A New Approach for Automating Analysis of Responses on Verbal Fluency Tests from Subjects At-Risk for Schizophrenia

Mary Pietrowicz[1], *Carla Agurto*[1], *Raquel Norel*[1], *Elif Eyigoz*[1], *Guillermo Cecchi*[1], *Zarina R. Bilgrami*[2], *and Cheryl Corcoran*[2,3]

[1]IBM Thomas J Watson Research Center
[2]Icahn School of Medicine at Mount Sinai
[3]Bronx VA MIRECC

Mary.Pietrowicz@ibm.com, Carla.Agurto@ibm.com, rnorel@us.ibm.com, ekeyigoz@us.ibm.com, gcecchi@us.ibm.com, zarina.bilgrami@mssm.edu, cheryl.corcoran@mssm.edu

Abstract

What if young people at risk for developing schizophrenia could be identified early, via a fast, automated, non-invasive test of language, which could be administered remotely? These youths could then receive intervention which might mitigate course and possibly prevent psychosis. Timed word fluency tests, in which individuals name words starting with a designated sound (typically F/A/S) or represent a given concept category (commonly animals/fruits/vegetables), have been used in the assessment of schizophrenia and its risk states, and in many other mental health conditions. Typically, psychologists manually record the number and size of valid phoneme clusters and switches observed in the phonemic tests and count the number of valid words belonging to a given category in the categorical tests. We present a new technique for automating the analysis of category fluency data and apply it to the problem of detecting youths at risk of developing schizophrenia, with best results over 85% accuracy when applying phonemic analysis to categorical data. The technique supports the separate quantification of structural and sequential phonemic similarity measures, supports an arbitrary range of pronunciations and dialects in the analysis, and may be extended to the assessment of other mental and physical health conditions, and their risk states.

Index Terms: verbal fluency, phonemic similarity, categorical similarity, structural similarity, sequential similarity, schizophrenia

1. Introduction

Schizophrenia is a mental disorder characterized by abnormal behavior, strange speech, decreased ability to understand reality, and possibly hallucinations and delusions [1]. Schizophrenia is typically diagnosed in the late teen years through the early thirties, but more subtle changes in cognition and social behavior may precede the diagnosis, often by years [2,3]. Schizophrenia patients exhibit language impairments, such as semantic network disorganization [4,5,6] and thought disorder, where thoughts lose coherence [7]. People with this disease often jump from one subject to another based on the sounds or associations of the words they have spoken [8]. The disease is typically diagnosed by clinical observation, often via language. Automated, systemic, unbiased tools help detect more subtle changes in language, which can help predict psychosis onset [9,10]. In this study, we explore the automated analysis of language to identify young people who are at-risk of developing schizophrenia, even before they show symptoms. These youths could then receive intervention which could possibly mitigate course and prevent psychosis.

Word fluency tests are common neuropsychological assessments, used both in clinical practice and in research [11]. Individuals are asked to produce as many words as possible in a given time, usually a minute. Semantic fluency tasks [12] ask for words from a given category such as animals or fruits, while phonemic fluency tasks [13] ask for words which start with a given letter, for example, “L”. In the standard analysis, the score is the total number of unique and correct words. Variations on this analysis include the number of repetitions, as well as clustering and switching properties [14], which measure the relationships between adjacent words (e.g., number of consecutive words beginning with a given sound or sound sequence, or number of times the beginning sound or sound sequence changes). Research has shown that schizophrenia patients produce more errors and fewer words in word fluency tests than healthy controls [15], and show more impairment in the semantic (categories) word fluency test [16]. Impairment on both kinds of fluency tasks extends to the early onset phase of schizophrenia, with schizophrenia patients demonstrating deficiency in semantic fluency compared to controls, even after age and IQ correction. For this reason, impaired semantic fluency has been proposed as an early marker of schizophrenia [17]. Research suggests that cultural background does not impact the language deficit due to schizophrenia, but that onset age does; and early onset results in more severe semantic degradation than later onset [18]. Some evidence suggests that selected antipsychotic drugs improve some aspects of cognitive function, including verbal fluency [15,18]. Our work extends the exploration of the relationship between semantic word fluency test performance and schizophrenia to those who are at risk for developing the disease, but who have not yet exhibited symptoms leading to a diagnosis. Our work also extends the analysis techniques commonly used on semantic fluency data to include both semantic and phonemic information, and presents a new way of representing and assessing phonemic quality and similarity which considers both structural and sequential information. Finally, our work presents new automated techniques for the assessment of phonemic similarity in word lists, which are directly applicable to both semantic and phonemic fluency tasks. We ask the following research questions, and present the results of our exploration:

RQ1: Can semantic fluency tasks be used to identify people who are at risk for developing schizophrenia?

