# Meta-Semantic Representation for Early Detection of Alzheimer's Disease

**Jinho D. Choi**
Computer Science
Emory University
Atlanta, GA, USA
jchoi31@emory.edu

**Mengmei Li**
Computer Science
Emory University
Atlanta, GA, USA
kate.li@emory.edu

**Felicia Goldstein**
Neurology
Emory University
Atlanta, GA, USA
fgoldst@emory.edu

**Ihab Hajjar**
Neurology
Emory University
Atlanta, GA, USA
ihajjar@emory.edu

## Abstract

This paper presents a new task-oriented meaning representation called *meta-semantics*, that is designed to detect patients with early symptoms of Alzheimer's disease by analyzing their language beyond a syntactic or semantic level. Meta-semantic representation consists of three parts, entities, predicate argument structures, and discourse attributes, that derive rich knowledge graphs. For this study, 50 controls and 50 patients with mild cognitive impairment (MCI) are selected, and meta-semantic representation is annotated on their speeches transcribed in text. Inter-annotator agreement scores of 88%, 82%, and 89% are achieved for the three types of annotation, respectively. Five analyses are made using this annotation, depicting clear distinctions between the control and MCI groups. Finally, a neural model is trained on features extracted from those analyses to classify MCI patients from normal controls, showing a high accuracy of 82% that is very promising.

## 1 Introduction

Our understanding of Alzheimers disease (AD) has evolved over the last few decades. Most notably is the discovery that AD has long latent preclinical and mild cognitive impairment (MCI) stages (Karr et al., 2018; Steenland et al., 2018). These stages are the focus of many prevention and therapeutic interventions. A key limitation in identifying these pre-dementia stages for clinical trial recruitment is the need for expensive or invasive testing like positron emission tomography or obtaining cerebrospinal fluid (CSF) analyses. Traditional cognitive testing is time-consuming and can be biased by literacy and test-taking skills (Fyffe et al., 2011). Recent advances in natural language processing (NLP) offer the unique opportunity to explore previously undetectable changes in the cognitive process of semantics that can be automated in clinical artificial intelligence (Beam and Kohane, 2016).

Limited prior studies have suggested the feasibility of detecting AD by analyzing language variations. One approach includes linguistically motivated analysis extracting lexical, grammatical, and syntactic features to detect language deficits in AD patients (Fraser et al., 2016; Orimaye et al., 2017). The other approach involves deep learning models to extract features from languages used by AD patients (Orimaye et al., 2016; Karlekar et al., 2018). The limitations of these studies are that most were developed based on dementia cases, so their ability to detect pre-dementia is still unknown. The impact of these methods is the highest in the cases where traditional cognitive measures are unable to clarify the patients cognitive status. Hence, we focus on these early MCI stages in this study.

We suggest a new meaning representation called *meta-semantics* that derives a knowledge graph reflecting semantic, pragmatic, and discourse aspects of language spoken by MCI patients. The objective of this representation is not to design yet another structure to capture more information but to sense aspects beyond the syntax and semantic level that are essential for the early detection of MCI patients. We hypothesize that patients in the pre-dementia stage do not necessarily make so much of grammatical mistakes compared to normal people but often have difficulties in elaborating or articulating their thoughts in language. To verify our hypothesis, we collect speeches from 50 normal controls and 50 MCI patients that standardized cognition tests fail to distinguish (Section 2), annotate meta-semantic representation on the transcripts of those speeches (Section 3), make several analyses to comprehend linguistic differences between the control and the MCI groups (Section 4), then develop a neural network model to detect MCI patients from normal controls (Section 5). To the best of our knowledge, this is the first time that a dedicated meaning representation is proposed for the detection of MCI.



*Proceedings of the First International Workshop on Designing Meaning Representations*, pages 82–91
Florence, Italy, August 1st, 2019 

## 2 Data Preparation

We analyzed data from 100 subjects collected as part of the B-SHARP, Brain, Stress, Hypertension, and Aging Research Program.[1] 50 cognitively normal controls and 50 patients with mild cognitive impairment (MCI) were selected based on neuropsychological and clinical assessments performed by a trained physician and a neuropsychologist. The two groups were matched on overall cognitive scores to examine how well our new meta-semantic indices would perform in the setting where standardized tests such as the Montreal Cognitive Assessment (Nasreddine et al., 2005) and the Boston Naming Test (Kaplan et al., 1983) failed to distinguish them. Table 1 shows demographics and clinical features of the control and the MCI groups.

| Type | Control | MCI | P-value |
|---|---|---|---|
| Age | 65.6 (±6.80) | 66.0 (±8.38) | 0.809 |
| Race | 54%; 44% | 58%; 42% | 0.840 |
| Sex | 62% | 60% | 1.000 |
| Education | 54% | 56% | 1.000 |
| MoCA | 24.2 (±2.15) | 23.9 (±2.00) | 0.502 |
| BNT | 14.0 (±1.43) | 13.8 (±1.23) | 0.550 |
| CDR | 0.01 (±0.07) | 0.43 (±0.18) | **<0.001** |
| FAQ | 1.00 (±1.62) | 1.71 (±2.57) | 0.103 |

Table 1: Demographics and clinical features of the two groups. Age: avg-years, Race: % African American; % Non-Hispanic Caucasian, Sex: % female, Education: % Bachelor's or above, MoCA (Montreal Cognitive Assessment): avg-score, BNT (Boston Naming Test): avg-score, CDR (Clinical Dementia Rating): avg-score, FAQ (Function Assessment Questionnaire): avg-score. The p-values are evaluated by the t-test except for race, sex, and education which are evaluated by the $\chi^2$ test.

