A Comparison of Syntax, Semantics, and Pragmatics in Spoken Language among Residents with Alzheimer's Disease in Managed-Care Facilities

Curry Guinn
Department of Computer Science
UNC Wilmington
guinnc@uncw.edu

Ben Singer
Department of Computer Science
UNC Wilmington
bjs4203@uncw.edu

Anthony Habash
Department of Computer Science
UNC Wilmington
redwolf98@gmail.com

***Abstract*— This research is a discriminative analysis of conversational dialogues involving individuals suffering from dementia of Alzheimer's type. Several metric analyses are applied to the transcripts of the Carolina Conversation Corpus in order to determine if there are significant statistical differences between individuals with and without Alzheimer's disease. Our prior research suggests that there exist measurable linguistic differences between managed-care residents diagnosed with Alzheimer's disease and their caregivers. This paper presents results comparing managed-care residents diagnosed with Alzheimer's disease to other managed-care residents. Results from the analysis indicate that part-of-speech and lexical richness statistics may not be good distinguishing attributes. However, go-ahead utterances and certain fluency measures provide defensible means of differentiating the linguistic characteristics of spontaneous speech between individuals that are and are not diagnosed with Alzheimer's disease. Two machine learning algorithms were able to classify the speech of individuals with and without dementia of the Alzheimer's type with accuracy up to 80%.**

Keywords—Natural language processing; NLP in healthcare, text classification; Alzheimer's disease.

I. Introduction

The Alzheimer's Association reports that an estimated 5.2 million Americans have Alzheimer's disease in 2014[2]. During the course of the disease, individuals lose cognitive-communication skills in predictable ways [3]. This decline in language facility can result in decreased social interaction and life satisfaction for persons with AD and their caregivers. In particular, persons with AD begin to feel a loss of their personal identity.

Individuals suffering from dementia of Alzheimer's type are characterized as being afflicted with a loss in cognitive and communicative functionality [4]. This condition is often reflected within their powers of communication with 88 – 95% of individuals with Alzheimer's portraying some degree of aphasia (language disability) and cognitive failure including the inability to grasp concepts, events of their past, or the ability to recognize individuals [5].

Spontaneous speech, the focus of this research, is characterized as allowing for no pre-emptive planning or memorization of a response and demanding the highest level of cognitive action and memory contemplation to produce accurate and effective responses. This research attempts to detect quantitative signs of degradation in speech and cognitive capacities within individuals suffering from Alzheimer's.

II. Language Pathologies in Dementia of the Alzheimer's Type

The vast majority of individuals with Alzheimer's are characterized by the degradation of their language and cognitive functionality, resulting in significant complications in vocal communication. How to properly analyze and treat individuals with Alzheimer's is an ongoing question. In the majority of cases of individuals diagnosed with Alzheimer's disease, the patient has likely suffered the condition for several years before the symptoms became evident enough for testing [6]. This late diagnosis makes a comparison of previous cognitive and communication ability within the patient with Alzheimer's prior to the onset of the disease extremely difficult if not impossible. Therefore, a broad, general comparison of the communication abilities of individuals with Alzheimer's to those of individuals who do not suffer the condition might prove more accessible in determining signs of cognitive and linguistic degradation.

III. Computational Models of Speech in Individuals with Alzheimer's

Some previous quantitative studies of the pathologies (in general) of speech include the use of pauses, fillers, formulaic speech, restarts, repeats, incomplete statements and speech disfluencies [7, 8, 9, 10]. The degradation of speech capacity exhibited within individuals with Alzheimer's should include these factors. Also, past research suggests that persons with Alzheimer's exhibit less lexical richness [10].

There is a complication in using these quantitative measures for contrasting individuals with and without dementia of the Alzheimer's type: namely, it can be hard to gauge an individual's pre-existing speech capacity. Also, pre-existing studies that claim effective discrimination of individuals with Alzheimer's and healthy individuals using these computational models have done so using a limited and size-controlled corpus for analysis, as well as using interview-style dialogue instead of spontaneous speech. For instance, the Bucks et al. study used 8 individuals with dementia of the Alzheimer's type [4]. The experiment described in this paper uses a substantially larger corpus (n = 56, with 28 managed-care residents diagnosed with Alzheimer's Disease and 28 residents not diagnosed with Alzheimer's Disease) than previous research.

