Empirical evaluation of language disorder in schizophrenia

Massoud Stephane, MD; Giuseppe Pellizzer, PhD; Charles R. Fletcher, PhD; Kate McClannahan, BA

Stephane and McClannahan — Psychiatry Service Line and the Domenici Research Center for Mental Illness, VA Medical Center; Stephane and Pellizzer — Brain Sciences Center, VA Medical Center; Stephane — Department of Psychiatry, University of Minnesota; Stephane and Fletcher — Department of Psychology, University of Minnesota; Pellizzer — Department of Neuroscience, University of Minnesota, Minneapolis, Minn.

Objective: Studies of the content of speech and of verbal hallucinations in schizophrenia point to dysfunction at multiple levels of language. In this study, we empirically evaluated language processes.

Methods: We examined the performance of 22 schizophrenia patients and 11 healthy control subjects with procedures designed to explore the sublexical, lexical, semantic, syntactic and discourse levels of language processing.

Results: Schizophrenia patients exhibit impairment in the recognition of incorrect, but not correct, linguistic stimuli at all but the sublexical level of language processing. The patients were not impaired in the recognition of nonlinguistic stimuli.

Conclusion: This language-specific differential impairment could explain speech abnormalities in schizophrenia. The nonrecognition of incorrect linguistic information would prevent patients from correcting the abnormal speech they may occasionally produce. A model of decreased power of linguistic computations (reduced number of operations) adequately accounts for this differential impairment.

Introduction

Language disorder in schizophrenia has been described since the early accounts of this illness. Further, it has been suggested that language and psychosis have a common evolutionary origin. Although schizophrenic speech is easily recognizable, it is difficult to define. Nonetheless, multiple studies have examined speech samples of schizophrenia patients and reported anomalies at multiple levels of language processing.

At the lexical level, anomalies such as pronounceable non-word and confusion of antonyms were noted. At the sentence level, Chaika described aberrations such as subject or verb incompatibilities with sound grammatical structure (e.g., "the house burnt the cow horrendously always"), failures to pronominalize or delete (e.g., "I gave my friend food so my friend would not go hungry"), errors in tense and article choice.

Many syntactic peculiarities were also reported at the sen-
tence and discourse level. For example, Pylyshyn found more use of passive voice, perfect tense and state verbs but less use of achievement verbs and qualifying subordinators (e.g., if, since). Additionally, patients with thought disorders exhibited less use of conjunction links (e.g., and, but), pronounomialization, reduction in the syntactic complexity and syntactical deviance.

Other discourse abnormalities include lack of discourse markers for sequencing (e.g., finally, thus), announcing contradiction (e.g., but, however) and similarity (e.g., and) and association between sentences subordinated by phonetic and semantic features rather than the topic. Reference failures (e.g., “a commuter and a skier are on a ski lift and he looks completely unconcerned”) were also reported in patients with thought disorders and in patients with schizophrenia and their unaffected siblings and parents. Grammatical deviance and incoherent discourse were also reported. Using the Hunt test (i.e., composing a multisentence text from a controlled set of input sentences), patients significantly misrepresented the meaning of the input sentences.

In addition to speech aberrations, language disorder is implicated in the pathogenesis of auditory verbal hallucinations (AVHs), a frequent symptom in schizophrenia.

Studies have shown a correlation between subvocal speech and AVH and between the activation of Wernicke’s area and AVH. Earlier theories hypothesized that patients are hearing their own subvocal speech. However, because blocking subvocal speech does not alleviate AVH, it was suggested that both subvocal speech and AVH are related to a central pathological process — that of anomalous speech generation. According to this theory, activation of Wernicke’s area constitutes the final common pathway for the hallucinatory experience. Further, a recent study showed that patients report hearing single words, sentences or conversation differentially, which probably indicates dysfunctions of corresponding levels of language neural resources.

