Chapter 11
Automated Method for Diagnosing Speech and Language Dysfunction in Schizophrenia

Anne Debra Tilaka, Joachim Diederich, Insu Song and Ai Ni Teoh

11.1 Introduction

The year 2011 marked one hundred years since the introduction of the diagnostic category schizophrenia and for a century, speech and language dysfunction (SLD) has been strongly associated with schizophrenia. At a descriptive clinical level, thought disorder in schizophrenics can be divided into positive and negative symptoms [25]. Positive thought disorder includes, among other things, the use of neologisms (new unusual words), incoherence, derailment (the speaker is losing track) and glossomania (association chaining, which may also occur in mania). Negative thought disorder means poverty of speech (e.g., reduced vocabulary) and correlates with other, non-linguistic negative symptoms (e.g., flat affect). SLD occurs in other mental health disorders as well, for instance, in mania and depression [17]. This study, however, focuses on schizophrenia, since incoherent speech and language is among the main symptoms of the disorder [17, 25].

A. D. Tilaka (✉)
School of Arts and Social Sciences, James Cook University, Singapore 574421, Singapore
e-mail: debradass@yahoo.com.sg

J. Diederich
School of Information Technology and Electrical Engineering, The University of Queensland, Brisbane, QLD 4072, Australia
e-mail: j.diederich@uq.edu.au

I. Song
School of Business/IT, James Cook University, Singapore Campus, Singapore 574421, Singapore
e-mail: insu.song@jcu.edu.au

A. N. Teoh
North Dakota State University, Fargo, ND 58102, USA

M. Lech et al. (eds.), *Mental Health Informatics*,
Studies in Computational Intelligence 491, DOI: 10.1007/978-3-642-38550-6_11,

11.1.1 The Importance of Speech and Language Disorders for the Assessment of Schizophrenia

There are speech and language differences between individuals affected by schizophrenia and non-psychotic subjects. Firstly, several studies demonstrated that schizophrenic patients communicate less compared to normal subjects [3, 15, 16, 23, 26]. Qualitatively, schizophrenic SLD includes two features which are commonly referred to as 'poverty of speech' (lack of speech or reduced vocabulary) and 'poverty of content' (lack of variety) [3, 11]. Schizophrenic SLD persists longitudinally, occurring in about 80–90 % of patients during the acute stages of hospitalization [22, 33, 37], and prevailing in about 39 % of the patients during the post-acute stages [3, 6, 22, 38].

Since SLD is associated with cognitive impairments [13, 18], it is regarded as a reliable clinical marker of a psychotic disorder, as it can be directly observed [10, 37]. SLD is also considered as an early vulnerability indicator of schizophrenia, since its presence in children and adolescents is a significant predictor of adult psychosis [4, 27, 31, 32]. Unlike positive SLD (e.g., derailment, incoherence), it is the negative type (poverty of speech and content) that is indicative of a poor prognosis [3, 5, 21, 24, 28]. Reliable and valid measures of SLD are critical for research and clinical practice.

11.1.2 Speech and Language Disorder Measures

11.1.2.1 Thought, Language and Communication Scale

Common measures of SLD are observer-rated scales, as they provide for a broad assessment of symptoms. The most widely used instrument is the Thought, Language and Communication scale (TLC) [2, 21, 28, 36] which is based on a phenomenological approach [28]. It consists of 18 items: poverty of speech, poverty of content, pressure of speech, distractibility, tangentiality, loss of goal, derailment, circumstantiality, illogicality, incoherence, neologisms, word-approximations, stilted-speech, clanging, preservation, echolalia, blocking and self-reference.

TLC was validated by demonstrating that schizophrenics obtained greater SLD an normal participants [1, 2]. Furthermore, compared with other clinical groups (mania and depression), schizophrenic patients consistently scored high over extended periods of time [1, 2]. Other TLC studies have produced similar results [3, 5, 11, 36, 37]. Among the 18 TLC items, schizophrenics normally obtain high scores on 'poverty of speech' and 'poverty of content'. These findings have been replicated numerous times [3, 5, 11, 33, 36, 37].

TLC has also been used for diagnostic purposes. The scale achieved an overall accuracy of 84 % in correctly determining schizophrenia and mania with a class validation rate of 80 % (83 % in a replication study) [3]. Although non-psychotic participants were included, the study did not provide classification rates for this group [1]. Cuesta and Peralta found that the TLC scale was able to differentiate between patients affected by schizophrenia or mania with an accuracy of approximately 80 % [11].

Many of the diagnostic studies that used TLC focused on comparing schizophrenia with mania and did not necessarily include non-clinical participants. An exception was the study by Berenbaum et al. [5] who refined the TLC scale. When the authors compared schizophrenics with non-psychotic participants, accuracy rates were a mere 64 % indicating that the original TLC was performing better. Overall, the vast majority of studies demonstrated that TLC has good discriminative and predictive properties.