IV. Literature Review and Analysis Methods

Several studies have used computational methods to examine the speech of individuals with dementia of the Alzheimer's type. Several projects have examined the level of speech degradation in relevance to age, gender, relationships, topics, role, and stages of dementia [4, 9]. Most of this research has focused on two dimensions of speech: lexical richness [4, 9, 11, 12] and/or speech fluency [4, 7, 8, 10].

A. *Corpus Annotation*

The corpus of spontaneous speech conversations involving individuals with Alzheimer's that will be used in this study is the Carolina Conversations Collection (CCC). The CCC Corpus, developed in a partnership between the Medical University of South Carolina and the University of North Carolina Charlotte, was constructed in an effort to provide useful data concerning speech patterns with regards to age, gender, social identities, health and illness stories, and explanatory models of disease [1].

The CCC corpus consists of over 400 transcribed conversations with 125 multiethnic, older individuals suffering from any number of possible conditions that have been individually categorized [1]. All subjects were living in managed-care facilities. These conversations were originally recorded in only audio format and eventually transcribed into text using the methods defined by Ten Have [13]. For the study described in this paper, we use the transcripts of 28 subjects that were diagnosed with Alzheimer's disease (77 total transcripts), and the transcripts of 28 subjects there had not been diagnosed with Alzheimer's Disease (204 total transcripts). To compensate for multiple transcripts involving the same person, all data gathered from transcripts involving the same subject was conglomerated to produce one set of statistics for each subject. A dialog fragment from the CCC is given in Figure 1.

Interviewer	When you was little did you work on the farm, in the fields?
Subject	Uh huh.
Interviewer	What did you do in the fields?
Subject	Cotton cotton cotton and horses and all kinds of stuff.
Interviewer	Did you mind picking cotton?
Subject	We, we just had to do it.
Interviewer	Yeah Okay. What did you do with the horses?
Subject	Fields, some where they uh, but they uh, but other--
Interviewer	Mmm hmm animals. What other animals were there? Like some pigs and hogs and chickens?
Subject	Um hmm.
Interviewer	Okay. Did y'all grow any types of like corn or something?
Subject	Mmm Hmm.

Figure 1 A fragment of a conversation within the Carolina Conversations Collection. Subject has dementia of the Alzheimer's type. Interviewer is a graduate student eliciting some life events from the subject.

In a previous publication, we describe a comparison of these conversations with conversations recorded among healthy adults [14, 15]. This study was a comparative analysis with the Switchboard Corpus, a collection of transcribed, spontaneous telephone conversations between healthy individuals [16]. The Switchboard Corpus was formulated as part of a project by Texas Instruments that was meant to address a growing need for large multispeaker databases of telephone bandwidth speech. Like with the CCC Corpus, conversations for the Switchboard Corpus were originally recorded via audio and then transcribed by hand into text using the methods defined by Ten Have [13].

B. *Syntactic Modeling*

Identifying the parts of speech within a speaker's dialogue allows for the break-down and categorization of the cognitive strength and capacity based on their lexical richness. This part-of-speech tagging is achieved using the Natural Language Toolkit, a series of analytical programs designed systematically to parse and tag raw text [17]. As the subject with Alzheimer's cognitive capacities fail, their ability to grasp facts such as names, places, and actions deteriorates. Past research reports a rise in the use of pronouns, and a decrease in the use of proper nouns, verbs, and adjectives [11].

C. *Semantic Modeling and Lexical Richness*

Several studies have established that a form of communication degradation caused by the onset of Alzheimer's is the weakening of an individual's vocabulary [4, 9, 12]

Three different forms of linguistic measurement are applied to the corpus to analyze the vocabulary richness employed by the speakers: Type Token Ratio (TTR), Brunet's Index (BI), and Honore's Statistic (HS) [4, 11]. These weighted measurements provide an applied approach to measuring the lexical richness of a dialog by providing algorithms for weighing the significance of unique vocabulary versus word count and total vocabulary.