The above studies point to possible dysfunctions at multiple levels of language processing in schizophrenia. However, 4 points need to be addressed before making such a conclusion. First, speech samples only partially reflect processes involved in speech generation. The speaker’s assumptions about the understanding of the listener and context information are both integrated in speech generation processes but cannot be detected in external speech. Therefore, studies using speech samples could lead to erroneous conclusions, especially at the sentence and discourse levels. Second, language processes are specialized but interconnected. Therefore, difficulties at a given level of language processing affect the function of another level. Consequently, it could be difficult to make conclusions regarding the specificity of the affected levels of language processing from AVH or studies using speech samples. Third, studies of speech content or the content of AVHs investigate language disorder in patients with thought disorders and hallucinations, respectively. Therefore, these studies provide a limited view of the extent of language problems in schizophrenia. Fourth, of the above findings, only discourse disruption and poverty of speech and of speech content are found to be specific to schizophrenia.

The motivation for this study is to address the above methodological limitations to clarify the nature of language disorder in schizophrenia. We investigated patients’ linguistic performance on tests designed to explore sublexical, lexical, sentence and discourse processing. This method avoids the first limitation because there is no hearer involved. The second limitation is addressed by investigating processes specific to each level of language processing separately. Finally, the method allowed for the investigation of linguistic operations independent of symptomatology as well as the examination of the relation between these operations and symptoms.

Methods

Subjects

Twenty-two patients (20 males, 2 females) meeting the Diagnostic and statistical manual of mental disorders, fourth edition (DSM-IV) criteria for schizophrenia or schizoaffective disorder participated in this study. They were recruited from the outpatient clinic at the Minneapolis VA Medical Center (VMAC). Diagnostic evaluation was carried out by a research assistant who was trained to use the Structured clinical interview for the DSM-IV (SCID). Consensus diagnosis was established with the treating psychiatrist or by 3 other psychiatrists when the treating psychiatrist was not available.

The severity of psychopathology was assessed with the Brief Psychiatric Rating Scale (BPRS) and the Positive and Negative Symptoms Scale (PANSS). The duration of illness was derived from reviews of patient records. All patients but 1 were medicated with atypical antipsychotic medications. The chlorpromazine equivalent doses of medications were estimated according to the methods of Woods and Van Kammen and Marder. Measures of premorbid intellectual functioning were obtained, using the National Adult Reading Test (NART).

We also studied 11 healthy control subjects (8 males, 3 females), recruited by advertisements posted at the University of Minnesota and Minneapolis VA Hospitals. Subjects were screened for mental illness with the SCID. Subjects had neither a history of mental illness nor a family history of schizophrenia in first-degree relatives.

All subjects were native English speakers and were right-handed. Handedness was assessed by the Edinburgh Handedness Inventory; none had neurological disease or major medical illness. The protocol was approved by the VAMC and the University of Minnesota Institutional Review Boards. All subjects understood the risks and benefits associated with the study and gave written informed consent.

The patient group did not differ significantly (p < 0.05) from the control group with respect to age, personal or parental level of education, and premorbid overall and verbal intelligence. However, premorbid intellectual capacity was marginally significantly lower (p = 0.05) in the patient group, which indicates latent difficulty preceding illness onset. Table 1 summarizes the demographic and clinical characteristics of the subjects.
**Linguistic task**

A linguistic task (Fig. 1) was designed based on standard psycholinguistic procedures. Subjects had to distinguish between correct and incorrect stimuli at the (A) sublexical, (B) lexical, (C) semantic, (D) syntactic and (E) discourse levels of language processing. The correct stimuli were meaningless pronounceable letter strings in A and common words, which have a similar mean frequency of occurrence in the English language in B, C, D and E. The mean frequency was 80, 81, 79 and 73, respectively. Incorrect stimuli were created by making unpronounceable letter strings in A, changing some letters or the order of letters to create pronounceable nonwords in B, changing the subject and object order to create nonsensical sentences in C, changing the order of words to create ungrammatical (and meaningless) sentences in D and making short stories with causally unrelated sentences in E.

