Available online at www.sciencedirect.com

SCIENCE DIRECT®

Psychiatry Research 133 (2005) 159–171

PSYCHIATRY RESEARCH

www.elsevier.com/locate/psychres

Dynamical quantification of schizophrenic speech

Fabrice Leroy[a,b,*], Laurent Pezard[b,c,d], Jean-Louis Nandrino[a,b], Daniel Beaune[a]

[a] *UFR de Psychologie, Université Charles de Gaulle (Lille 3), Domaine Universitaire du Pont de Bois, F59653 Villeneuve d'Ascq., France*
[b] *Institut International Erasme - Université Charles de Gaulle, Lille 3, France*
[c] *Neurosciences Cognitives et Imagerie Cérébrale, LENACNRS UPR 640, Hôpital de la Salpêtrière, 47 Bd de l'Hôpital, F75651 Paris cedex 13, France*
[d] *Institut de Psychologie, Université René Descartes, 71 Bd Ed. Vaillant, F92774 Boulogne-Billancourt cedex, France*

Received 13 September 2003; received in revised form 26 April 2004; accepted 29 July 2004

Abstract

Schizophrenic speech has been studied both at the clinical and linguistic level. Nevertheless, the statistical methods used in these studies do not specifically take into account the dynamical aspects of language. In the present study, we quantify the dynamical properties of linguistic production in schizophrenic and control subjects. Subjects' recall of a short story was encoded according to the succession of macro- and micro-propositions, and symbolic dynamical methods were used to analyze these data. Our results show the presence of a significant temporal organization in subjects' speech. Taking this structure into account, we show that schizophrenics connect micro-propositions significantly more often than controls. This impairment in accessing language at the highest level supports the hypothesis of a deficit in maintaining a discourse plan in schizophrenia.

Keywords: Schizophrenia; Language; Dynamics; Complexity; Symbolic methods; Discourse plan

1. Introduction

The discourse of psychotic patients such as schizophrenic (positive or undifferentiated symptomatology) or manic patients is associated with neologism, fading, obstruction, incoherence, hyperphrasia and associative loosening (Andreasen, 1979a,b). Specific linguistic analysis of the discourse of schizophreni patients demonstrates that syntax is globally spared within sentences (Rochester and Martin, 1979), but that cohesion is impaired between sentences (King et al., 1990; Thomas et al., 1990). Patients thus display a disorganization of semantic systems rather than a lack of semantic knowledge (Goldberg et al., 1998; Paulsen et al., 1996). These language impairments in schizophrenia have been related to cognitive deficits such as working memory and attention impairments (Docherty et al., 1996; Sullivan et al., 1997), inability to structure discourse (Hoffman et al., 1986) or lack of conceptual sequencing (Docherty et al., 2000). Such deficits

* Corresponding author. UFR de Psychologie, Domaine Universitaire du Pont de Bois, F59653 Villeneuve d'Ascq., France.
E-mail addresses: Fabrice.Leroy@univ-lille3.fr (F. Leroy), laurent.pezard@univ-paris5.fr (L. Pezard), nandrino@univ-lille3.fr (J.-L. Nandrino).

0165-1781/$ - see front matter © 2004 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.psychres.2004.07.009

would thus lead to a specific organization of schizophrenic discourse characterized by thematic shifts to idiosyncratic ideas (Harrow et al., 1983) and high semantic priming effects (Minzenberg et al., 2002; Kwapil et al., 1990). Moreover, symptoms such as poverty of speech, perseveration and inappropriate responses have been interpreted within the general model of action self-monitoring (Frith, 1992; Frith et al., 2000). These characteristics thus argue for an impairment of the overall organization of discourse rather than an elementary linguistic problem.

Two important interpretations have been proposed to explain the macroscopic impairment of schizophrenic speech:

(1) Context-processing deficits are defined as an impaired ability to internally represent, maintain and update context information (Cohen and Servan-Schreiber, 1992). Schizophrenic patients depict such a deficit in the pragmatic use of contextual indices (Titone et al., 2000; Mesure et al., 1998; Cohen et al., 1999), in lexical ambiguity tasks (Copland et al., 2002) and in sentence-completion tasks (Bazin et al., 2000). It has been correlated with structural cerebral abnormalities (McCarley et al., 1999). This has been confirmed by studies of deficits in the integration of contextual data, at a neuropsychological level (Cohen and Servan-Schreiber, 1992; Chapman et al., 1976; Plagnol et al., 1996), as well as at a neurological level (Sitnikova et al., 2002; Salisbury et al., 2002). In this framework, a unifying hypothesis has been proposed suggesting an inability to maintain contextual information, and to use this information to inhibit inappropriate responses (Servan-Schreiber et al., 1996; Braver et al., 1999).

(2) Language-production models explain a message generation with the creation of a discourse plan that includes the topic of the discourse and the information to be conveyed (Levelt, 1989). The interaction between language-production processes and thought disorders could explain speech disorders in schizophrenia (Barch and Berenbaum, 1996). In this framework, negative thought disorders, such as reduced verbosity or increased pausing, reflect a deficit in generating a discourse plan, and discourse-coherence disturbances, such as tangential responses or distractible speech, reflect a deficit in maintaining a discourse plan and in monitoring the ongoing content of speech (Barch and Berenbaum, 1997).