Type-Token Ratio (TTR) provides a comparison to the total vocabulary used in a dialogue (V) to the total word count (N) of the dialogue.

$$\mathbf{TTR = V/N}$$

However, Type-Token Ratio does not account for the variations in word count, which is an uncontrolled factor in spontaneous conversation. Brunet's Index (BI) is unique from Type-Token Ratio in that it attempts to quantify the vocabulary used without considering the word count. Brunet's Index will then likely show a more applicable result when applied.

$$\mathbf{BI = N^{V(-0.165)}}$$

Honore's Statistic (HS) attempts a deeper analysis by accounting for words that are only used once (V_1), indicating a higher lexical richness.

$$\mathbf{HS = (100 \log N) / (1 - V_1/V)}$$

D. *Disfluency Modeling*

Distinct from a language's lexicon, disfluencies reflect a failure of concept more than a failure of vocabulary, where the speaker is uncertain, unclear, or doubtful of what they are trying to communicate. Spontaneous speech by its nature is commonly very disfluent [9]. Our analysis will track several features associated with disfluency.

1) *Filler Words*

Fillers are non-word and short phrase utterances that serve a communicative purpose with several possible meanings depending on their place [9]. When placed at the beginning or end of a dialogue they could possibly be "hints" given by the speaker in order to indicate that they had trouble understanding something or that they desire input. They could also serve to indicate that the speaker has misspoken and desires to take back what was said or reword it, which is shown more the case when they occur in the middle of a dialog than at the ends of it [9]. In general though, they do indicate some form of cognitive lapse, where the speaker fails to communicate properly. Hence we theorize that the rate of fillers would be higher for individuals with Alzheimer's. In the following example from a speaker with Alzheimer's disease, we see repeated use of the filler "uh"

> *But, uh, potato bread, we, uh,-- uh- pared the potatoes the night before. And, uh,- then,- uh, cut it all up into little pieces like that.*

2) *Repetitions*

Repetition is a notable coping mechanism for cognitive failure where the speaker will repeat a stated word or words in order to allow the information provided within the discourse to become more evident within their cognitive process. Repeats usually occur surrounding small pauses within an utterance where the speaker recaps or reaffirms to themselves what they are trying to communicate and then continue forth by picking up where they left off as in the following example:

> *"That was a very… very intriguing speech we just heard."*

Repeats are a natural coping method for individuals who experience a form of cognitive lapse. As such, it is arguable that individuals with Alzheimer's whose cognitive functionality has suffered degradation would be prone to use repeats more often than healthy individuals.

3) *Incomplete Words*

Incomplete words are instants within the corpus where the speaker begins to pronounce a word but then inexplicably stops before completing the pronunciation, indicated within the CCC corpus by the use of a dash or tilde at the end of a word as in the following example:

> *"Until my Mom got home and uh, as gro~ as I was growing up, it was, I had a great mother."*

The cause of this behavior is creditable as a lapse in cognitive ability where the speaker intercepts their thought process before saying something wrong, or it can serve as an indicator that the speaker is unsure of what they are trying to communicate and stop altogether.

4) *Syllables per Minute*

The rate of speech is provided within the CCC corpus using the metric of syllables-per-minute for both the Interviewers and the Subjects with Alzheimer's disease. This form of measurement is indicated to be an effective means of comparing the speech capacity of individuals [10].

E. *Pragmatic Features*

Part of our study is to investigate dialogue moves that may facilitate conversation with persons with dementia of the Alzheimer type. Pope and Davis [1] have suggested that the use of short (often two-syllable) go-ahead phrases will increase the fluency of dialogs with persons with dementia.

1) *Go-ahead utterances*

Go-ahead utterances are instances in dialogue in which a speaker provides responses that do not add anything in a conversation beyond a minimal response. Go-ahead utterances usually serve as an indicator by one of the speakers that they either have nothing to input within a conversation or wish for another speaker to continue. Go-aheads also serve as means of validation that a person is listening to what someone is saying, or that they agree or disagree with what is being said as in the following dialog example:

Ms. April:	*It was a nonprofit organization.*
Ms. March:	*Uh huh.*
Ms. April:	*And they didn't provide any of that.*
Ms. March:	*Awww.*
Ms. April:	*So, all my health coverage is still my …*

V. RESULTS AND ANALYSIS

A. *Differences between subjects and interviewers*

In a prior study using the CCC, we calculated the average for each subject with Alzheimer's and for each conversational partner (for shorthand, we will call the person *without* Alzheimer's the "Interviewer" even though there is no formal interview being conducted) [14, 15]. The conversational partner was most often a graduate nursing student. In that study, there were 80 total transcripts with 31 different individuals with Alzheimer's disease and 57 different interviewers.