No significant group differences were found in age, race, sex, or education between these two groups. The MCI group performed significantly worse on the Clinical Dementia Rating (Morris, 1994), but did not differ as much on the Function Assessment Questionnaire (Pfeffer et al., 1982) assessing instrumental activities of daily living.

### 2.1 Speech Task Protocol

We conducted a speech task protocol that evaluated subjects' language abilities on 1) natural speech, 2) fluency, and 3) picture description, and collected audio recordings for all three tasks from each subject. For this study, the audio recordings from the third task, picture description, were used to demonstrate the effectiveness of the meta-semantics analysis on detecting MCI. All subjects were shown the picture in Figure 1, *The Circus Procession*, copyrighted by McLoughlin Brothers as part of the Juvenile Collection, and given the same instruction to describe the picture for one minute. Visual abilities of the subjects were confirmed before recording.


Figure 1: The image of "*The Circus Procession*" used for the picture description task.

### 2.2 Transcription

Audio recordings for the picture description task (Section 2.1) from the 100 subjects in Table 1 were automatically transcribed by the online tool, Temi,[2] then manually corrected. Table 3 shows transcripts from a normal control and an MCI patient whose MoCA scores are matched to 29 (out of 30 points). For the annotation of meta-semantic representation in Section 3, all transcripts were tokenized by the open-source NLP toolkit called ELIT.[3] Table 2 shows general statistics of these transcripts from the output automatically generated by the part-of-speech tagger and the dependency parser in ELIT.

| Type | Control | MCI | P-value |
|---|---|---|---|
| T | 174.32 (±40.14) | 175.04 (±48.01) | 0.936 |
| S | 11.34 (±3.08) | 11.22 (±3.73) | 0.862 |
| N | 36.32 (±8.62) | 38.06 (±12.25) | 0.418 |
| V | 27.10 (±7.44) | 24.50 (±6.93) | 0.077 |
| C | 7.74 (±4.25) | 7.54 (±4.42) | 0.820 |
| RN | 2.36 (±1.82) | 1.64 (±1.67) | **0.044** |
| CM | 4.52 (±2.74) | 4.30 (±2.15) | 0.659 |

Table 2: Statistics of transcripts from the two groups. The avg-count and the stdev are reported for each field. T: tokens, S: sentences, N: nouns, V: verbs, C: conjuncts, RN: relative clauses and non-finite modifiers, CM: clausal modifiers or complements. The p-values are evaluated by the t-test.

[1] B-SHARP: http://medicine.emory.edu/bsharp

[2] Temi: https://www.temi.com

[3] ELIT: https://github.com/elitcloud/elit

| Control | MCI |
|---|---|
| This is a what looks like a circus poster. The title is the Circus Procession. There's an off color green background. On the left-hand side is elephant in a costume peddling a tricycle, operating a tricycle. On the right side is another elephant with holding a fan. He's dressed in an outfit with a hat and a cane. There are two people in the background and they could be either men or women. And then there are three, I'll take that back. And then the foreground is a clown in a white suit with red trim. It was copyrighted in 1988 by the McLoughlin Brothers, New York or NY. Um, there's a black border. Um, the, there are shadows represented by some brown color at the bottom. | It's a circus poster. Going left to right is an elephant standing on its side legs, and a, um, vest, a tie and a red Tuxedo coat, and um yellow cap with a black band holding what appears to be a fan in its trunk. The elephant has glasses and a cane. Um, the top, says the Circus Procession. To the left of the elephant is a clown in a white and red costume with red and black paint on his face, red hair or shoes. And there appear to be three like soldiers, um gray suits, yellow trim, um, um, red hair. To the left of them, there's another elephant, riding a bicycle. This elephant has pants to red bicycle. He's got a regular coat of his and a red bow tie. |

Table 3: Transcripts from a normal control and an MCI patient whose MoCA scores are 29 points.

No significant group differences were found in text-level counts (tokens and sentences), grammatical categories (nouns and verbs), or syntactic structures (conjuncts, clausal modifiers or complements), except for the relative clauses and non-finite modifiers whose p-value is less than 0.05. The MCI group used notably a fewer number of verbs although the difference to the control group was not significant.

## 3 Meta-Semantic Representation

We organized a team of two undergraduate students in Linguistics to annotate meta-semantic representation on the transcripts from Section 2.2 such that every transcript was annotated by two people and adjudicated by an expert. The web-based annotation tool called BRAT was used for this annotation (Stenetorp et al., 2012), where the entire content of each transcript was displayed at a time. Figure 2 shows a screenshot of our annotation interface using BRAT on the control example in Table 3.

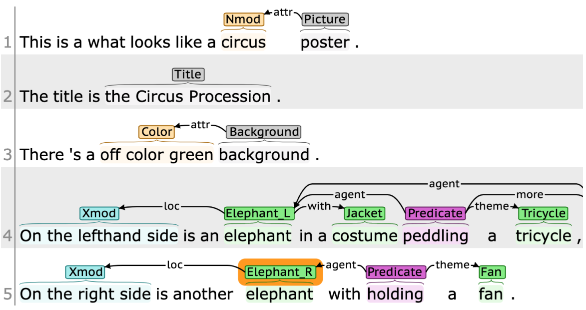


Figure 2: A screenshot of our annotation interface using the web-based tool BRAT on the first five sentences of the control example in Table 3.