RQ2: Can phonemic analysis techniques applied to semantic fluency task data help improve detection of people at risk for developing schizophrenia?

2. Verbal Fluency Dataset

The dataset included word lists collected via a semantic fluency test administered to 36 (14 male) healthy young people, and 53 others (25 male) who were at-risk for developing schizophrenia. The prompt for the test was, "Tell me the names of as many animals as you can. Name them as quickly as possible. Any animals will do; they can be from the farm, the jungle, the ocean, or house pets. For instance, you could begin with dog… Ready? Begin." The subjects were all English speakers, who acquired their language from more than 12 countries and from every populated continent in the world. The data collection sites included multiple cities in the Northeastern United States; Toronto, Canada; and Melbourne, Australia. The vocabulary from the speakers in Melbourne was notably different from the North American speakers (e.g., "komodo dragon" and "wallaby" vs "cat" and "squirrel"). To account for these differences, we first modeled the North American data separately, and then ran separate analysis using all the data for comparison. The North American subset included 15 healthy people, and 37 at-risk for developing schizophrenia, for a total of 52 people. Both the control and at-risk populations were young, ranging in age from 14-31, which represents the time in which schizophrenia symptoms typically appear. 35 people in the at-risk population had other mental health diagnoses (e.g., anxiety, cannabis use disorder, PTSD, panic disorder, social phobia, anorexia, and others), and about half of the at-risk population took medications. Only one person in the control group took medication, and none in the control group had mental health diagnoses.

The at-risk group was diagnosed for being at-risk based on the presence of attenuated psychotic symptoms (APS) using the "Structured Interview for Psychosis-Risk Syndromes/Scale of Psychosis-Risk Symptoms" (SIPS/SOPS) [19]. This is the gold standard measure for diagnosing APS, and has excellent psychometric properties. Research staff are trained and certified in the administration of this test, and consensus ratings are achieved with expert input. All at-risk individuals participate in a prospective cohort study to determine outcome, and this includes schizophrenia onset, assessed using SIPS/SOPS. For this and other at-risk APS cohorts, ~20% develop schizophrenia within 1-2 years.

One of the authors, Cheryl Corcoran MD, is an expert in schizophrenia risk and has published in the ethics of this research. All of the participants in this study were help-seeking with clinical symptoms, such that automated methods have only been applied thus far in order to identify language markers that may enhance prediction beyond the APS categorization. Analysis are of de-identified data. These are preliminary proof-of-principle studies and would need to be done in respect to biomarker development before they could be used for general screening in clinical contexts.

3. Description of Features

Both phonemic and semantic features were extracted from the semantic fluency test data and are described below.

3.1. Phonemic Information Features

Word lists contain both phonemic word structure and word sequence information. The structural information includes the amount, variation, and overall quality of phonemic similarity, diversity, and complexity across words, *independent of the word order*. We mapped the phonemic information in each word onto a graph to capture phonemic structure such that the top-level nodes in the graph contain the collections of words which start with the same sound (1st-degree phonemic similarity), and the second level nodes in the graph contain the collections of words which begin with the same 2 sounds (2nd-degree phonemic similarity), and so on (see Figure 1). The graph could be arbitrarily deep, but in our observation, most of the phonemic information is contained in the top two or three levels of the graph (1-3 consecutive sounds) for semantic fluency tests. In contrast, when this technique is applied to phonemic fluency tests (in which subjects are asked to name words which start with a given sound), significant information is present at much deeper levels, easily to a depth of 6. As an example, statistical analysis of the phonemic structure of the graph in Figure 1 reveals 10 words with three 1st-degree similarity nodes having an average of 3.33 words (σ=2.05) per node. The words at this level vary in length with a mean of 4.00 sounds per word (σ=0.89). The second level in the tree has 5 words with two 2nd-degree similarity nodes, with a mean of 2.50 words per node (σ=0.50). The words at this level also have a mean of 4.00 sounds per word (σ=0.63).