11.1.2.2 Clinical Language Disorder Ratings Scale

The Clinical Language Disorders Rating Scale (CLANG) utilises a linguistic approach and is structured according to linguistic levels: syntax, semantics and production. The majority of the 17 items are described in linguistic terms (excess phonetic-association, abnormal syntactic-structure, excessive syntactic-constraints, lacking semantic-association, referential failures, discourse failures, excessive details, lacking details, prosodic speech, abnormal prosody, pragmatic disorder, dysfluency, dysarthria, and paraphasic error), except for three that include phenomenological terms (poverty of speech, pressure of speech, and neologisms). Generally, speech refers to motor production and language refers to content. For CLANG, transcripts of audio-recorded speech are used, and in this sense, this study is both on speech as well as on language.

CLANG was first validated with a sample of 204 Hong Kong Chinese schizophrenic patients [10]. Another study involved British participants. In the study, schizophrenic patients obtained higher SLD scores compared to normal participants [9]. In comparison to other clinical groups (mania and depression), schizophrenic subjects obtained the highest SLD scores [8]. Qualitatively, over half of CLANG's items can be linked with TLC items, which contributes to its convergent validity [10]. Furthermore, CLANG's production factor (poverty of speech, lack of details and aprosody) correlates with TLC's poverty of speech and poverty of content. Schizophrenics scored highest on the CLANG item 'lack of details', which parallels TLC's 'poverty of content' [8].

When CLANG was used for prediction (in this case differentiating schizophrenia from depression), it achieved an overall accuracy of 76 % [8]. In another study, CLANG achieved a perfect accuracy of 100 % in correctly identifying non-clinical participants, and 70 % accuracy in identifying schizophrenic participants in a mixed group of normal and various clinical subjects (psychosis) [9]. The authors conclude that, when other clinical groups are combined with schizophrenic participants, accuracy rates are compromised. Although studies using CLANG are limited, the scale has shown good performance when discriminating between normal and psychotic participants, and moderately good performance when differentiating between participants with other psychotic disorders.

11.1.2.3 Strengths and Weaknesses of the Scales

Observer-rated scales have been instrumental in SLD research and are in widespread clinical use. The main advantage is their simplicity with concise definitions and instructions for raters [1, 10]. TLC is straight forward to use [19], while CLANG requires a basic understanding of general linguistic concepts [10]. Both scales have Likert scores, offer flexible scoring with separate items, have subsections, and include total scores [10, 19]. Statistically, both TLC [3, 5, 7, 11, 33, 36, 37] and CLANG [8–10] have good reliability and validity. Furthermore, observer scales, if applied live, have the advantage of face-to-face interactions that allow better interpretation of speech, because of the accompanying non-verbal expressions [1].

Nevertheless, observer-rated scales suffer from shortcomings. Subjectivity can bias the assessment, as the result of different levels of expertise, background, and training, or cultural and personal factors come into play [29]. For example, one study that employed six psychiatrists with different levels of training resulted in widely distributed scores [5]. Cultural background can also impact evaluations. In one study, American psychiatrists mistakenly identified some of the colloquialisms of the British participants as indicators of language impairments [5].

In clinical practice, it is difficult to ascertain if high inter-rater reliability among raters really exists, while research practice requires standardized training and testing procedures [10, 16, 19] that provide this information. However, even in a research setting, poor inter-rater reliability on certain items is not always identified, especially if items occur infrequently [3, 12, 20].

According to a Canadian survey, many psychiatrists do not use scales because they are apprehensive about their usefulness [30]. The standard practice is to conduct a Mental Status Examination (MSE) [35] and consult the DSM-IV-TR [17]. Although the MSE and DSM can be criticized for lacking detailed Likert-like ratings [35], they remain practical in clinical settings, where time constraints are paramount.

Notwithstanding the strengths of scales, their weaknesses warrant an investigation into other methods of evaluating SLD. One possibility is to look beyond manual methods towards automated techniques.

11.1.3 Advantages of Automated Measures

It is now possible for automated methods to offer benefits similar to those of standardized observer-rated scales. Fully automated programmes are user-friendly, since they remove the labour of scoring and the generation of reports that are interpretative in nature [34]. The greatest advantage of automated methods is the objectivity they provide, which contrasts with the inherent subjectivity of scales [34].

Since automated programmes offer time and cost saving advantages [34], mental health professionals may welcome the idea of using such measures in everyday practice. Unlike clinicians who need substantial time and training to gain experience in assessing SLD efficiently, automated programmes use “machine learning” to mimic experienced clinical decision making processes [14].

11.1.3.1 Ex-Ray

Mental health practitioners, consciously and unconsciously, analyse the language of their patients, identify patterns and use this information for clinical assessment, or classification, using DSM IV or ICD10. Machine learning techniques can be applied in psychiatry to analyse data, including speech and language. This chapter investigates a novel approach to psychiatric classification and diagnostic screening that utilises widely available data. This computational approach is compared to observer rated scales.