Meta-semantic representation involves three types of annotation, entities (Section 3.1), predicate argument structures (Section 3.2), discourse attributes (Section 3.3), as well as few other miscellaneous components (Section 3.4). The following sections give a brief overview of our annotation guidelines.

### 3.1 Entities

To analyze which and how objects in the picture are described by individual subjects, every object mentioned in the transcript is identified as either a predefined entity or an unknown entity. All nominals including pronouns, proper nouns, common nouns, and noun phrases are considered potential mentions. Table 4 shows the list of 50 predefined entities that are frequently mentioned in the transcripts.

| Main Entity | Sub Entities |
|---|---|
| `Picture` | `Background`, `Border`, `Copyright`, `Parade`, `Shadow`, `Title` |
| `Elephant_L` | `EL_Beanie`, `EL_Collar`, `EL_Head`, `EL_Jacket`, `EL_Pants`, `EL_Tie`, `EL_Tricycle`, `EL_Trunk` |
| `Elephant_R` | `ER_Fedora`, `ER_Coat`, `ER_Vest`, `ER_Cane`, `ER_Fan`, `ER_Glasses`, `ER_Head`, `ER_Collar`, `ER_Pants`, `ER_Tie`, `ER_Hand`, `ER_Feet`, `ER_Trunk`, `ER_Hanky` |
| `Men` | `Man_L`, `Man_M`, `Man_R`, `M_Boots`, `M_Costume`, `M_Cross`, `M_Flag`, `M_Hat`, `M_Plume`, `M_Sword` |
| `Clown` | `CL_Face`, `CL_Hair`, `CL_Head`, `CL_Pants`, `CL_Ruffle`, `CL_Shoes`, `CL_Suit` |

Table 4: Predefined entities, where the main entities indicate the 5 conspicuous objects in Figure 1 and the sub entities indicate objects that belong to the main entities.

In the example below, five mentions are found and can be linked to four entities as follows:

> An *elephant*$_1$ is holding a *fan*$_2$. To the leftside of *him*$_3$, another *elephant*$_4$ is riding a *tricycle*$_5$.

- $\{elephant_1, him_3\} \rightarrow$ `Elephant_R` (elephant on the right)
- $\{elephant_4\} \rightarrow$ `Elephant_L` (elephant on the left)
- $\{fan_2\} \rightarrow$ `ER_Fan`
- $\{tricycle_5\} \rightarrow$ `EL_Tricycle`

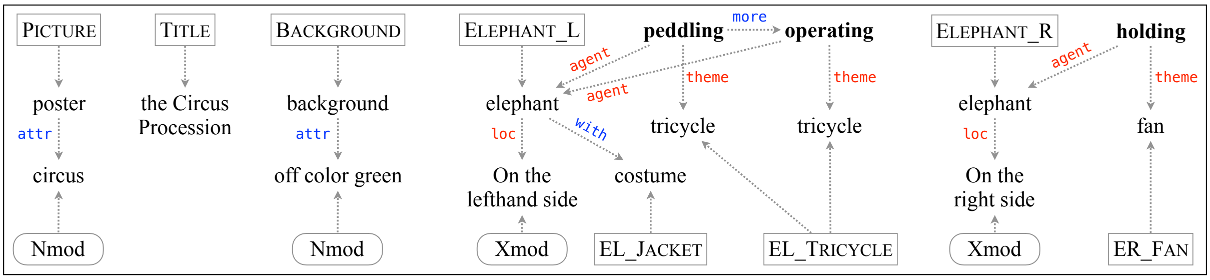


Figure 3: Visualization of meta-semantic representation on the first 5 sentences of the control example in Table 3.

The entity `Men` is a group of three people including `Man_L`, `Man_M`, and `Man_R` (man on the left, middle, and right) as its sub entities. Such a group entity is defined because subjects regularly describe them together as one unit. `Picture` often refers to the types of the picture that subjects view it as (e.g., *poster* in Figure 2). Special kinds of entities, `Title` and `Copyright`, are also defined that are annotated on the literals (e.g., *the Circus Procession* in Figure 2, *McLoughlin Brothers, 1888, N.Y.*) to see if subjects indeed recognize them correctly. Any object that is either ambiguous or not predefined is annotated as an unknown entity.

It is worth mentioning that unlike mention annotation for coreference resolution in OntoNotes (Pradhan et al., 2012) where whole noun phrases are annotated as mentions, function words such as articles or determiners and modifiers such as adverbs or adjectives are not considered part of mentions in our annotation, which is similar to abstract meaning representation (Banarescu et al., 2013). Such abstraction is more suitable for spoken data where the usage of these function words and modifiers is not so consistent.

## 3.2 Predicate Argument Structures

To analyze semantics of the entities as well as their relations to one another, predicate argument structures are annotated. Note that meta-semantic representation is entity-centric such that expressions that do not describe the picture are discarded from the annotation (e.g., *When I was young, circus came to my town all the time*). Such expressions do not help analyzing subjects' capabilities in describing the picture although they can be used for other kinds of analyses which we will explore in the future.

Following the latest guidelines of PropBank (Bonial et al., 2017), both verbal predicates, excluding auxiliary and modal verbs, and nominal predicates, including eventive nouns and nouns from light-verb constructions, are considered in our representation.

