stated perhaps most succinctly as, "you shall know a word by the company it keeps" (Firth, 1957). The difficulty of representing the different senses of a word was an issue in the early days of vector space representations, although numerous solutions have been proposed since (Reisinger and Mooney, 2010; Huang et al., 2012; Guo et al., 2014; Wu and Giles, 2015).

Words	Distance (AD corpus)	Distance (CT corpus)
getting, running	**0.225**	0.524
getting, overflowing	**0.224**	0.511
getting, falling	**0.191**	0.384
getting, reaching	0.395	**0.349**
getting, ask	0.588	**0.482**

Table 2: Cosine distances for the words in Figure 4 in each of the two vector representations (trained on AD corpus and CT corpus).

In this section, we perform an analysis based on methods for unsupervised word sense discovery. We take a step back, and rather than considering the final vector representation for a word, we look at all the context vectors that contribute to the final vector and perform cluster analysis on them. Different clusters will represent different contexts, and by assumption different word senses. We use k-means clustering with a Euclidean distance metric. The optimal number of clusters k is chosen manually, by the silhouette method (Rousseeuw, 1987).

Our example for this section is the word *three*. The clusters for $k = 5$ are shown in two-dimensions in Figure 5a. The lemmatized context words associated with each point are given in Figure 5b. Our interpretation of the clusters

[2] http://www.oed.com/view/Entry/77994

is subjective, but in general we see that both AD participants and controls use the word *three* to describe the three-legged stool (turquoise cluster), as well as the number of dishes and the number of people in the room (purple cluster). Of the smaller outlier clusters, two consist of only contexts from the AD corpus. The orange cluster is made up of examples where participants described three cups, which is not a semantically accurate representation of the picture (there are three dishes, but one is a plate). In the green cluster, a single AD participant repeats the word *three* (thus creating two instances with very similar contexts) and uses the context word *woman*, which is unusual (the unlemmatized transcript reads *one two three three women*). Interestingly, although they were placed in a separate cluster, we also see this "counting" use of *three* in context vectors 4 and 7 (both from the AD corpus). Finally, the red cluster consists of two cases of repeating the word *three* in the context of the stool — one from the CT corpus and one from the AD corpus.

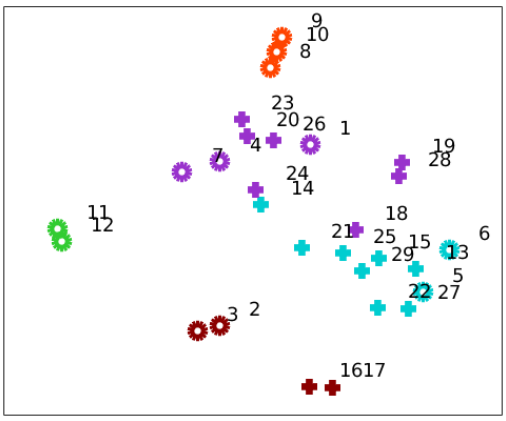
To summarize, in this example we see word senses (contexts) that are used by both AD participants and controls, and then we see other rare senses that appear only in the AD corpus. In this particular case, those rare senses correspond to semantic errors, although more work will be needed to see if that result generalizes to other words.

6. Limitations

Models based on raw word co-occurrence counts are perhaps the most basic distributional models of semantics, and it is known that performance is usually improved by (a) transforming the raw counts using methods like positive pointwise mutual information, or (b) learning predictive models using neural networks (Baroni et al., 2014). However, as the model increases in complexity, we face the issue of whether the Alzheimer's model and control model can still be compared directly. The literature on comparing different semantic spaces is relatively sparse, with some exceptions (Zuccon et al., 2009). In future work we plan to build on the baseline we have presented here by exploring different vector space models and methods for comparing them.

Furthermore, certain aspects of our methodology, such as choosing the optimal number of clusters, involve human intervention and have some degree of subjectivity. Work on automatically choosing k and then evaluating the purity of clusters and identifying outlier clusters is currently under way.

Moving away from computational details and looking at the big picture, it is clear that this study faces the same problem as many others: trying to study individual variation at the population level. We do not expect that every person with AD will say the boy is putting cookies into the jar, or that there are three cups on the counter. Rather, we expect that most people with AD will start to make semantic errors of some kind. In doing this analysis, we have picked up on semantic errors that we did not find using our previous approach (Fraser et al., 2015), and which have not been reported, to our knowledge, in any previous work using this data set. However, our approach here was similar in some ways to a case study, where we dug deep into a few representative examples. The true value will lie in scaling our methods to detect and count general semantic irregularities, which can then be used as input to a system for screening, longitudinal assessment, or diagnostic support.



(a) Two-dimensional visualization of the clusters using t-SNE. Circles represent contexts from the AD corpus, while plus signs represent contexts from the CT corpus.