This study uses a computational method called Ex-Ray [14] which uses a particular machine learning technique: Support Vector Machines (SVMs). SVMs are a class of algorithms which are well-suited to learning classification and regression tasks. SVMs have also been used for ranking problems, e.g., a ranking with regard to the severity of a disorder. SVMs have been utilised in a wide variety of tasks, including text and image classification as well as bio-medical applications. SVMs utilise kernels to work in a high-dimensional feature space. In binary classification tasks, the margin between the two classes is maximized in order to find the best possible separation. An SVM algorithm finds an optimal decision boundary in the multi-dimensional feature space by finding a hyperplane, which has maximum distance from prototypical samples, called “support vectors.” The learned hyperplane is then used as a decision boundary of the SVM classifier.

11.1.4 Objective and Hypotheses

Since the computational method offers similar advantages as observer-rated scales and overcomes some of their weaknesses, the objective of this study is to compare Ex-Ray with TLC and CLANG in its ability to determine SLD. As an automated and objective instrument not affected by subjectivity and inter-rater issues, the computational method should achieve a higher classification accuracy than the observer-rated scales on a sample of randomised schizophrenic and normal speech samples.

The minor hypotheses of this study are that each of the measures (TLC, CLANG and Ex-Ray) will be able to correctly separate schizophrenic from normal participants by use of speech samples. The main hypothesis is that Ex-Ray will have the same performance as the scales or even outperform these in terms of accuracy and quality of receiver operating characteristic (ROC) curves.

11.2 Method

11.2.1 Participants

A total of fifty-four subjects (n = 27 schizophrenics; n = 27 controls) participated in this study. Schizophrenic subjects were recruited from two centres of the Singapore Association of Mental Health (SAMH) by a supervisor who was unaware of the purpose of the study. These participants were at the post-acute stage of the DSM-IV-TR4 diagnosis without other comorbid issues and were taking neuroleptics.

Non-psychotic participants were recruited from a batch of first year psychology students of James Cook University Australia (Singapore campus). The exclusion criteria were: no recent history of mental illness, depression, substance abuse or alcohol intoxication. Normal participants were rewarded with one course credit point to fulfil part of the requirements for some core psychology modules, while schizophrenia participants were rewarded with a cash token of two Singapore dollars.

The age of schizophrenic participants ranged from 21 to 62 (M = 43.9, SD = 10.5), with 13 males and 14 females. The ethnic composition of the sample was 81 % Chinese, 12 % Indians and 7 % Malays (reflecting the demographics of Singapore). All the participants could converse in English, even though at home, the majority (74 %) spoke other languages (e.g., Chinese, Tamil and Malay) and a minority (26 %) spoke mostly English. The age of the non-psychotic participants ranged from 18 to 37 (M = 22.6, SD = 4.8), with 11 males and 16 females. Among them, 78 % were Chinese, 19 % were Indians and 4 % were Malays. All participants were conversant in the English language, with the majority (86 %) speaking English at home, while the rest (14 %) used mainly other languages.

11.2.2 Apparatus

Ex-Ray performed well in a preliminary SLD study [14], achieving a 77 % accuracy in distinguishing schizophrenics from normal participants. TLC consists of 18 phenomenological items, each to be rated on a 5-point Likert scale (0 = no more than one instance of SLD, 1 = mild SLD, 2 = moderate SLD, 3 = severe SLD, and 4 = extreme SLD). As stated above, TLC has good psychometric properties with high inter-rater reliability ($k > 0.80$) for common items such as poverty of speech and content and a high intra-class correlation coefficient that ranged from 0.78 to 0.85 in previous studies [8, 24]. CLANG has 17 linguistic items to be rated on a 4-point Likert scale (0 = normal speech, 1 = mild SLD, 2 = moderate SLD, and 3 = severe, pervasive SLD). CLANG has high internal reliability for the subscales (the entire scale's alpha coefficient is 0.76) with an intra-class correlation of 0.88 for the full scale [18]. Convergent validity was established with the Brief Psychiatric Rating Scale (BPRS) [18] and the Schneiderian First-Rank Symptom (SFRS) [9], while criterion validity was established with TLC [18].

11.2.3 Procedure

Ethical clearance was obtained from the James Cook University (JCU) Human Research Ethics Committee.

11.2.3.1 Interview

After obtaining informed consent, an unstructured interview was conducted with each participant. Interviews were approximately 20 min in length, conducted in English and audio recorded. A list of open-ended questions had been prepared earlier by referring to the manuals of TLC and CLANG to guide the interview. The interviews with schizophrenic participants were carried out in a Singapore Association of Mental Health room with a clinician present, while the interviews with non-clinical participants took place in a JCU research lab.

There were two teams of interviewers (Team A and B), each comprised of two persons. The interviewers were not blind to the purpose of the study. As there were two locations, participants from one SAMH centre were interviewed by Team A, while those in the other centre were interviewed by Team B. In each of the locations, half of the schizophrenic participants were interviewed by the first interviewer, while the other half of the participants were interviewed by the second person. The normal participants were divided into two groups and interviewed in a similar manner by the two teams that had conducted the schizophrenic group interviews.