Once predicates are identified, arguments are annotated with the following thematic roles (in the examples, predicates are in italic, arguments are in brackets, and thematic roles are in subscripts):

- `agent`: Prototypical agents
  e.g., An [elephant]$_{\text{agent}}$ is *holding* a fan.
- `theme`: Prototypical patients or themes
  e.g., An elephant is *holding* a [fan]$_{\text{theme}}$.
- `dative`: Recipients or beneficiaries e.g., The soldier is *bringing* a flag to the [circus]$_{\text{dative}}$.
- `adv`: Adverbial modifiers
  e.g., That elephant is [actually]$_{\text{adv}}$ *walking*.
- `dir`: Directional modifiers
  e.g., Feathers are *coming* out of the [hat]$_{\text{dir}}$.
- `loc`: Locative modifiers e.g., The *clown* is *dancing* in between the [elephants]$_{\text{loc}}$.
- `mnr`: Locative modifiers
  e.g., Soldiers are *marching* [proudly]$_{\text{mnr}}$.
- `prp`: Purpose or clausal modifiers e.g., The clown is *dancing* to [tease]$_{\text{prp}}$ the elephants.
- `tmp`: Temporal modifiers e.g., This seemed to be a poster *made* in the early [1900s]$_{\text{tmp}}$.

If an argument is a preposition phrase, the thematic role is annotated on the preposition object such that in the example above, only the head noun [hat] is annotated as `dir` instead of the entire preposition phrase "out of the hat".[4] As shown in the `prp` example, a predicate can be an argument of another predicate. Note that modifiers do not need to be arguments of only predicates but entities as well (e.g., the *elephant* on the [tricycle]$_{\text{loc}}$, a *poster* from way back in [1990s]$_{\text{tmp}}$).

The choice of these thematic roles are observational to the transcripts. No instance of `dative` is found in our dataset but the role is still kept in the guidelines for future annotation.

[4] See the `case` relation in Section 3.4 for more details about how prepositions are handled in our annotation.

### 3.3 Discourse Attributes

To analyze discourse aspects of the transcripts, six labels and one relation are annotated as follows (in the examples, attributes are indicated in brackets):

**ambiguous** Objects contextually ambiguous to identify are annotated with this label. For example, both [elephant] and [something] are annotated as `ambiguous` because it is unclear which elephant and object they refer to. Also, [blue] likely refers to the vest of `Elephant_R` but not specified in this context; thus, it is also annotated as `ambiguous`.

That [elephant] is holding [something].
The elephant with [blue] on is walking.

**opinion** Descriptions subjective to that particular subject are annotated with this label. For example, 'red' is considered an objective fact agreed by most subjects whereas [fancy] is considered a subjective `opinion`, not necessarily agreed by others. Similarly, [like a millionaire] is considered subject's `opinion` about the elephant's costume.

The 'red 'tie with the [fancy] shirt.
That elephant is dressed up [like a millionaire].

**emotion** Expressions that carry subjects' emotions or their views on objects' emotions are annotated with this label.

That clown looks [happy].
The elephant makes me [sad].

**certain** Adverbials or modals that express certainty are annotated with this label.

Those people [must] be women.
This is [obviously] an old poster.

**fuzzy** Adverbials or modals that express fuzziness are annotated with this label.

The elephant carries [some kind of] balloon.
I am [not sure] if the elephant is marching.

**emphasis** Adverbials used for emphasis are annotated with this label.

That tricycle is [very] big.
That clown is [definitely] enjoying this.

**more** Additional descriptions from appositions and repetitions from repairs are annotated with this relation (in the examples, ones in the brackets have `more` relations to the ones in italic):

There are *elephants*, two [elephants]$_{\text{more}}$, here.
This is the Circus *Profession*, [Procession]$_{\text{more}}$.
That one is holding an *umbrella*, or a [fan]$_{\text{more}}$.

[elephants] is an apposition that adds `more` information to *elephants*. [Procession] is a prototypical repair case that fixes the prior mention of *Profession*. [fan] may not be considered a repair in some analysis, but it is in ours because it attempts to fix the earlier mention of *umbrella* in a speech setting.

### 3.4 Miscellaneous

Two additional modifiers, `Nmod` and `Xmod` are annotated. `Nmod` are modifiers of nominals that modify entities with the `attr` relation:

A [polka dot]$_{\text{attr}}$ *dress*.
Very [big]$_{\text{attr}}$ [red and yellow]$_{\text{attr}}$ *pants*.

`Xmod` are any other types of modifiers, mostly adverbials and prepositions. If adverbials, they are annotated with the `adv` relation in Sec 3.2. If prepositions, they are annotated with the `case` relation:

There is a [seemingly]$_{\text{adv}}$ *dancing* clown.
Feathers are coming [out of]$_{\text{case}}$ the *hat*.

Finally, possessions of entities are annotated with the `with` relation regardless of verbs such as *have* or *get* for the consistency across different structures. In both of the following sentences, [jacket] has the `with` relation to the *elephant*.

The *elephant* with a blue [jacket]$_{\text{with}}$.
The *elephant* has a blue [jacket]$_{\text{with}}$.

## 4 Meta-Semantic Analysis

Given the annotation in Section 3, several analyses are made to observe how effective meta-semantic is to distinguish the control ($C$) and MCI ($M$) groups.

### 4.1 Entity Coverage Analysis

We anticipate that most subjects in $C$ and $M$ would recognize the main entities whereas a fewer number of sub entities would be commonly recognized by $M$ than $C$. For each entity $e_i$, that is the $i$'th entity in Table 4, two counts $c_i^c$ and $c_i^m$ are measured such that they are the numbers of subjects in $C$ and $M$ whose transcripts include at least one mention of $e_i$. For instance, the entity $e_7$ = `Title` is mentioned by $c_7^c = 37$ subjects in $C$ and $c_7^m = 40$ subjects in $M$ in our annotation.