Purple	1	okay there be person in the
	4	uh two dish dish sit on
	7	have two two four five door
	19	oh I see people in there
	20	dish there be dish set on
	23	flower there be dish leave to
	24	of the on of the cupboard
	26	back there be dish on the
	28	can see those thing
Red	2	xxx doing xxx three legged stool
	3	doing xxx three legged stool I
	16	fall off a three prong stool
	17	off a three prong stool his
Turquoise	5	stand on a legged stool and
	6	depart from that legged stool he
	13	a stool a legged stool and
	14	upend on the legged stool uh
	15	be a a legged stool and
	18	it be a it be a
	21	his sister the legged stool be
	22	be on a legged stool which
	25	stool be a legged stool um
	27	up on a legged stool which
	29	it be a legged stool and
Orange	8	running there be cup and uh
	9	the floor and cup three bowl
	10	and three cup bowl there
Green	11	be one two three woman has
	12	one two three woman has some

(b) The lemmatized contexts corresponding to each point.

Figure 5: Clustering of contexts for the word *three*.

7. Conclusion

We have presented preliminary results showing that the changes in word usage that occur in Alzheimer's disease

can be detected through analysis of the resulting semantic space. We examined these differences through visual analysis of the vectors themselves, two-dimensional representations of the vector spaces, and cluster analysis of the individual context vectors. Many of the differences are consistent with previous work on language changes in AD. Future work will focus on how these methods can be applied to automated scoring of the picture description task, or generating meaningful features for a diagnostic classifier.

8. Acknowledgements

The authors would like to thank Tong Wang and Maria Yancheva for their helpful discussions. This work was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) grant number RGPIN-2014-06020.

9. References

Adlam, A.-L. R., Bozeat, S., Arnold, R., Watson, P., and Hodges, J. R. (2006). Semantic knowledge in mild cognitive impairment and mild Alzheimer's disease. *Cortex*, 42(5):675–684.

Ahmed, S., de Jager, C. A., Haigh, A.-M., and Garrard, P. (2013). Semantic processing in connected speech at a uniformly early stage of autopsy-confirmed Alzheimer's disease. *Neuropsychology*, 27(1):79.

Appell, J., Kertesz, A., and Fisman, M. (1982). A study of language functioning in Alzheimer patients. *Brain and language*, 17(1):73–91.

Baroni, M., Dinu, G., and Kruszewski, G. (2014). Don't count, predict! A systematic comparison of context-counting vs. context-predicting semantic vectors. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 238–247.

Beach, T. G., Monsell, S. E., Phillips, L. E., and Kukull, W. (2012). Accuracy of the clinical diagnosis of Alzheimer disease at National Institute on Aging Alzheimer Disease Centers, 2005–2010. *Journal of Neuropathology & Experimental Neurology*, 71(4):266–273.

Becker, J. T., Boiler, F., Lopez, O. L., Saxton, J., and McGonigle, K. L. (1994). The natural history of Alzheimer's disease: description of study cohort and accuracy of diagnosis. *Archives of Neurology*, 51(6):585–594.

Bird, S., Klein, E., and Loper, E. (2009). *Natural Language Processing with Python*. O'Reilly Media.

Breedin, S. D., Saffran, E. M., and Schwartz, M. F. (1998). Semantic factors in verb retrieval: An effect of complexity. *Brain and Language*, 63:1–31.

Bullinaria, J. A. and Levy, J. P. (2012). Extracting semantic representations from word co-occurrence statistics: stop-lists, stemming, and SVD. *Behavior Research Methods*, 44(3):890–907.

Chenery, H. J. and Murdoch, B. E. (1994). The production of narrative discourse in response to animations in persons with dementia of the Alzheimer's type: Preliminary findings. *Aphasiology*, 8(2):159–171.

Croisile, B., Ska, B., Brabant, M.-J., Duchene, A., Lepage, Y., Aimard, G., and Trillet, M. (1996). Comparative study of oral and written picture description in patients with Alzheimer's disease. *Brain and language*, 53(1):1–19.

Firth, J. R. (1957). A synopsis of linguistic theory. *Studies in Linguistic Analysis*, pages 1–32.

Forbes-McKay, K. E. and Venneri, A. (2005). Detecting subtle spontaneous language decline in early Alzheimer's disease with a picture description task. *Neurological Sciences*, 26:243–254.

Fraser, K. C., Meltzer, J. A., and Rudzicz, F. (2015). Linguistic features identify Alzheimer's disease in narrative speech. *Journal of Alzheimer's Disease*, 49(2):407–422.

Giffard, B., Desgranges, B., Nore-Mary, F., Lalevée, C., de la Sayette, V., Pasquier, F., and Eustache, F. (2001). The nature of semantic memory deficits in Alzheimer's disease. *Brain*, 124(8):1522–1532.

Giles, E., Patterson, K., and Hodges, J. R. (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.

Goodglass, H. and Kaplan, E. (1983). *Boston diagnostic aphasia examination booklet*. Lea & Febiger Philadelphia, PA.