11.2.3.2 TLC and CLANG Scoring

The content of each interview was independently scored by two raters on every item of TLC and CLANG scales. A global score was also given for each scale based on the overall content of the interview. Prior to conducting the ratings, a practice session was held to facilitate the understanding of items and to discuss potential discrepancies in rating behaviour. All audio recorded samples were de-identified and randomized. To ensure blind ratings, interviews conducted by Team A were rated by Team B and vice versa. Half of these samples were rated with TLC first followed by the CLANG; the remainder were rated in reverse order. The two global ratings awarded by the raters for each participant were averaged into one rating per participant and scale.

11.2.3.3 Ex-Ray Learning

Speech samples were manually transcribed by the interviewers. The transcripts were handed to an Ex-Ray programmer for analysis. The transcripts were segmented into words, each word was stemmed, and functional words (such as "to" and "the") were removed. The words were ranked according to document frequency, and words with less than two occurrences were removed. The selected words were then used as attributes, and thus each transcript was represented as a vector of the frequencies of words.

In Ex-Ray, the input features are normalized frequencies of words. Recordings of interviews are transcribed into text documents and each document is segmented into a list of words. Words in the whole corpora are ranked and infrequent words are discarded. The normalized frequency (normalized to the length of documents) of each word in the ranked list of words are then used as input features of each document. That is, an input feature $\mathbf{x} = (v_1, \ldots, v_d) \in \Re^d$ is a d-dimensional vector and each feature value v_i is a normalized frequency of a word:

$$v_i = \text{Frequency of the } i\text{th Word} \,/\, \text{The Length of Document} \qquad (11.1)$$

where the length of document is the L2-Norm of word frequencies in a document. The task is then to find a function $f : \mathbf{x} \rightarrow D$ by use of SVM learning that maps the input feature vector $\mathbf{x} \in \Re^d$ to a class label $D \in \{+1, -1\}$.

11.2.3.4 Statistical Analysis

The data from all participants was included in the statistical analysis. Data analyses were performed using the Predictive Analytics Software (PASW Statistics, v.18) with alpha level set at 0.05. Only averaged global scores of TLC and CLANG were used for analysis. For the assessment of group membership, participants were identified as schizophrenics if they scored 1 or more on the averaged global score of each scale, or as control if they scored less than 1.

Standardised global ratings were calculated for TLC (M = 0.95, SLD = 1.14), CLANG (M = 0.79, SLD = 0.95), and Ex-Ray (M = −0.06, SLD = 0.97), because of the different Likert scores in the two observation scales. Cronbach's alphas were computed on the averaged global scores of TLC and CLANG for inter-rater reliability within Team A and Team B, respectively.

In order to obtain the classification accuracy of each measure, (ROC) curve analyses were conducted. Finally, to test for differences between the ROC curves of Ex-Ray and TLC, as well as between Ex-Ray and CLANG, z-score analysis was performed by manually calculating the critical ratio z.

11.3 Results

11.3.1 Differences in Word Frequencies Between the Subject and Control Groups

The mean relative-frequency of unique words in each transcript differed significantly between the subject and control groups (see Figs. 11.1, 11.2, 11.3). The mean relative-frequency $\mu 1$ of the control group was larger than the mean relative-frequency $\mu 2$ of the schizophrenia group ($N1 = 27$, $N2 = 27$, $\mu 1 = 0.3172$, $\mu 2 = 0.2720$, $SE1 = 0.0160$, $SE2 = 0.0088$, μ diff $= 0.045$, SE_diff $= 0.0182$, $t = 2.469$, $df = 40.6$, $p < 0.009$).

In addition, there was a difference in the number of new words added to the overall corpus with each 100 word segment. The average number of new words added to the control group corpus was consistently larger than the average number of new words added to the subject group corpus in 22 of the 100-word segments ($N = 22$, μ diff $= 2.78$, SE_diff $= 0.327$, $t = 8.60$, $df = 21$, $p < 0.002$).

Since the transcripts obtained from the schizophrenia and control groups differed in length, blocks of 100 word segments were formed by use of the transcripts

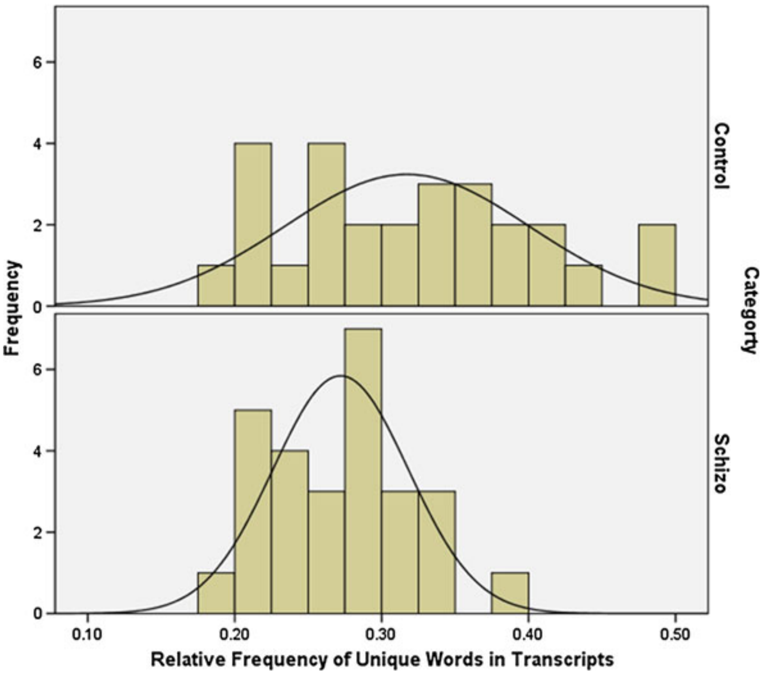

Fig. 11.1 Histogram of relative frequencies of unique words in each transcript for the subject (schizophrenia) and control groups