The stimuli were presented individually in the centre of a monitor placed in front of the subjects. The stimuli appeared in sets of 5 at a rate of 1 Hz. Each set was followed by a 10-second response window. The subjects were instructed to read the stimuli silently and distinguish between 2 sets of stimuli: (A) sets with all pronounceable stimuli versus sets with 1 unpronounceable stimulus, (B) all real English words versus sets with 1 nonword, (C) a semantically sound sentence versus a semantically unsound sentence and (D) a grammatical nonsensical sentence versus ungrammatical nonsensical sentence.

In E, each set of 5 words comprised a semantically and syntactically correct sentence (test sentence). This sentence was preceded by the presentation of 2–3 sentences (context sentences), which constituted the beginning of a short story. The context sentences appeared individually for a duration that allows reading at a rate of 3 words per second (normal reading speed). The test sentence was presented 1 word at a time, as described above. Subjects had to decide whether the test sentence was an acceptable continuation to the preceding context sentences. Acceptable continuations were constituted by sentences that were causally coherent, and unacceptable continuations were constituted by causally incoherent sentences.

Twenty correct and 20 incorrect trials were presented in A–E in a randomized block design. Subjects were asked to respond at the end of each trial (5-stimuli set) by pressing buttons marked Yes or No on a response pad with their index or middle fingers, respectively.

Because language operations call on working memory, we matched for working memory load across levels of language processing, using trials with the same number of stimuli in A–E. In C–E, the same syntactic structure was used (“The,” noun, verb, “the,” noun) to match for syntactic complexity in the test sentences. Although the distinction between correct and incorrect linguistic stimuli may not be a function that language processors exert in a natural state, it would require a sound operation of these processors.
for, this task allows us to evaluate the integrity of language processors.

Nonlinguistic task

A nonlinguistic task was designed to evaluate subjects’ general ability to detect anomalies. In this task, a set of 5 Kanji symbols were presented with display parameters similar to above. In 50% of the trials, the same symbol was presented 5 times, and in 50%, 1 of the symbols was different from the other 4 symbols. Subjects were required to distinguish between the 2 types of trials.

Reaction time (RT) control task

This task was designed to evaluate the response latency related to decision-making processes unrelated to linguistic operations. The numbers 1 and 2 were presented randomly, and subjects were required to press response buttons marked 1 or 2 with their index or middle fingers, respectively, according to the number displayed. The above 3 tasks were administered in random order.

Data analyses

Data were analyzed with standard statistical techniques. The analyses of the proportion of correct responses were performed on the angular transformed values to stabilize the variance.

A mixed-model analysis of variance (ANOVA) with the between-subjects factor group (2 levels) and the within-subjects factors condition (6 levels) and stimulus type (2 levels) was used. Post hoc analyses were performed as needed. The RT of correct responses data were analyzed with the ANOVA model. The average RT across repetitions was calculated with the harmonic mean, which is robust to potential outliers. Only trials with RTs greater than 100 ms were used in the analyses.

The RT in the linguistic task was adjusted according to the RT control task to yield the linguistic processing time. The adjusted RT \( \bar{Y}_{ijkl} \) was calculated as follows:

\[
\bar{Y}_{ijkl} (adj) = \bar{Y}_{ijkl} + b (\bar{X}_j - \bar{X})
\]

where \( \bar{Y}_{ijkl} \) is the average RT of subject \( i \) of Group \( j \) for Condition \( k \) and Response type \( l \), \( \bar{X}_j \) is the average RT of Group \( j \) in the control task, \( \bar{X} \) is the overall average RT in the control task, and \( b \) is the slope between the RT in the control task and the RT in the experimental task.

Rank correlation analyses (Spearman’s rho) were used to evaluate the relation between symptomatology, medication, and the experiment outcome measures. The analyses were performed with SPSS 13.0. Effects with \( p < 0.05 \) were considered significant.