Figure 4 shows how many entities are commonly mentioned by each percentage range of the subjects in $C$ and $M$. For example, six entities are commonly mentioned by 55∼75% of the subjects in $C$ whereas only three entities are commonly mentioned by the same range of the subjects in $M$. These percentage ranges are analyzed as follows:

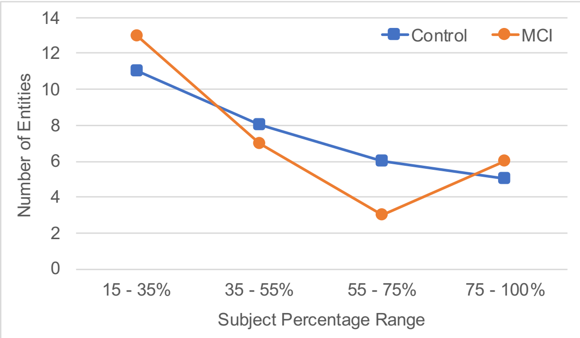


Figure 4: Entity coverage analysis.

**High range (75∼100%)** No significant group difference is found between $C$ and $M$. 5 entities, `Elephant_R`, `Elephant_L`, `EL_Tricycle`, `Clown`, and `Men`, are commonly mentioned by $C$, whereas 6 entities (all of above + `Title`) are commonly mentioned by $M$ in this range.

**Mid range (35∼75%)** Subjects in $M$ start not recognizing certain entities recognized by subjects in $C$ in this range. 14 entities are commonly mentioned by $C$ whereas 10 entities are mentioned by $M$. When the range is fine-grained to 45∼75%, the difference becomes even more significant such that 10 entities are commonly mentioned by $C$ whereas only 5 entities are mentioned by $M$ in that range.

**Low range (15∼35%)** Similar to the high range, no significant difference is found between the two groups. 11 and 13 entities are commonly recognized by $C$ and $M$, respectively in this range.

For the whole range of 15∼75%, the plot from $C$ can be well fitted to a linear line with $R^2 = 0.9524$, whereas the one from $M$ cannot, resulting significantly lower $R^2 = 0.5924$. The plot from $M$ rather shows an inverted Gaussian distribution, implying that the majority of $M$ tends not to mention about entities that are not immediately conspicuous which is not necessarily the case for subjects in $C$.

## 4.2 Entity Focus Analysis

This analysis shows which entities are more frequently mentioned (focused) by what subject group. For each entity $e_i$ and its counts $c_i^c$ and $c_i^m$ in Section 4.1, the proportions $p_i^c$ and $p_i^m$ are measured such that $p_i^c = c_i^c/|C|$ and $p_i^m = c_i^m/|M|$, where $|C| = |M| = 50$ (Table 1). Then, the relative difference $d_i^r$ for the $i$'th entity is measured as follow:

$$d_i^r = \frac{p_i^c - p_i^m}{\max(p_i^c, p_i^m)}$$

Thus, if $d_i^r$ is greater than 0, $e_i$ is more commonly mentioned by $C$; otherwise, it is by $M$. Figure 4 shows the entities that are significantly more mentioned by $C$ (blue) and $M$ (red), where $|d|_i^r \geq 0.2$. 6 entities, `CL_Pants`, `M_Boots`, `ER_Glasses`, `EL_Collar`, `ER_Trunk`, `M_Flag`, `EL_Pants`, `ER_Vest`, and `EL_Jacket`, are noticeably more mentioned by $C$, whereas only 2 entities, `EL_Tie` and `EL_Hat`, are by $M$, which are focused on those two small parts of the left elephant. Additionally, $M$ mentions more about the `Background`, which is not a specific object but an abstract environment.

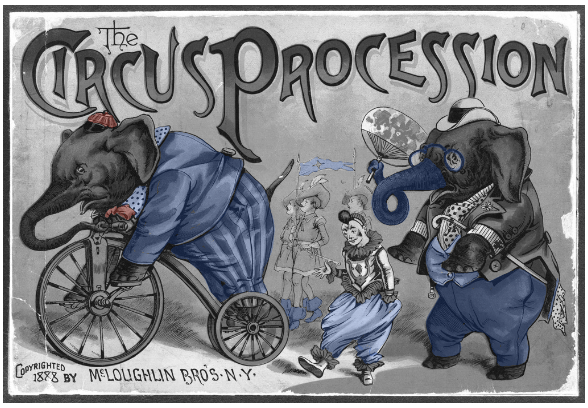


Figure 5: Entity focus analysis. Entities focused by $C$ and $M$ are colored in blue and red, respectively.

## 4.3 Entity Density Analysis

This analysis shows the proportion of the description used for each object in the transcript. Meta-semantic representation forms a graph comprising many isolated subgraphs. In Figure 3, there are 5 subgraphs, where the largest subgraph has 7 vertices (the one with `Elephant_L`) and the smallest subgraph has only 1 vertex (the one with `Title`).