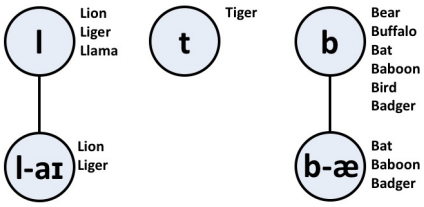

Figure 1: Phonemic Similarity Graph Example. The following list of animal names are projected onto a similarity graph: "lion, liger, tiger, llama, bear, bat, baboon, badger, bird, buffalo." The top level nodes contain words which start with the same sound, and the nodes at the bottom of the graph contain words which start with the same two sounds.

The word lists are projected onto similarity graphs using information from a phonemic dictionary, in this case the Penn Forced Aligner [20], extended to accommodate the animal naming test. Many words have multiple pronunciations, so to resolve this variability, we first build the graphs with all possible pronunciations from the dictionary. Then, we prune the tree with an algorithm that prefers pronunciations which produce the deepest trees first, and prefers nodes with the largest numbers of words in them second. It will prune a childless node with a single word in it in favor of an alternate pronunciation belonging to a node with many words and child nodes. This technique favors phonemic similarity, and tends to select the more common (in practice) pronunciation. Alternate algorithms could prune based on regional accent, given this information. We represent phonemic structure in each word list via a vector of 37 features, containing the maximum graph depth, the number of nodes, the number of nodes at each level (0-6), the per-level number of words per

node (μ and σ), and the per-level number of sounds per word (μ and σ). The graph structure enables efficient data representation and processing of all phonemic features, structural and sequential.

Phonemic word sequence information measures the sonic relationships between adjacent words, and therefore depends on word ordering. Given the sequence in Figure 1, we can examine the first sound in each word, and discover 3 different first sounds across 10 words, and 3 switches from one first sound to another in the sequence. Also, we see that the last 6 words all start with the sound "b." This consecutive string of words starting with the same sound is called a cluster [14]. We can look for clusters and switches over arbitrary sequences of sounds at arbitrary depths in the tree, and measure the distance between clusters. In this list, "bat," "baboon," and "badger" form a cluster of 3 sequential words which share the same first 2 sounds in the word. We represent word sequence information in each word list via a vector of 30 features containing the 1) number of switches across the first 1-6 consecutive sounds, 2) the cluster sizes, or number of words in each cluster of 1-6 consecutive sounds (μ and σ), and 3) the size of the words in each cluster (μ and σ). In addition to structural and sequential phonemic information, we include the total number of animal names given, the number of unique animal names, and the ratio of valid animal names to total animal names given.

3.2. Semantic Information Features

We used the Global Word Vector representation (GloVe) [21] to identify the associations of the given animal names with the following semantic categories: *mammals, birds, reptiles, amphibians, fish, insects, crustaceans, arachnids, echinoderms, worms, mollusks, and sponges*. This model is an unsupervised algorithm trained with the Wikipedia 2014 and Gigaword 5 corpora such that each of the 6-billion words provided by GloVe is represented via 300 dimensions. We selected the types of animals as semantic categories to learn whether specific content could help characterize the subjects at high risk for schizophrenia. To assess the similarity of the given words to these categories, the cosine distance between each animal word listed by the participant and the categories of interest was calculated. We computed 6 statistical descriptors from the distribution of values obtained, including median, interquartile range from 25th to 75th percentiles (IQR), 10th (pct10) and 90th (pct90) percentiles, and 3rd and 4th moments (skewness and kurtosis).

4. Experiments and Results

This section presents an analysis of candidate features, describes the classification experiments, and presents the classification results.

4.1. Analysis of Features

To gain insight into the ability of the features to distinguish young people at risk of developing schizophrenia from healthy young people (controls), we ran t-tests for all features across the entire dataset, and again for just the North American portion of the dataset. Table 1 shows the top 3 semantic features, top 3 phonemic features using the complete names all animals (e.g., "komodo dragon"), and top 3 phonemic features using only the first word of the given animal names (e.g., "komodo").