Results

Proportion of correct responses

The average proportion of correct responses for each group, condition and stimulus type is plotted in Fig. 2, and the average performance was above the chance level of 0.5 in each case. The results of the ANOVA are presented in Table 2. This analysis showed a significant triple interaction group \( \times \) condition \( \times \) stimulus type, which indicates that condition and stimulus type had a differential effect on the performance of the patient group, compared with the control group.

To investigate the source of the interaction, we analyzed

![Fig. 2: Average proportion of correct responses for each group, condition and stimulus type. The error bars represent the standard error of the mean. The ordinate is in an angular scale. Overlapping errors have been omitted for clarity.](image-url)
the data for each stimulus type separately. The analysis of the responses for the “correct stimuli” showed that there was a significant effect of condition ($F_{1,31} = 19.307, p < 0.0005$) but no significant effect of group or interaction group $\times$ condition ($F_{1,31} = 0.066, p = 0.800$ and $F_{1,31} = 0.725, p = 0.606$, respectively). In contrast, the analysis of the responses for the incorrect stimuli showed significant effects of group, condition, and group $\times$ condition interaction ($F_{1,31} = 8.677, p = 0.006$, $F_{1,31} = 13.065; p < 0.0005$ and $F_{5,155} = 4.055; p = 0.002$, respectively). Further analyses of the performance in the incorrect stimuli trials were done in each condition separately. These analyses showed that the proportion of correct responses differed significantly between the patients and the control subjects in the lexical, semantic, syntactic and discourse conditions ($t_{n} = 2.263, p = 0.031; t_{n} = 3.655, p = 0.001; t_{n} = 2.375, p = 0.024$; and $t_{n} = 2.977, p = 0.006$, respectively) but not in the sublexical and the nonlinguistic conditions ($t_{n} = 0.873, p = 0.389$ and $t_{n} = 0.940, p = 0.355$, respectively).

Therefore, in all but the sublexical level of language and with the incorrect stimuli only, the patients had significantly smaller proportion of correct responses than the control subjects.

Adjusted RT of correct responses

The average adjusted RT of correct responses for each group, condition and stimulus type is plotted in Fig. 3. The results of the ANOVA are presented in Table 3. This analysis also showed a significant triple interaction group $\times$ condition $\times$ stimulus type. This result indicates that condition and stimulus types had a differential effect on the performance of the patient group, compared with the control group.

We applied the same strategy used above to investigate the source of the interaction. The analysis of the “correct stimuli” responses showed that there was a significant effect of condition ($F_{5,155} = 10.773, p < 0.0005$), but no significant effect of group or interaction group $\times$ condition ($F_{5,155} = 0.010, p = 0.920$ and $F_{5,155} = 0.475, p = 0.795$, respectively). In contrast, the analysis of the incorrect stimuli responses showed significant effects of condition and group $\times$ condition interaction ($F_{5,155} = 14.412, p < 0.0005$ and $F_{5,155} = 2.653, p = 0.025$, respectively). The main effect of group was not significant ($F_{1,31} = 0.205, p = 0.654$). Thus, the adjusted RT of correct responses of incorrect stimuli were compared in each condition separately; however, no significant difference was found between the patients and the control subjects in any condition (all $t_{n}$ with $p > 0.05$). Nevertheless, the inspection of Figure 3 shows that the RT of correct responses was longer for the patients compared with the control subjects in the semantic, syntactic and discourse conditions.

Clinical linguistic correlations

Rank correlation (Spearman’s rho) analyses were performed between the clinical measurements (BPRS, negative and positive symptoms scores, thought disorder scores, duration of illness, and the chlorpromazine equivalent doses of medication) and the experiments outcome measures (proportion of correct responses, and the adjusted RT). The latter measures were those obtained with the incorrect stimuli and averaged across the lexical, semantic, syntactic and discourse conditions. Table 4 summarizes the correlation analyses results. No significant correlation was found between the clinical characteristics of patients or their medication status and the outcome measures of the linguistic experiment.