Guo, J., Che, W., Wang, H., and Liu, T. (2014). Learning sense-specific word embeddings by exploiting bilingual resources. In *Proceedings of the 25th International Conference on Computational Linguistics (COLING)*, pages 497–507.

Hakkani-Tür, D., Vergyri, D., and Tür, G. (2010). Speech-based automated cognitive status assessment. In *Proceedings of the 11th Annual Conference of the International Speech Communication Association (INTERSPEECH)*, pages 258–261.

Huang, E. H., Socher, R., Manning, C. D., and Ng, A. Y. (2012). Improving word representations via global context and multiple word prototypes. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics*, pages 873–882. Association for Computational Linguistics.

Kempler, D. (1995). Language changes in dementia of the Alzheimer type. *Dementia and Communication*, pages 98–114.

Kim, M. and Thompson, C. K. (2004). Verb deficits in Alzheimer's disease and agrammatism: Implications for lexical organization. *Brain and Language*, 88(1):1–20.

Kirshner, H. S. (2012). Primary progressive aphasia and Alzheimer's disease: brief history, recent evidence. *Current neurology and neuroscience reports*, 12(6):709–714.

Lira, J. O. d., Minett, T. S. C., Bertolucci, P. H. F., and Ortiz, K. Z. (2014). Analysis of word number and content in discourse of patients with mild to moderate Alzheimer's disease. *Dementia & Neuropsychologia*, 8(3):260–265.

MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk: Volume I: Transcription format and programs*. Lawrence Erlbaum Associates.

MacWhinney, B. (2007). The Talkbank Project. In Beal, J., Corrigan, K., and Moisl, H. L., editors, *Creating and Digitizing Language Corpora*, pages 163–180. Springer.

Manning, C. D., Raghavan, P., Schütze, H., et al. (2008).

Introduction to Information Retrieval, volume 1. Cambridge University Press.

Mikolov, T., Chen, K., Corrado, G., and Dean, J. (2013). Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.

Monsch, A. U., Bondi, M. W., Butters, N., Salmon, D. P., Katzman, R., and Thal, L. J. (1992). Comparisons of verbal fluency tasks in the detection of dementia of the Alzheimer type. *Archives of Neurology*, 49(12):1253–1258.

Nicholas, M., Obler, L. K., Albert, M. L., and Helm-Estabrooks, N. (1985). Empty speech in Alzheimer's disease and fluent aphasia. *Journal of Speech, Language, and Hearing Research*, 28(3):405–410.

Pakhomov, S. V., Smith, G. E., Chacon, D., Feliciano, Y., Graff-Radford, N., Caselli, R., and Knopman, D. S. (2010). Computerized analysis of speech and language to identify psycholinguistic correlates of frontotemporal lobar degeneration. *Cognitive and Behavioral Neurology*, 23:165–177.

Perry, R. J. and Hodges, J. R. (1999). Attention and executive deficits in Alzheimer's disease. *Brain*, 122(3):383–404.

Reilly, J., Troche, J., and Grossman, M. (2011). Language processing in dementia. *The handbook of Alzheimer's disease and other dementias*, pages 336–368.

Reisinger, J. and Mooney, R. J. (2010). Multi-prototype vector-space models of word meaning. In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL)*, pages 109–117. Association for Computational Linguistics.

Rousseeuw, P. J. (1987). Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20:53–65.

Salmon, D. P., Butters, N., and Chan, A. S. (1999). The deterioration of semantic memory in Alzheimer's disease. *Canadian Journal of Experimental Psychology*, 53(1):108.

Snowdon, D. A., Kemper, S. J., Mortimer, J. A., Greiner, L. H., Wekstein, D. R., and Markesbery, W. R. (1996). Linguistic ability in early life and cognitive function and Alzheimer's disease in late life: findings from the Nun Study. *Journal of the American Medical Association*, 275(7):528–532.

Socher, R., Huval, B., Manning, C. D., and Ng, A. Y. (2012). Semantic compositionality through recursive matrix-vector spaces. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pages 1201–1211. Association for Computational Linguistics.

Van der Maaten, L. and Hinton, G. (2008). Visualizing data using t-SNE. *Journal of Machine Learning Research*, 9:2579–2605.

Van Dongen, S. and Enright, A. J. (2012). Metric distances derived from cosine similarity and Pearson and Spearman correlations. *arXiv preprint arXiv:1208.3145*.

Wu, Z. and Giles, C. L. (2015). Sense-aaware semantic analysis: A multi-prototype word representation model using wikipedia. In *Proceedings of the 29th AAAI Conference on Artificial Intelligence (AAAI-15)*, pages 2188–2194.

Zou, W. Y., Socher, R., Cer, D. M., and Manning, C. D. (2013). Bilingual word embeddings for phrase-based machine translation. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1393–1398.

Zuccon, G., Azzopardi, L. A., and Van Rijsbergen, C. (2009). Semantic spaces: Measuring the distance between different subspaces. In *The Proceedings of the 3rd International Symposium on Quantum Interaction*, pages 225–236.