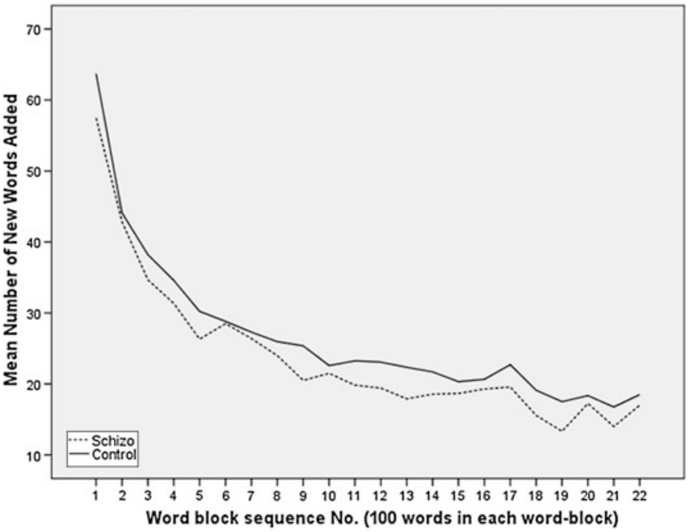

Fig. 11.2 Mean number of new words added to the corpus with each 100-word block (schizophrenia and control groups)

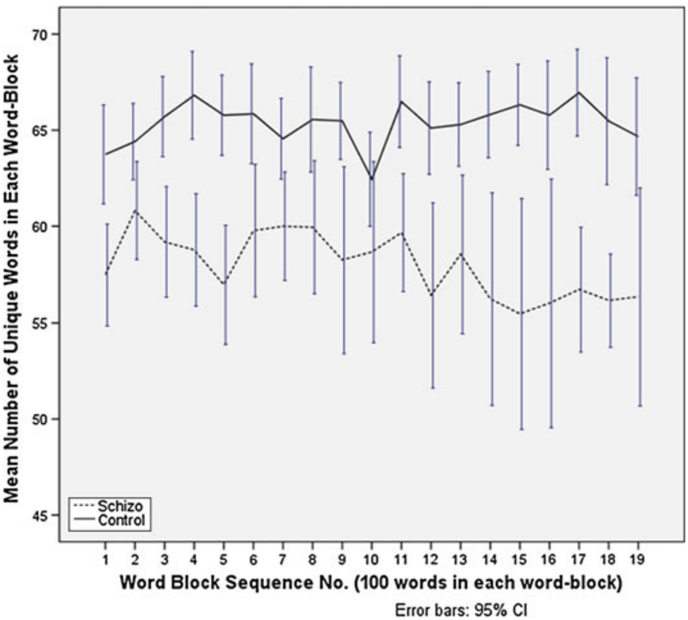

Fig. 11.3 Mean number of unique words in each 100-word block of transcripts for the schizophrenia and control groups

from both groups. The mean number of unique words in each 100 word segment again differed significantly between the subject and control groups ($N1 = 43$, $N2 = 22$, $\mu 1 = 63.9$, $\mu 2 = 57.8$, $SE1 = 0.486$, $SE2 = 0.394$, $\mu\_diff = 6.07$, $SE\_diff = 0.737$, $t = 8.24$, $df = 63$, $p < 0.001$).

11.3.2 Ex-Ray Versus TLC and CLANG

The Cronbach's alpha inter-rater reliability for TLC for Team A and Team B was 0.85 and 0.95, respectively, while that of CLANG was 0.84 and 0.96 for Team A and Team B.

Thus, the first hypothesis which stated that TLC would correctly discriminate between schizophrenic and normal participants was supported, as ROC analysis revealed a high value for area under the curve (AUC) for TLC that was not due to chance ($AUC = 0.98$, $SE = 0.02$, $p < 0.001$). The ROC curve (see Fig. 11.4) had a sensitivity of 0.96, specificity of 0.96, and an optimal cut-off point of −0.62. This indicates that TLC differentiated between schizophrenic and normal participants with a high accuracy of 98 %. It correctly identified 96 % of the schizophrenics, and 96 % of the normal participants.

The second hypothesis stated that CLANG would correctly classify schizophrenic and normal participants. This hypothesis was also supported with a high AUC value ($AUC = 0.97$, $SE = 0.02$, $p < 0.001$), a sensitivity of 0.96, a specificity of 0.96 at an optimal cut-off point of −0.56. CLANG differentiated between schizophrenic and non-clinical participants with an accuracy of 97 %, and it correctly predicted 96 % of schizophrenic patients and 96 % of normal participants.