In Table 1, we present a number of metrics comparing speakers without Alzheimer's disease (Interviewers) and speakers with dementia of the Alzheimer's type (Subjects) along with the p value associated with a T-test comparing the two groups. The part-of-speech metrics are computed per word (in other words, 40.7% of the words spoken by the Interviewers are nouns). Incomplete words, filler phrases, and repeats are also calculated per word. The percentage of go-ahead utterances is averaged per utterance.

Table 1 Metrics contrasting caregivers (Interviewers) with subjects with Alzheimer's disease

Metric	Interviewers (Mean & SD)	Subject w/ Alz. (Mean & SD)	T-Test (p value)
Noun Rate	0.407 (0.054)	0.397 (0.050)	0.1978
Verb Rate	0.080 (0.017)	0.083 (0.014)	0.2010
Adjective Rate	0.157 (0.013)	0.020 (0.019)	0.1434
Pronoun Rate	0.162 (0.030)	0.179 (0.025)	0.1166
Type-Token Ratio	0.414 (0.140)	0.406 (0.114)	0.3865
Brunet's Index	13.23 (1.896)	13.21 (1.823)	0.4740
Honore's Statistic	615.1 (93.7)	623.2 (93.9)	0.3508
% of Go-Ahead Utt.	**34.88 (15.4)**	**41.18 (16.3)**	**0.0384**
Repeats	**0.0035 (0.004)**	**0.0068 (0.006)**	**0.0014**
Incomplete Words	**0.0067 (0.007)**	**0.0114 (0.010)**	**0.0054**
Filler phrases	**0.012 (0.017)**	**0.026 (0.025)**	**0.0702**
Syllables per minutes	**216.1 (42.1)**	**199.0 (70.6)**	**0.0801**

.

As can be seen from Table 1, part-of-speech was not a good metric for distinguishing between speakers with and without dementia of the Alzheimer's type. This finding contrasts with Singh's earlier work suggesting that persons with Alzheimer's use more pronouns [4]. Our analysis does indicate a slightly higher use of pronouns; however, the p-value of the T-test makes this statistic inconclusive.

Also, unlike Singh, the semantic complexity of the two speakers was not statistically different as measured by the three lexical richness measures: type-token ratio, Brunet's index, and Honore's statistic. However, when we compared the lexical richness of the dialogs within the CCC to dialogs within the Switchboard corpus, there was a significant difference in Honore's statistic: (615.1 for the CCC corpus; 652.3 for the Switchboard corpus, T-test p-value = 0.0053). We conjecture that the interviewer in the CCC corpus may be trying roughly to match the lexical richness of a person with Alzheimer's; however, we would need a collection of dialogs with the interviewer and persons not afflicted with Alzheimer's disease in order to determine if this may be the case. The CCC corpus does not support this analysis.

The most statistically significant differences between the two groups of speakers are related to the use of go-ahead phrases and several fluency measures. The Interviewer in the CCC corpus uses significantly fewer go-ahead phrases than the subject with Alzheimer's. The person with Alzheimer's also exhibits more overt signs of difficulty in fluency with higher rates of incomplete words, filler words, repeated words, and slower rate of speech.

B. *Differences between managed-care residents diagnosed with Alzheimer's and those residents that have not been diagnosed with Alzheimer's*

Because of the sometimes vast differences between the interviewers and the subjects with Alzheimer's, comparing those two groups may be misleading. Interviewers were often graduate nursing students and were characteristically much younger, healthier, and had a higher level of education than the subjects. To follow up on our earlier study, we now present results contrasting residents within the managed-care facilities based on whether the resident had or had not been diagnosed with Alzheimer's disease. For this study, we compared the statistics from 28 residents with Alzheimer's disease and 28 residents not diagnosed with AD, and the results are summarized in Table 2.