Discussion

Differential impairment in schizophrenia

In this study, schizophrenia patients performed as well as healthy control subjects in recognizing correct linguistic stimuli but were impaired in recognizing incorrect linguistic stimuli. They also exhibited a bias to consider incorrect stimuli as correct. This difficulty was encountered at the lexical, semantic, syntactic and discourse levels but not at the sublexical level. Similarly, patients exhibited slower linguistic processing speed at the semantic, syntactic and discourse levels. This last finding was not statistically significant; nonetheless, it indicates that the impairment in accuracy was not related to speed–accuracy trade-off.
The capacity to recognize incorrect linguistic stimuli depends on the ability to detect anomalies, the degree of difficulty of the task and intact language processing systems. The patients did not differ from the control subjects in the detection of nonlinguistic anomalies. Therefore, the performance difference is language-specific and involves all but the sublexical levels of language processing. Further, because both the nonlinguistic and sublexical tasks are on opposite ends of the spectrum with respect to difficulty, it can be concluded that the differentiation between patients and control subjects was related to the specificity of the investigated language processes rather than to the degree of difficulty of the task. Because the patients performed similarly to the control subjects in the sublexical and nonlinguistic tasks, a general bias to respond “yes” in the patient group cannot account for the findings. Consequently, these findings reflect impairment of language disorder in schizophrenia.

Silent reading requires the activity of most neural resources implicated in language generation. Therefore, this finding is served in this study.

For example, several studies point to a reduced ability to detect anomalies, the degree of difficulty of the task. Because the patients performed similarly to the control subjects in the sublexical and nonlinguistic tasks, a general bias to respond “yes” in the patient group cannot account for the findings. Consequently, these findings reflect impairment of language disorder in schizophrenia.

Generally, it is accepted that language requires coordination between neurolinguistic resources specific to each level of language processing, with several interrelated cognitive functions, such as verbal working memory and attention. Many of these cognitive and psycholinguistic operations were the subject of extensive investigations and were found to be impaired in schizophrenia. However, in our view, most of these deficits do not adequately explain the type of impairment observed in this study.

For example, several studies point to a reduced ability to represent, maintain and update context information at the sublexical, lexical and sentence levels. These deficits should have equal effects with both correct and incorrect linguistic stimuli. Therefore, they do not adequately explain the dissociation of performance found in this study. The same argument would apply to impairment of verbal working memory and the impairment of attention and executive functions.

Language also requires access to episodic and semantic memories. Semantic memory was investigated intensively in schizophrenia research, with paradigms such as semantic priming and semantic categorization tasks. The literature on semantic priming shows conflicting results. The priming effect is theoretically related to the spread of activation within the semantic memory network. The incorrect stimuli in this study constitute a clear break in the semantic relatedness (e.g., a nonword such as “Lerstrave” is not semantically related to any English word). Therefore, increased semantic priming, if such is the case in schizophrenia, cannot be retained as an adequate explanation of the deficits in this study.

Chen and colleagues used a semantic categorization task in which subjects decided whether words belonged to a given semantic category. The authors used words with either different degrees of semantic relatedness or words that were unrelated to the semantic category in question. They found that schizophrenia patients, relative to healthy control subjects, have longer RT for words outside the semantic category. The study by Chen and colleagues, in line with our data, points to the same difficulty that schizophrenia patients have — recognizing what is not the case. Their findings were not replicated.

The findings in this study could reflect an impairment in detecting errors occurring at some language processing levels. Earlier research showed that normal subjects have the ability to monitor and correct the errors they make during a...
It has been shown that a subset of schizophrenia patients are less likely to correct errors in the absence of visual cues. This finding leads to the theory that schizophrenia patients have a deficient internal monitor of "willed actions," which could explain a wide range of psychotic symptoms such as delusions, thought insertion and broadcasting, and verbal hallucinations. Recently, numerous electrophysiological studies have reported smaller error-related negativity (ERN) in schizophrenia, supporting the theory of error monitoring deficits in this illness.