In line with the third hypothesis that Ex-Ray would correctly classify schizophrenic and non-clinical participants, the AUC value was significantly higher than

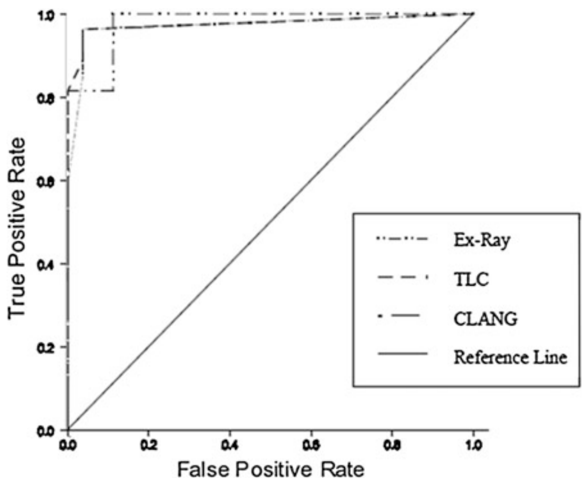

Fig. 11.4 Receiver operating characteristic *curves* of the thought, language and communication (TLC) scale, the clinical language disorder rating (CLANG) scale, and Ex-Ray

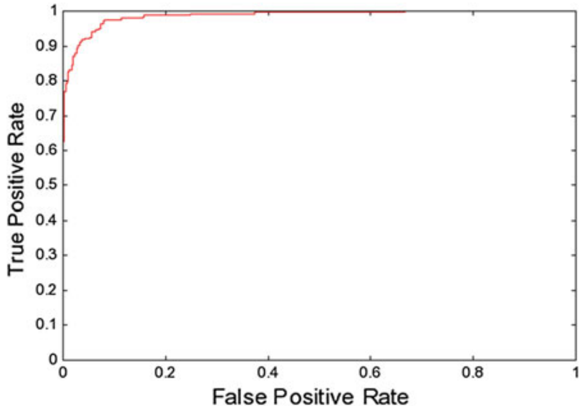

Fig. 11.5 ROC *curve* for Ex-Ray only if 100-word blocks are used

0.5 (AUC = 0.98, SE = 0.01, $p < 0.001$), with a sensitivity of 0.89, specificity of 0.89 and an optimal cut-off point of −0.13. As expected, Ex-Ray discriminated between schizophrenics and normal participants with an accuracy of 98 %. In addition, it correctly recognised 89 % of the schizophrenic cases and 89 % of non-psychotic participants.

It was also hypothesized that Ex-Ray would outperform both the TLC and CLANG scales in differentiating between schizophrenics and non-clinical participants. Contrary to this hypothesis, z-score analysis showed that the accuracy rate of Ex-Ray did not differ significantly from that of TLC, as the obtained z-score (0.08) was less than the critical value at alpha level of 0.05 (one-tailed). Similarly, with an obtained z score (0.25) that is less than the critical value, Ex-Ray did not differ significantly from CLANG in its accuracy.

Ex-Ray, however, performed at a very high level. If blocks of 100 words were used for classification instead of entire transcripts, the classification accuracy was very high with an AUC of 0.98 (see Fig. 11.5).

11.4 Discussion

This study compared observer-rated scales, such as TLC and CLANG, with Ex-Ray, an automated method for evaluating SLD. All minor hypotheses stating that TLC, CLANG and Ex-Ray could differentiate between schizophrenics and normal participants were confirmed. However, the main hypothesis that Ex-Ray would surpass the performance of observer-rated scales with its predictive accuracy was not supported, as all three measures obtained similarly high accuracy rates. In this study, the abilities of these observer-rated scales to correctly identify schizophrenic participants and to differentiate them from controls is consistent with past efforts that used TLC [3, 5, 11] as well as those that used CLANG [8, 9].

For Ex-Ray, the high classification rates obtained indicate that it has diagnostic value as a SLD measure, comparable to the performance of observer-rated scales.

Why did the computational method perform so well? The difference in unique words used (Figs. 11.1, 11.2, 11.3) is evidence for poverty of speech (and possibly poverty of content) in the schizophrenic group. Since there is a significant difference in the vocabulary used by both groups, SVMs that take word frequencies as input ("bag of words") can easily determine a decision boundary between both groups. Further analyses of the data may also reveal positive symptoms in the speech samples obtained from the schizophrenics, which would make the classification task even easier.

There are several explanations why Ex-Ray did not surpass the performance of the observer-rated scales, despite being a more objective measure. Firstly, there was an unexpectedly high inter-rater reliability among the raters (Cronbach's alpha was 0.90 on both the scales). This value is higher than those reported in previous studies [3, 10]. The scales' high inter-rater reliability and classification accuracy may be attributed to the small sample, compared to past studies [3, 10]. Next, as the ratings were not conducted live, the raters had more time to refer to manuals to determine scores. Since the interviews allowed for open-ended responses, some participants mentioned coming to 'school' or to the 'centre'. Indirectly, this may have provided information about the clinical background of the participants. Also, all raters were trained together and had similar background and experience as psychology students.