Table 2 Metrics contrasting managed-care residents without Alzheimer's disease with residents with Alzheimer's disease

Metric	**No** Diagnosis of Alzheimer's (Mean & SD)	Diagnosis of Alzheimer's (Mean & SD)	T-Test (p value)
Noun Rate	**0.368 (0.043)**	**0.386 (0.037)**	**0.0426**
Verb Rate	**0.076 (0.014)**	**0.083 (0.012)**	**0.0389**
Adjective Rate	**0.014 (0.004)**	**0.017 (0.007)**	**0.0520**
Pronoun Rate	0.171 (0.024)	0.171 (0.024)	0.4581
Type-Token Ratio	**0.222 (0.096)**	**0.301 (0.102)**	**0.0021**
Brunet's Index	**15.82 (1.63)**	**14.70 (1.477)**	**0.0043**
Honore's Statistic	**699.3 (73.1)**	**661.4 (94.0)**	**0.0477**
% of Go-Ahead Utt.	**16.65 (12.1)**	**35.7 (14.0)**	**< 0.0001**
Pauses	**0.0099 (0.023)**	**0.0354 (0.074)**	**0.0508**
Repeats	0.0158 (0.018)	0.0165 (0.010)	0.3694
Incomplete Words	**0.0039 (0.004)**	**0.0148 (0.015)**	**0.0002**
Filler phrases	0.024 (0.016)	0.034 (0.038)	0.1147
Syllables per minutes	182.5 (45.4)	179.6 (32.3)	0.3848

1) Discussion of lexical diversity

As can be seen in Table 2, there are some contrasting results from Table 1. First, there were statistically significant differences detected in syntax and lexical richness; however, these results are somewhat contradictory. When looking at noun, verb, and adjective rates as well Type Token Ratio, it appears that the subjects with Alzheimer's disease are using *more diverse* language than those subjects without the disease. However, this is likely accounted for by the fact that conversations with the Alzheimer's subjects contained almost 50% fewer words in total compared to the other group. Statistics that are sensitive to length (like Type Token Ratio) may not be quite as accurate in these situations. Notice that both Brunet's Index and Honore's Statistic account for the size of the data in a better way; and, in fact, note that the subjects without AD have statistically better scores on both measures. Past research has suggested that the lexical diversity of persons with AD is diminished [4, 11].

2) Discussion of Disfluencies

Repeated words and filler phrases occur at roughly the same rates for both groups with no statistically significant differences. Note, too, that the number of syllables per minute is not statistically different for the two groups. (There was a difference when comparing Interviewers with subjects.) However, the number of pauses per word for subjects with AD is significantly higher. The most significant disfluency difference between the two groups is the presence of incomplete words. Subjects with AD had a much higher incidence of words that were truncated in mid-speech.

3) Discussion of Go-Aheads

The most prominent statistical difference between the two groups was the use of short Go-Ahead phrases by subjects with Alzheimer's disease. What is particular striking is the relative *lack* of Go-Aheads of the subjects without AD. What this difference implies is that conversational participants without AD are contributing more to the conversation per turn. In the earlier study looking at Interviewers (Table 1), the interviewers also used relatively high rates of Go-Aheads. This high rate can perhaps be accounted for by the task the interviewers were trying to carry out: the elicitation of stories and information from the subjects. The interviewers were trying to get the subjects to talk about themselves; the use of Go-Aheads encourages that behavior [1, 7].

VI. Learning Algorithms for Classifying Alzheimer's versus non-Alzheimer's Speech

Using the features that were observed to be statistically significant, we applied two learning algorithms: a Bayesian classifier and a Decision Tree classifier, to see whether it was possible to distinguish between transcripts of residents with and without Alzheimer's disease based on these features. Training and testing were performed using leave-one-out testing (i.e., training on all data but one, testing on that one piece of data, and rotating in this manner for all data points). For purposes of both algorithms, data was discretized to values of Very Low, Low, Medium, High, and Very High, by computing which quintile a particular data point was in for each range of data. Table 3 presents these results.

Table 3 Precision and recall results for classification of Alzheimer's versus non-Alzheimer's transcripts using a decision tree classifier and a Naïve Bayes classifier

Learning Algorithm (28 Alz; 28 non-Alz)	Alzheimer's		Non-Alzheimer's	
	Prec.	Recall	Prec.	Recall
Decision Tree	66.7	66.7	67.9	67.9
Naïve Bayes	80.8	0.75	79.3	82.1

Both classification algorithms perform significantly better than random. The Naïve Bayes classifier indicates that the most informative features were: Pauses, Go-Aheads, Fillers, and Incomplete words.

VII. Conclusions and Future Work

We have shown that a number of features in the speech of persons with dementia of the Alzheimer's type who are in managed-care facilities are statistically significantly different from the speech of other individuals in managed-care. Using these features, we have also employed classification algorithms to classify the speech of an individual as either indicating the symptoms of Alzheimer's disease or not. The Bayesian classifier had a precision approaching 80% precision and recall.