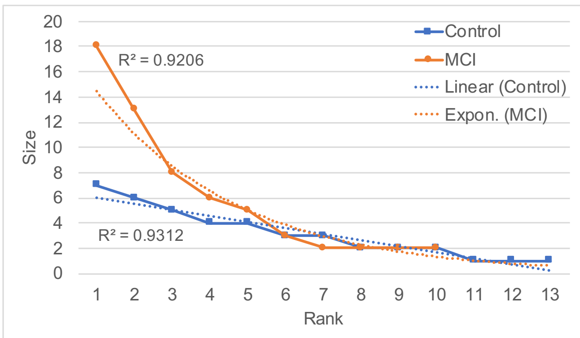


Figure 6: Plots of size lists derived from meta-semantic representation annotated on the control and MCI examples in Table 3, where $x$ and $y$ axises are ranked indices and sizes of the subgraphs, respectively.

Let $G^t$ be a graph derived from meta-semantic representation annotated on the $t$'th transcript. $G^t$ can be represented by a list of its subgraphs sorted in descending order with respect to their sizes such that $G^t = [g_1^t, \ldots, g_k^t]$ where $|g_i| \geq |g_j|$ for all $0 < i < j \leq k$. The size of a subgraph is determined by the number of vertices. For the graph in Figure 3, $G = [g_1, \ldots, g_5]$ such that $|G| = k = 5$, $|g_1| = 7$, and $|g_5| = 1$. Given $G^t$, the size list $L^t$ can be derived such that $L^t = [|g_1^t|, \ldots, |g|_k^t]$. Figure 6 shows plots of the size lists from the graphs derived by meta-semantic representation annotated on the control and MCI examples in Table 3. The control plot can be well-fitted to a linear line with $R^2 = 0.9312$, whereas the MCI plot is better fitted to an exponential curve with $R^2 = 0.9206$.

| $\overline{SSE}_d$ | Control | MCI | P-value |
|---|---|---|---|
| $d = 1$ | 12.10 (±12.37) | 17.50 (±22.43) | 0.1394 |
| $d = 2$ | 5.18 (±4.81) | 7.08 (±7.54) | 0.1370 |
| $d = 3$ | 3.03 (±2.44) | 4.01 (±3.83) | 0.1278 |
| $d = 4$ | 2.36 (±2.14) | 2.55 (±2.12) | 0.6661 |
| $d = 5$ | 1.89 (±1.78) | 1.78 (±1.54) | 0.7391 |

Table 5: The average sums of squared errors by fitting each size list to degrees 1-5 of polynomial functions.

Table 5 shows the average sums of squared errors $\overline{SSE}_d$ by fitting each size list $L^t = [l_1^t, \ldots, l_k^t]$ to polynomial functions $f_d(x)$ of degrees $d = [1, .., 5]$ where $n = 50$ for both $C$ and $M$:

$$\overline{SSE}_d = \frac{1}{n} \sum_{t=1}^{n} \sum_{i=1}^{|L^t|} (f_d(i) - l_i^t)^2$$

The control plots fit to lower degree functions more reliably than the MCI plots, although not statistically significant, implying that subjects in $C$ distribute their time more evenly to describe different entities than subjects in $M$ who tend to spend most of their time to describe a couple of entities but not so much for the rest of the entities.

### 4.4 Predicate Argument Analysis

Figure 7 shows the average percentages of predicates and their thematic arguments annotated on the transcripts. Subjects in $C$ generally form sentences with more predicate argument structures although the differences are not statistically significant. Not enough instances of the modifiers (e.g., `mnr`, `loc`) are found to make a meaningful analysis for those roles. Although predicate argument structures may not appear useful, these structures make it possible to perform the entity density analysis in Section 4.3 and potentially other types of analyses, which we will conduct in the future.

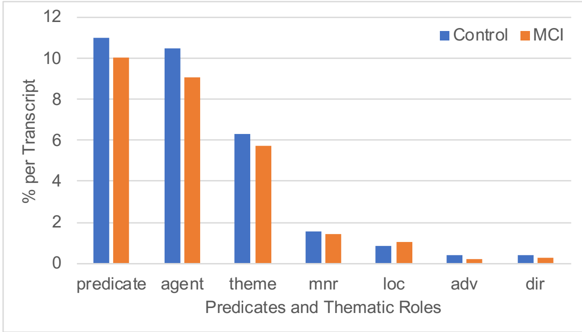


Figure 7: Predicate argument analysis.

### 4.5 Discourse Attribute Analysis

Figure 8 shows the average percentages of discourse attributes. Notice that $M$ makes over twice more `ambiguous` mentions than $C$, implying that MCI patients do not elaborate as well. Moreover, $M$ makes more `fuzzy` expressions and frequently uses `more` relations to repair, which makes their speeches less articulated. On the other hand, $C$ makes more subjective `opinion` and `certain` expressions with `emphasis`, which makes their speeches sound more confident. These are essential features to distinguish $M$ from $C$, makes this analysis more "meta-semantics".

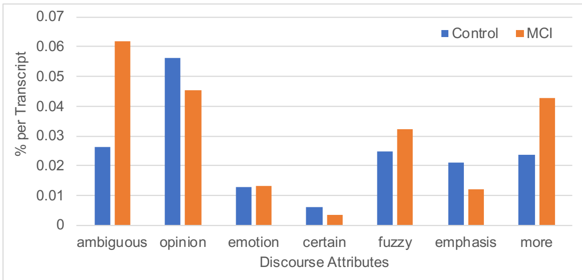


Figure 8: Discourse attribute analysis.

## 5 Experiments

### 5.1 Inter-Annotator Agreement

The annotation guidelines summarized in Section 3 are established through multiple rounds of double annotation and adjudication. During the final round, the entity annotation, the predicate argument annotation, and the discourse attribute annotation reach the F1 scores of 88%, 82%, and 89% respectively for the inter-annotator agreement, which yield high-quality data ready for training statistical models.