Ex-Ray correctly identified 89 % of the schizophrenic and non-clinical participants. This equates to incorrect classifications of 11 % for each group, which is higher than the incorrect classifications (4 %) obtained by both the scales. The higher number of classification errors by Ex-Ray can be partially explained by cultural language differences. The speech samples that Ex-Ray used came from a majority of participants who were Chinese (80 %), followed by Indians and Malays (6 %). English spoken by the Singaporean sample is peppered with ethnic words to describe local food, songs, and movies. Such words were elicited due to the nature of the questions. The human raters were familiar with most of the local "lingo" and relied on their judgment and background to decide if an expression was, indeed, an instance of a SLD. Obviously, Ex-Ray has no access to this kind of background knowledge.

In a follow-up study with a different control group, which matched the schizophrenic group in terms of ethnicity, socio-economic status and educational background, the Ex-Ray performance was confirmed with an overall accuracy that matched the results in this study. It is, therefore, a valid conclusion that support vector machines perform very well on this type classification task.

Acknowledgments The authors wish to thank all those involved in the data collection: Lynette Tham, Esther Hong, and Sabrina Yeo, and all the mental health professionals at the Singapore Mental Health Association, as well as the participants in this study. Thank you also to Margaret Lech, Adam Vogel, and Susan Wright for providing comments on earlier versions of this document.

References

1. Andreasen, N.C.: Thought, language, and communication disorders: i. Clinical assessment, definition of terms, and evaluation of their reliability. Arch. Gen. Psychiatry **36**(12), 1315–1321 (1979). doi:10.1001/archpsyc.1979.01780120045006
2. Andreasen, N.C.: Scale for the assessment of thought, language, and communication (TLC). Schizophr. Bull. **12**(3), 473–482 (1986)
3. Andreasen, N.C., Grove, W.M.: Thought, language, and communication in schizophrenia: Diagnosis and prognosis. Schizophr. Bull. **12**(3), 348–359 (1986). doi:10.1093/schbul/12.3.348
4. Bearden, C.E., Rosso, I.M., Hollister, J.M., Sanchez, L.E., Hadley, T., Cannon, T.D.: A prospective cohort study of childhood behavioural deviance and language abnormalities as predictors of adult schizophrenia. Schizophr. Bull. **26**(2), 395–410 (2000)
5. Berenbaum, H., Oltmanns, T.F., Gottesman, I.I.: Formal thought disorder in schizophrenics and their twins. J. Abnorm. Psychol. **94**(1), 3–16 (1985)
6. Breier, A., Berg, P.H.: The psychosis of schizophrenia. Biol. Psychiatry **46**, 361–364 (2003)
7. Buchanan, R.W., Strauss, M.E., Breier, A., Kirkpatrick, B., Carpenter, W.T.J.: Attentional impairments in deficit and nondeficit forms of schizophrenia. Am. J. Psychiatry **154**(3), 363–370 (1997)
8. Ceccherini-Nelli, A., Crow, T.J.: Disintegration of the Components of Language as the Path to a Revision of Bleuler's and Schneider's Concepts of Schizophrenia, vol. 182 (2003) First published. doi:10.1192/bjp.02.211
9. Ceccherini-Nelli, A., Turpin-Crowther, K., Crow, T.J.: Schneider's first rank symptoms and continuous performance disturbance as indices of dysconnectivity of left- and right-hemispheric components of language in schizophrenia. Schizophr. Res. **90**(1), 203–213 (2007)
10. Chen, E.Y.H., Lam, L.C.W., Kan, C.S., Chan, C.K.Y., Kwok, C.L., Nguyen, D.G.H., Chen, R.Y.L.: Language disorganisation in schizophrenia: Validation and assessment with a new clinical rating instrument. Hong Kong J. Psychiatry **6**(1), 4–13 (1996)
11. Cuesta, M.J., Peralta, V.: Does formal thought disorder differ among patients with schizophrenic, schizophreniform and manic schizoaffective disorders? Schizophr. Res. **10**(2), 151–158 (1993)
12. Cuesta, M.J., Peralta, V.: Thought disorder in schizophrenia. Testing models through confirmatory factor analysis. Eur. Arch. Psychiatry Clin. Neurosci. **249**(2), 55–61 (1999)
13. DeLisi, L.E.: Speech Disorder in Schizophrenia: Review of the Literature and Exploration of its Relation to the Uniquely Human Capacity for Language, vol. 27 (2001) First published
14. Diederich, J., Al-Ajmi, A., Yellowlees, P.: Ex-ray: Data mining and mental health. Appl. Soft Comput. J. **7**(3), 923–928 (2007)
15. Docherty, N.M., Cohen, A.S., Nienow, T.M., Dinzeo, T.J., Dangelmaier, R.E.: Stability of formal thought disorder and referential communication disturbances in schizophrenia. J. Abnorm. Psychol. **112**(3), 469–475 (2003)
16. Docherty, N.M., DeRosa, M., Andreasen, N.C.: Communication Disturbances in Schizophrenia and Mania, vol. 53 (1996) First published. doi:10.1001/archpsyc.1996.01830040094014
17. DSM-IV-T.R.: Diagnostic and Statistical Manual of Mental Disorders, Text Revision, 4th edn. American Psychiatric Association
18. Elvevag, B., Foltz, P.W., Rosenstein, M., DeLisi, L.E.: An automated method to analyze language use in patients with schizophrenia and their first-degree relatives. J. Neurolinguistics **23**(3), 270–284 (2009)
19. Elvevag, B., Foltz, P.W., Weinberger, D.R., Goldberg, T.E.: Quantifying incoherence in speech: An automated methodology and novel application to schizophrenia. Schizophr. Res. **93**(1–3), 304–316 (2007)