Our next main goal is to determine what dialogue behaviors enhance the fluency of individuals with Alzheimer's disease. Does the use of go-aheads and paraphrasing by conversational participants increase the fluency and general conversational ability of a person with Alzheimer's disease? This future study will involve looking at more than just syntactic measures as it will require looking at the relative "success" of each dialogue turn.

REFERENCES

[1] C. Pope and B. Davis, “Finding a balance: The Carolinas Conversation Collection”, *Corpus Linguistics and Linguistic Theory* 7(1), 143—161, 2011.

[2] Alzheimer’s Association. *2014 Alzheimer’s Disease Facts and Figures*. Downloaded on 07/05/14 from www.alz.org, 2014.

[3] M. Fried-Oken, C. Rowland, G. Baker, M. Dixon, C. Mills, D. Schultz, D., and B. Oken. “The Effect of Voice Output on AAC-Supported Conversations of Persons with Alzheimer’s Disease”. *ACM Transactions on Accessible Computing*, Vol. 1, No. 3, Article 15, 2009.

[4] R. Bucks, S. Singh, J. Cuerden, and G. Wilcock, “Analysis of spontaneous, conversational speech in dementia of Alzheimer type: Evaluation of an objective technique for analysing lexical performance”, *Aphasiology* 14(1), 71-91, 2000.

[5] I. Thompson. Language in dementia. *International Journal of Geriatric Psychiatry*, 2, 145-161, 1987.

[6] S. Ray, M. Britschgi, C. Herbert, Y. Takeda-Uchimura, A. Boxer, K. Blennow, L. Friedman, D. Galasko, M. Jutel, A. Karydas, J. Kaye, J. Leszek, B. Miller, L. Minthon, J. Quinn, G. Rabinovici, W. Robinson, M. Sabbagh, Y. So, D. Sparks, M. Tabaton, J. Tinklenberg, J. Yesavage, R. Tibshirani, R. and T Wyss-Coray, “Classification and prediction of clinical Alzheimer's diagnosis based on plasma signaling proteins”, *Nat Med* 13(11), 1359—1362, 2007.

[7] B. Davis, B. and M. Maclagan,”Examining pauses in Alzheimer's discourse”, *American journal of Alzheimer's Disease and other dementias* 24(2), 141—154, 2009.

[8] M. Snover, B. Dorr, B. and R. Schwartz,, “A lexically-driven algorithm for disfluency detection”, in *Proceedings of HLT-NAACL 2004: Short Papers*, pp. 157—160, 2004.

[9] H. Bortfeld, S. Leon, S.; J. Bloom, M. Schober, and S. Brennan, “Disfluency rates in conversation: Effects of age, relationship, topic, role, and gender”, *Language and Speech* 44(2), 123-147, 2001.

[10] J. Yaruss, “Real-time analysis of speech fluency: Procedures and reliability training”, *American Journal of Speech-Language Pathology* 7(2), 25, 1998.

[11] S. Singh, *Neural Networks In Spontaneous Speech Assessment Of Dysphasic Patients*, Ph.D. Dissertation, University of the West of England, Bristol, UK, 1996.

[12] S. Singh, S. and T. Bookless, “Analysing spontaneous speech in dysphasic adults”, *International Journal of Applied Linguistics* 7(2), 165—181, 1997.

[13] P. Ten Have, *Doing conversation analysis*, Sage Publications Ltd., 1999.

[14] C. Guinn and A. Habash, “Language Analysis of Speakers with Dementia of the Alzheimer's Type”, *AAAI Fall Symposium: Artificial Intelligence for Gerontechnology, 2012.*

[15] A. Habash, C. Guinn, D. Kline, and L. Patterson. “Language Analysis of Speakers with Dementia of the Alzheimer's Type”. *Annals of the Master of Science in Computer Science and Information Systems at UNC Wilmington*, 6(1) paper 11, 2012.

[16] J. Godfrey, E. Holliman, and J McDaniel. "SWITCHBOARD: Telephone speech corpus for research and development." In *Acoustics, Speech, and Signal Processing*, ICASSP-92., 1992 IEEE International Conference on, vol. 1, pp. 517-520. IEEE, 1992.

[17] S. Bird, E. Klein, E. and E. Loper, E., *Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit*, O'Reilly, Beijing, 2009.