### 5.2 Data Split

The 100 transcripts from Section 2 are split into 5 folds where each fold contains 10 transcripts from the control group and another 10 transcripts from the MCI group (so the total of 20 transcripts). To evaluate our model that takes a transcript annotated with meta-semantic representation as input and predicts whether or not it is from the MCI group, 5-fold cross validation is used, which is suitable for experimenting with such a small dataset.

### 5.3 Features

For each transcript, three types of features are extracted from the meta-semantic analysis in Section 4 for the classifications of Control vs. MCI:

- **Entity Types**: A vector $e \in \mathbb{R}^{1\times|E|}$ is created where $|E| = 50$ is the total number of predefined entities in Table 4, and each dimension $i$ of $e$ represents the occurrence of the corresponding entity such that $e_i = 1$ if the $i$'th entity appears in the transcript; otherwise, $e_i = 0$.

- **Entity Densities**: A vector $d \in \mathbb{R}^{1\times|P|}$ is created where $P = \{1, 2, 3\}$ ($|P| = 3$) consisting of degrees used for the entity density analysis in Section 4.3 (in this case, the polynomial functions with degrees 1, 2, and 3 are used) such that $d_i$ is the sum of the squared error measured by comparing the size list $L$ of this transcript to the fitted polynomial function of the degree $i$.

- **Labels**: A vector $b \in \mathbb{R}^{1\times|N|}$ is created where $N$ contains counts of predicates, thematic roles, and discourse attributes in Sections 3.2 and 3.3 ($|N| = 16$) such that $b_i$ is the count of the corresponding component in the transcript.

### 5.4 Classification

The feature vector $x = e \oplus d \oplus b$ is created by concatenating $e$, $d$, and $b$, and gets fed into a classifier. Figure 9 illustrates the feed-forward neural network used for the classification between the control and the MCI groups. Let the size of the feature vector $x$ be $s = |E| + |P| + |L|$. Then, the input vector $x \in \mathbb{R}^{1\times s}$ is multiplied by the weight matrix $W_0 \in \mathbb{R}^{s\times d_0}$ and generates the first hidden vector $h_1 = x \cdot W_0$. The hidden vector $h_1 \in \mathbb{R}^{1\times d_0}$ is multiplied by another weight matrix $W_1 \in \mathbb{R}^{d_0\times d_1}$ and generates the second hidden vector $h_2 = h_1 \cdot W_1$. Finally, $h_2 \in \mathbb{R}^{1\times d_1}$ is multiplied by the last weight matrix $W_2 \in \mathbb{R}^{d_1\times d_2}$ where $d_2$ is the number of classes to be predicted, and generates the output vector $y = h_2 \cdot W_2 \in \mathbb{R}^{1\times d_2}$. In our case, the sizes of the hidden vectors are $d_0 = 200$ and $d_1 = 100$, and the size of the output vector is $d_2 = 2$. Note that we have experimented with simpler networks comprising only one or no hidden layer, but the one with two hidden layers shows the best results.

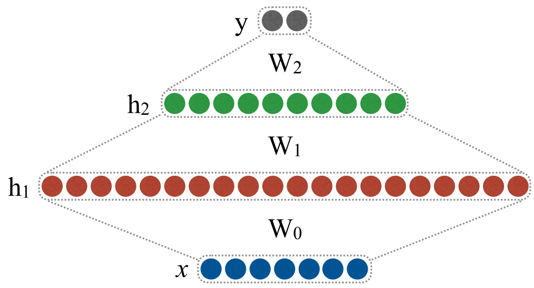


Figure 9: Feed-forward neural network used for the classification of the control vs. MCI group.

The two dimensions $y_m$ and $y_c$ in the output vector are optimized for the likelihoods of the subject being control or MCI, respectively. The average of 82% accuracy is achieved by the 5-fold cross-validation (Section 5.2) with this model. Considering these are subjects that the standardized tests such as MoCA or Boston Naming Test could not distinguish (Table 1), this result is very promising.

## 6 Related Work

Reilly et al. (2010) found that neurodegenerative disorders could deteriorate nerve cells controlling cognitive, speech and language processes. Verma and Howard (2012) reported that language impairment in AD could affect verbal fluency and naming, that requires integrity of semantic concepts, before breaking down in other facets of the brain. Tillas (2015) showed that linguistic clues captured from verbal utterances could indicate symptoms of AD.

Toledo et al. (2018) investigated the significance of lexical and syntactic features from verbal narratives of AD patients by performing several statistical tests based on 121 elderly participants consisting of 60 patients with AD and 61 control subjects. In this work, immediate word repetitions, word revisions, and coordination structures could be used to distinguish patients with AD from the control group. Mueller et al. (2018) recently found that AD patients often depicted less informative discourse, greater impairment in global coherence, greater modularization, and inferior narrative structure compared to the normal control group.

## References


Laura Banarescu, Claire Bonial, Shu Cai, Madalina Georgescu, Kira Griffitt, Ulf Hermjakob, Kevin Knight, Philipp Koehn, Martha Palmer, and Nathan Schneider. 2013. Abstract Meaning Representation for Sembanking. In *Proceedings of the 7th Linguistic Annotation Workshop and Interoperability with Discourse*, LAW-ID'13, pages 178–186.

A. L. Beam and I. S. Kohane. 2016. Translating Artificial Intelligence Into Clinical Care. *Journal of the American Medical Association*, 316(22):2368–2369.