Table 1: Statistical significance (p-values) for classification between healthy controls and high-risk subjects (best features). "L0" stands for Level 0 in the graph, and "L3" stands for Level 3 in the graph.

Group	Type of feature	Relevant Feature	Pval
North America (52)	**Phonemic 1 word**	L0: mean cluster size	.01
		L0: mean word size per cluster	.01
		L3: mean number of words	.03
	Phonemic all words	L0: mean cluster size	.01
		L0: mean word size per cluster	.01
		L3: σ word size	.03
	Semantic	Amphibians (pct90)	.02
		Crustaceans (IQR)	.02
		Fishes (IQR)	.03
All Subjects (89)	**Phonemic 1 word**	L0: mean number of words	.03
		L0: σ number of unique words	.05
		L0: σ cluster size	.05
	Phonemic all words	L0: mean number of words	.02
		L0: σ number of unique words	.04
		L0: σ cluster size	.05
	Semantic	Fishes (IQR)	.0009
		Mammals (median)	.002
		Crustaceans (IQR)	.002

Results suggest that both semantic and phonemic features may be useful in identifying the at-risk subjects. For the phonemic tests, the analysis suggests that using just the first word of an animal name may identify at-risk subjects nearly as well as the full name. Results also suggest that phonemic differences exist between the North American subjects and the Australian subjects, and that possibly each linguistic group should be examined separately. For the North American subjects, the mean cluster size, considering just the first sound, is an important factor (L0: mean cluster size); this finding supports a prior finding that schizophrenia patients often jump from one subject to another based on the sounds or associations of the words they have spoken [8]. These results suggest a new finding, that the complexity or size of the words in the clusters are also important in distinguishing at-risk subjects (L0: mean word size per cluster). Finally, for the North American subjects, phonemic similarity at sequences of 3 or more sounds is important to at-risk subject identification, both in the number and complexity of the words themselves (L3: mean number of words, L3: σ word size). When considering all subjects, the best phonemic features are slightly different, but similar. In this group, the variation in cluster sizes, considering the first sound, is important (L0: σ cluster size). Also, the average number of words starting with the same sound is important here, regardless of the ordering (L0: mean number of words), and also the variation in the number of unique words which start with each letter (L0: σ cluster size). Phonemic similarity and quality appear to be important markers in the semantic fluency task.

For the semantic features, water animals (fish and crustaceans) seem to be important markers in distinguishing at-risk populations in both population groups. The categorical difference between the groups for Mammals and Amphibians, however, suggests again that linguistically similar groups should be analyzed individually for best results. Some of the features still have considerable overlap between conditions, however, even with $p < 0.001$; Figure 2 shows an example of this, using the best feature, as ranked by p-value.

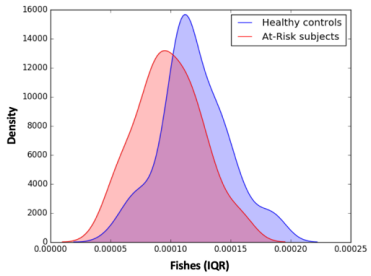

Figure 2: Distribution density overlap between conditions for the semantic "Fishes" feature.

4.2. Classification Experiments

To explore the ability of phonemic features to classify at-risk population vs. controls, we first addressed possible vocabulary differences between the Australian and North American speakers by testing classification on the North American group only, and then on the entire dataset. Within these two subgroups, we explored the effects of using only structural phonemic features vs. all phonemic features. Finally, we explored the effects of approximating the animal name using the first word only in a multi-word name, vs. using all of the words in a multi-word animal name.

Table 2: Classifier Results Using Only Phonemic Features with 5-Fold Cross Validation. The "1-word"cases use the first word only in an animal name as an approximation; "All words" use the complete animal name. "S"cases use only the structural phonemic features; "S+T" cases use both the structural and sequential (time) features. Classifiers include: SVM(Support Vector Machine), NN(Nearest Neighbor), RF(Random Forest).