20. Faber, R., et al.: Comparison of schizophrenic patients with formal thought disorder and neurologically impaired patients with aphasia. Am. J. Psychiatry **140**(10), 1348–1351 (1983)
21. Fraser, W.I., King, K.M., Thomas, P., Kendell, R.E.: The Diagnosis of Schizophrenia by Language Analysis, vol. 148 (1986) First published. doi:10.1192/bjp.148.3.275
22. Harrow, M., Marengo, J.T.: Schizophrenic Thought Disorder at Followup: Its Persistence and Prognostic Significance, vol. 12 (1986) First published. doi:10.1093/schbul/12.3.373
23. Holzman, P.S., Shenton, M.E., Solovay, M.R.: Quality of thought disorder in differential diagnosis. Schizophrenia bulletin; schizophrenia. Bulletin **12**(3), 360–372 (1986)
24. Jampala, V.C., Taylor, M.A., Abrams, R.: The diagnostic implications of formal thought disorder in mania and schizophrenia: A reassessment. Am. J. Psychiatry **146**(4), 459–463 (1989)
25. Kuperberg, G.R.: Language in schizophrenia. Part I. An introduction. Lang. Linguist. Compass **4**(8), 576–589 (2010)
26. Liddle, P.F., Ngan, E.T.C., Caissie, S.L., Anderson, C.M., Bates, A.T., Quested, D.J., White, R., Weg, R.: Thought and Language Index: An Instrument for Assessing Thought and Language in Schizophrenia, vol. 181 (2002) First published. doi:10.1192/bjp.181.4.326
27. Makowski, D., Waternaux, C., Lajonchere, C.M., Dicker, R., Smoke, N., Koplewicz, H., Min, D., Mendell, N.R., Levy, D.L.: Thought disorder in ado-escent-onset schizophrenia. Schizophr. Res. **23**(2), 147–165 (1997)
28. McKenna, P., Oh, T.M.: Schizophrenic Speech: Making Sense of Bathroots and Ponds that Fall in Doorways. Cambridge University Press, Cambridge (2005)
29. Myers, D.: Social Psychology, 9th edn. McGraw-Hill International Edition, New York (2008)
30. Oluboka, O., Stewart, S., Haslam, D., Wodlinger, J., Adams, S.: The Use of Rating Scales by Canadian Psychiatrists: Qualitative and Quantitative Evidence. http://www.priory.com/psychiatry/Rating_Scales_Canadian_Psychiatry.htm. http://www.priory.com/psychiatry/Rating_Scales_Canadian_Psychiatry.htm. Accessed 10 Jan 2010
31. Ott, S.L., Roberts, S., Rock, D., Allen, J., Erlenmeyer-Kimling, L.: Positive and negative thought disorder and psychopathology in childhood among subjects with adulthood schizophrenia. Schizophr. Res. **58**(2), 231–239 (2002)
32. Parnas, J., Schulsinger, F., Schulsinger, H., Mednick, S.A., Teasdale, T.W.: Behavioral Precursors of Schizophrenia Spectrum: A Prospective Study, vol. 39 (1982) First published. doi:10.1001/archpsyc.1982.04290060020005
33. Peralta, V., Cuesta, M.J., de Leon, J.: Formal thought disorder in schizophrenia: A factor analytic study. Compr. Psychiatry **33**(2), 105–110 (1992)
34. Roid, G.H., Johnson, W.B.: Computer assisted psychological assessment. In: Rey-nolds, C.R. (ed.) Comprehensive Clinical Psychology 3 (1998)
35. Sadock, B.J., Sadock, V.A.: Synopsis of Psychiatry: Behavioural Sciences in Clinical Psychiatry, 9th edn. Lippincott Williams & Wilkins, Philadelphia (2003)
36. Simpson, D.M., Davis, G.C.: Measuring thought disorder with clinical rating scales in schizophrenic and nonschizophrenic patients. Psychiatry Res. **15**(4), 313–318 (1985)
37. Taylor, M.A., Reed, R., Berenbaum, S.A.: Patterns of speech disorders in schizophrenia and mania. J. Nerv. Mental Dis. **182**(6), 319–326 (1994)
38. Wykes, T., Leff, J.: Disordered speech: Differences between manics and schizophrenics. Brain Lang. **15**(1), 117–124 (1982)