Claire Bonial, Kathryn Conger, Jena D. Hwang, Aous Mansouri, Yahya Aseri, Julia Bonn, Timothy O'Gorman, and Martha Palmer. 2017. Current Directions in English and Arabic PropBank. In Nancy Ide and James Pustejovsky, editors, *Handbook of Linguistic Annotation*, pages 737–769. Springer.

Kathleen C. Fraser, Jed A. Meltzer, and Frank Rudzicz. 2016. Linguistic Features Identify Alzheimer's Disease in Narrative Speech. *Journal of Alzheimer's Disease*, 49(2):407–422.

Denise C. Fyffe, Shubhabrata Mukherjee, Lisa L. Barnes, Jennifer J. Manly, David A. Bennett, and Paul K. Crane. 2011. Explaining Differences in Episodic Memory Performance among Older African Americans and Whites: The Roles of Factors Related to Cognitive Reserve and Test Bias. *Journal of the International Neuropsychological Society*, 17(4):625–638.

Edith Kaplan, Harold Goodglass, and Sandra Weintraub. 1983. *The Boston Naming Test*. Philadelphia: Lea and Febiger.

Sweta Karlekar, Tong Niu, and Mohit Bansal. 2018. Detecting Linguistic Characteristics of Alzheimer's Dementia by Interpreting Neural Models. In *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics*, NAACL'18, pages 701–707.

Justin E. Karr, Raquel B. Graham, Scott M. Hofer, and Graciela Muniz-Terrera. 2018. When does cognitive decline begin? A systematic review of change point studies on accelerated decline in cognitive and neurological outcomes preceding mild cognitive impairment, dementia, and death. *Psychology and Aging*, 33(2):195–218.

John C. Morris. 1994. The Clinical Dementia Rating (CDR): current version and scoring rules. *Neurology*, 43(11):2412–2414.

Kimberly Diggle Mueller, Bruce P. Hermann, Jonilda Mecollari, and Lyn Turkstra. 2018. Connected speech and language in mild cognitive impairment and Alzheimer's disease: A review of picture description tasks. *Journal of Clinical and Experimental Neuropsychology*, 40(1):917–939.

Ziad S. Nasreddine, Natalie A. Phillips, Valérie Bédirian, Simon Charbonneau, Victor Whitehead, Isabelle Collin, Jeffrey L. Cummings, and Howard Chertkow. 2005. The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. *Journal of the American Geriatrics Society*, 53(4):695–699.

Sylvester O. Orimaye, Jojo S-M. Wong, Karen J. Golden, Chee P. Wong, and Ireneous N. Soyiri. 2017. Predicting probable Alzheimer's disease using linguistic deficits and biomarkers. *BMC Bioinformatics*, 18(34).

Sylvester Olubolu Orimaye, Jojo Sze-Meng Wong, and Judyanne Sharmini Gilbert Fernandez. 2016. Deep-Deep Neural Network Language Models for Predicting Mild Cognitive Impairment. In *Proceedings of the Workshop on Advances in Bioinformatics and Artificial Intelligence: Bridging the Gap*.

R. I. Pfeffer, T. T. Kurosaki, C.H. Jr. Harrah, J. M. Chance, and S. Filos. 1982. Measurement of Functional Activities in Older Adults in the Community. *Journal of Gerontology*, 37(3):323–329.

Sameer Pradhan, Alessandro Moschitti, Nianwen Xue, Olga Uryupina, and Yuchen Zhang. 2012. CoNLL-2012 Shared Task: Modeling Multilingual Unrestricted Coreference in OntoNotes. In *Proceedings of the Sixteenth Conference on Computational Natural Language Learning: Shared Task*, CoNLL'12, pages 1–40.

Jamie Reilly, Amy Rodriguez, Martine Lamy, and Jean Neils-Strunjas. 2010. Cognition, Language, and Clinical Pathological Features of Non-Alzheimer's Dementias: An Overview. *Journal of Communication Disorders*, 43(5):438–452.

Kyle Steenland, Liping Zhao, Samantha E. John, Felicia C. Goldstein, Allan Levey, and Alvaro Alonso. 2018. A 'Framingham-like' Algorithm for Predicting 4-Year Risk of Progression to Amnestic Mild Cognitive Impairment or Alzheimer's Disease Using Multidomain Information. *Journal of Alzheimer's Disease*, 63(4):1383–1393.

Pontus Stenetorp, Sampo Pyysalo, Goran Topić, Tomoko Ohta, Sophia Ananiadou, and Jun'ichi Tsujii. 2012. brat: a Web-based Tool for NLP-Assisted Text Annotation. In *Proceedings of the Demonstrations at the 13th Conference of the European Chapter of the Association for Computational Linguistics*, EACL'12, pages 102–107.

Alexandros Tillas. 2015. Language as Grist to the Mill of Cognition. *Cognitive Processing*, 16(3):219–243.

Cíntia Matsuda Toledo, Sandra M. Aluisio, Leandro Borges dos Santos, Sonia Maria Dozzi Brucki, Eduardo Sturzeneker Trés, Maira Okada de Oliveira, and Letícia Lessa Mansura. 2018. Analysis of macrolinguistic aspects of narratives from individuals with Alzheimer's disease, mild cognitive impairment, and no cognitive impairment. *Alzheimer's Dementia*, 10:31–40.

Myank Verma and Robert J. Howard. 2012. Semantic memory and language dysfunction in early Alzheimer's disease: a review. *International Journal of Geriatric Psychiatry*, 27(12):1209–1217.