Group	# of Words	Feature Set	Chance	SVM | NN | RF
North America (52)	1 word	S	0.50	0.76 | 0.76 | **0.83**
		S+T	0.50	0.80 | **0.85** | 0.80
	All words	S	0.50	0.77 | 0.79 | **0.85**
		S+T	0.50	0.80 | **0.83** | **0.83**
All subjects (89)	1 word	S	0.50	0.68 | **0.71** | 0.70
		S+T	0.50	0.66 | **0.78** | 0.75
	All words	S	0.50	0.59 | 0.70 | 0.70
		S+T	0.50	0.60 | **0.77** | 0.69

We also eliminated from consideration the features which did not vary. Since the subjects were naming animals and were not trying to name words which began with the same sound, virtually all of the phonemic similarity information occurred in the first 4 sounds of the words. This reduced the feature vector to 22 structural phonemic features, 20 sequential phonemic features, and 3 general features. We ran 3 kinds of classifiers, including linear SVM, nearest neighbor, and random forest, using 5-fold cross validation, feature selection by p-value within fold, and random upsampling of the minority class to balance conditions. The results (see Table 2) show that 1) modeling linguistically similar groups of people separately (e.g., North American vs. all speakers) yields better results (up to 20% better), 2) including both structural and sequential phonemic features gives better results than using structural features alone (up to 9% better), 3) using the complete word name yields better results than using just the 1st word in the name (up to 9% better), and 4) Nearest Neighbor and Random Forest classifiers work best on this non-linear data.

To explore the ability of semantic analysis to identify the at-risk population, we applied Lagrangian Support Vector Machine (LSVM), Nearest Neighbor, and Random Forest classifiers to the GloVe features (animal name categories) described in the previous section, using feature selection by p-value within fold, and 5-fold cross validation. The classifiers identified the at-risk subjects at rates up to 11% above chance (see Table 3), using only the semantic concepts relating to animal groups.

Table 3: Classifier Results Using Only Semantic Features with 5-Fold Cross Validation. Classifiers include: LSVM (Lagrangian Support Vector Machine), Nearest Neighbor (NN), and Random Forest(RF)

Group	Chance	LSVM | NN | RF
North American (52)	0.71	0.58 | **0.77** | 0.75
All Subjects (89)	0.60	0.65 | **0.71** | 0.65

5. Discussion and Conclusions

We explored two research questions in this work, the first question (RQ1) examining whether semantic fluency tasks could be used to identify people who are at risk for developing schizophrenia, but who have not yet been diagnosed with the disease. The results suggest that these tasks can identify the at-risk population. The second question (RQ2) asked specifically whether phonemic analysis techniques applied to a semantic fluency task would be effective in identifying the at-risk population; and the results suggest that phonemic analytics are very good at discerning the at-risk population, and may be even better than some semantic analysis. The scale of semantic analysis could be greatly expanded, however, because we only explored semantic animal categories, since the data we were given involved naming animals. Expanded analysis of semantic features in the future could reveal more of the story. Future work can dig deeper to explore larger sample sizes, and a variety of populations, to see whether larger sample sizes and large-scale modeling techniques could improve detection across heterogeneous populations. Improved techniques for balancing sample sizes across conditions could also be explored. Future work could also incorporate auto-generated accent and dialect representations in the phonemic dictionary, to match the subject's accent or dialect of origin, if that were known. In general, the scale of the existing phonemic dictionary could be expanded via models which convert standard dictionary representations to phonemic representations, and output a range of pronunciations for various regional accents. Then, spoken word fluency tests (audio and video recordings) would enable analysis of acoustic, linguistic, and expressive gesture alongside text. Given continued longitudinal data collection, future studies could analyze the rates at which an at-risk population is diagnosed with schizophrenia, and the linguistic elements most likely to predict the probability of this diagnosis.

6. References

[1] World Health Organization "Schizophrenia," Available at: "http://www.who.int/mental_health/management/schizophrenia/en," last accessed 4/4/2018.

[2] J. McGrath, S. Saha, D. Chant, and J. Welham, "Schizophrenia: a concise overview of incidence, prevalence, and mortality," *Epidemiol. Rev*., vol. 30, no. 1, pp. 67-76, 2008.

[3] C. Hollis and L. Palaniyappan, "Schizophrenia and psychosis," in *Rutter's Child and Adolescent Psychiatry*, Wiley-Blackwell, 2015, pp. 774-792.