Monitoring for error is an integral part of the act of speaking. It allows a person to correct errors he or she may inadvertently produce while speaking. In this study, the observed impairment in recognizing the erroneous nature of the linguistic material being read (or generated) could result in impaired correction of these errors and would lead to the generation of speech with linguistic aberrations. Therefore, a deficient error monitoring, operating at specific aspects of language processing, could be retained as an explanation of the findings. However, although error recognition is a prerequisite for error correction, impairment in the latter is not necessarily related to impairment in the former. Further, a recent study indicates that the neural activity during conflict detection and making an error (error monitoring) are dissociated.

The findings can be adequately explained by a model of diminished computational power (a decrease in the number of linguistic operations) in schizophrenia. Let us assume that, for a language stimulus to be correct, it must have properties. If the computation capacity is decreased, only a subset of these properties would be evaluated. Consequently, all correct stimuli are recognized because they satisfy properties. However, the incorrect stimuli that satisfy properties but not the remaining properties are included in the correct category. Another possibility is that, if properties are satisfied, the stimulus is considered correct, whereas if a property is not satisfied, additional computations are performed to ensure that this particular property is incongruent. If only evaluations are performed, incorrect stimuli requiring computations are recognized; however, incorrect stimuli needing computations may be considered correct. Therefore, both scenarios would result in adequate recognition of correct stimuli, but in impairment of recognition of incorrect linguistic stimuli and an overinclusion of these stimuli in the correct category.

The model of decreased computational power is consistent with several studies pointing to excessive synaptic pruning and generalized disconnectivity in schizophrenia and with some electrophysiological findings. It has been shown, although inconsistently, that patients show significantly less negativity of N400 or less positivity of P600 with incongruent stimuli during semantic and syntactic processes, respectively. The amplitude of these ERP components is generally related to the difficulty of the task, or to the degree of semantic constraint, in the case of N400; the lower the semantic constraint, or the higher the difficulty, the higher the computations, the larger the amplitude. Therefore, the small amplitude of N400 and P600 could reflect schizophrenia patients’ failure to perform the additional computations associated with incorrect stimuli and thus support the second scenario of the model.
Finally, these findings should be taken in light of 4 considerations: 1) the patients in this study were medicated; nonetheless, there was no correlation between performance and the medications dose. Further, the differential impairment (incorrect vs. correct stimuli) and the normal performance on the nonlinguistic and sublexical procedures indicate that the observed effects are specific to certain operations rather than related to medication effects; 2) the small size of the control group might have obscured smaller effects in the sublexical and nonlinguistic tasks. If this were the case, the impairment at the lexical–discourse levels would be disproportionately higher than those of the nonlinguistic and sublexical tasks; 3) the sample consists mainly of males. Given the subtle sex-dependent differences in schizophrenia, our findings would be robust for the male population but may not be applicable to the female population; 4) whereas most findings would be robust for the male population but may be applicable to the female population.

**Conclusion**

Schizophrenia patients are impaired in recognizing incorrect, but not correct, linguistic information at most language processing levels. This impairment could explain why these patients produce abnormal speech. The nonrecognition of incorrect linguistic information would prevent them from correcting abnormal speech they may occasionally produce. A model of diminished computational power, possibly due to disconnection, adequately explains this impairment.

**Acknowledgements:** We received grant support from the VA Medical Center, the Minnesota Medical Foundation and the MIND Institute. We thank Christa Surerus-Johnson for the clinical evaluation of subjects included in this study.

**Competing interests:** None declared.

**Contributors:** Drs. Stephane, Pellizzer and Fletcher designed the study. Drs. Stephane, Pellizzer and Ms. McClannahan acquired and analyzed the data. Drs. Stephane and Pellizzer wrote the article and all authors revised it. All authors gave final approval for the article to be published.

**References**