[4] M.S. Aloia, M.L. Gourovitch, D.R. Weinberger, and T.E. Goldberg, "A investigation of semantic space in patients with schizophrenia," *J. Int. Neuropsychol. Soc*, vol. 2, no. 4, pp. 267-273, 1996.

[5] J. S. Paulsen, R. Romero, A. Chan, A.V. Davis, R.K. Heaton, and D.V. Jeste, "Impairment of the semantic network in schizophrenia," *Psychiatry Res.*, vol. 63, no. 2-3, pp. 109-121, 1996.

[6] A. Pawelczyk, M. Kotlicka-Antczak, E. Lojek, A. Ruszpel, and T. Pawelczyk, "Schizophrenia patients have higher-order language and extralinguistic impairments, " *Schizophr. Res.,* vol. 192, pp. 274-280, 2018.

[7] C.M. Adler *et al.*, "Comparison of ketamine-induced thought disorder in healthy volunteers and thought disorder in schizophrenia," *Am. J. Psychiatry*, vol. 156, no. 10, pp. 1646-1649, 1999.

[8] Michael A. Covington, Congzhou He, Cati Brown, Lorina Naci, Johathan T. McClain, Bess Sirmon Fjordbak, James Semple, and John Brown, "Schizophrenia and the structure of language: The linguist's view," Schizophrenia Research, vol. 77, pp. 85-98, 2005.

[9] G. Bedi *et al.*, "Automated analysis of free speech predicts psychosis onset in high-risk youths," *NPJ Schizophr*, vol. 1, p. 15039, 2015.

[10] C.M. Corcoran *et al.*, "Prediction of psychosis across protocols and risk cohorts using automated language analysis," *World Psychiatry*, vol. 17, no. 1, pp. 67-75, 2018.

[11] Z. Shao, E. Janse, K. Visser, and A.S. Meyer, "What do verbal fluency tasks measure? Predictors of verbal fluency performance in older adults," *Front. Psychol.*, 2014.

[12] A. L. Benton, "Differential behavioral effects in frontal lobe disease," Neuropsychologia, vol. 6, no. 1, pp.53-60, 1968.

[13] F. Newcombe, "Missile wounds of the brain: A study of psychological deficits," Oxford University Press, 1969.

[14] A.K. Troyer, M. Moscovitch, and G. Winocur, "Clustering and switching as two components of verbal fluency: Evidence from younger and older healthy adults," *Neuropsychology*, vol. 11, no. 1., American Psychological Association, US, pp. 138-146, 1997.

[15] F. Galaverna, A.M. Bueno, C.A. Morra, M. Roca, and T. Torralva, "Analysis of errors in verbal fluency tasks in patients with chronic schizophrenia," *Eur. J. Psychiatry*, vol. 30, no. 4, pp. 305-320, 2016.

[16] W.S. Kremen, L.J. Seidmman, S.V. Faraone, and M.T. Tsuang, "Is there disproportionate impairment in semantic or phonemic fluency in schizophrenia?," *J. Int. Neuropsychol. Soc*., vol. 9, no. 1, pp. 79-88, 2003.

[17] T.J. Phillips, A.C.D. James, T.J. Crow, and S.L. Collinson, "Semantic fluency is impaired but phonemic and design fluency are preserved in early-onset schizophrenia," *Schizophr. Res*., vol. 70, no. 2-3, pp. 215-222, 2004.

[18] C. Sumiyoshi, M. Matsui, T. Sumiyoshi, I. Yamashita, S. Sumiyoshi, and M. Kurachi, "Semantic structure in schizophrenia as assessed by the category fluency test: effect of verbal intelligence and age of onset.," *Psychiatry Res.*, vol 105, no. 3, pp. 187-199, Dec. 2001.

[19] T.J. Miller, T.H. McGlashan, S.W. Woods, K. Stein, N. Driesen, C.M. Corcoran, R. Hoffman, and L. Davidson, *Psychiatr Q.*, 70(4):273-87, Winter 1999.

[20] The Penn Phonetics Lab Forced Aligner, available at "https://babel.ling.upenn.edu/phonetics/old_website_2015/p2fa/index.html," last accessed 4/4/2018.

[21] J. Pennington, R. Socher, and C. Manning, "Glove: Global Vectors for Word Representation," in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 1532-1543, 2014.