These two models allow us to hypothesize that the temporal organization of schizophrenic speech should be impaired. Our study is thus an attempt to characterize the temporal organization of schizophrenic speech using specific dynamical methods.

Quantitative studies of linguistic data, based on free speech samples and oral interviews, found low complexity (frequency, depth and locus of embedded propositions), low integrity (syntactic and semantic errors) and dysfluency (number of pause fillers, false starts and repeated words) in schizophrenic speech (Morice and Ingram, 1982; Thomas et al., 1990). Nevertheless, these studies have usually used computer-assisted grammatical analysis and statistical methods based on counting the occurrence of specific items. Such a procedure neglects the temporal dimension of speech. Because of its inherent temporal dimension, speech can be considered as a dynamical process (Zellner Keller and Keller, 2000; Elman, 1995; Port and Van Gelder, 1995) and can be studied using concepts and numerical methods provided by nonlinear dynamics.

Nonlinear dynamics permit the study of complex phenomena evolving with time (Kaplan and Glass, 1995) and has been considered as a promising framework in psychiatry (Globus and Arpaia, 1994; Ehlers, 1995; Nandrino et al., in press). In the case of schizophrenia, the time evolution of biological, clinical and behavioral indices has been characterized: electroencephalographic (EEG) studies have shown modifications in brain dynamics in schizophrenic patients in resting condition (Kim et al., 2000; Jeong et al., 1998) and during sleep stages (Röschke et al., 1995, 1994; Röschke and Aldenhoff, 1993; Elbert et al., 1992); the time course of psychotic derealization has been associated with a nonlinear dynamical system (Tschacher et al., 1997) and a loss of consistency in response selection and ordering has been observed in a simple choice task (Paulus et al., 1996). Nevertheless, few studies have been devoted to the analysis of temporal and dynamical aspects of speech in schizophrenia.

Nonetheless, linguistic processes in psychopathology have been studied as dynamical phenomena in patient–therapist interactions. These studies showed that the patient–therapist interactions can be quantified using symbolic encoding (Rapp et al., 1991), and that critical transitions occur between periods of stability during the therapeutic process (Schiepek et al., 1997; Kowalik et al., 1997).

In the present article, schizophrenic speech was compared with control speech in a simple immediate-recall task. Speech production was encoded into sequences of discrete symbols according to the linguistic level of each proposition (Kintsch and Van Dijk, 1978). Symbolic sequences were then studied using both classical (i.e., counting) and dynamical methods. The dynamical methods allow both the quantification of the amount of "disorder" in the sequences and a description of transitions between symbols. According to the speech deficiencies and to the memory deficits observed in schizophrenia (Condray et al., 1996; Salamé et al., 1998), we hypothesized that recall complexity should be diminished in patients compared with controls. Moreover, on the basis of context-processing deficits and language-production models, we expected a specific temporal organization in patients' recall.

2. Materials and methods

2.1. Subjects

A group of 10 psychiatric patients was selected according to DSM-IV criteria for schizophrenia. They comprised four women and six men, 21 to 48 years of age (mean: 32.9, S.D.: 10.09) and included three negative, four positive and three undifferentiated patients as defined by the Positive And Negative Syndrome Scale (PANSS, Kay et al., 1987). They were all hospitalized under the same conditions and treated with neuroleptic drugs, with the same dosage. Patients were selected according to their ability, first, to understand the experimental instructions and the text, and then, to recall enough verbal material during the task.

A group of 10 control subjects (3 women and 7 men) matched for age (21 to 49 years old; mean: 33.9, S.D.: 11.17) and socio-cultural level with the patients was also selected.

2.2. Experimental paradigm and data recording

2.2.1. The story

The experimental task was immediate recall of a short story (an English translation is given in Appendix A). In order to make the comprehension and the retrieval processes easy for the schizophrenic patients, the story was composed of simple sentences related to a common daily situation and with no special emotional content.

During the experimental task, subjects were first asked to read the story in a loud voice to ensure their active participation. Immediately after reading, they were asked to recall the story's content. Oral recollections were recorded for further analysis.

2.2.2. Macro-propositions and micro-propositions

Both the story and the transcript of subjects' recollections were analyzed according to a hierarchical model of text comprehension and production (Kintsch and Van Dijk, 1978). The basic elements of the model are propositions defined as the minimal semantic unit on which a false or true assessment is possible. Propositions are composed of a predicate and one or several arguments. They can be described as in the following example:

Peter works in a bank.

"WORKING" is the predicate, with two arguments: "Peter" and "bank". Thus, the above utterance yields the proposition:

WORKING(Peter, bank).

According to Kintsch and Van Dijk (1978), the semantic structure of texts can be described both at the local microlevel and at a more global macrolevel. Propositions can thus be classified according to these two levels: main propositions articulating the topic of the discourse are considered as "macro-propositions" while the other propositions dealing with details are called "micro-propositions".

The hierarchical structure of the story was established by non-psychiatric subjects. In this pre-test, subjects were asked to assess the essential/unessential character of each proposition after reading the entire text. This procedure led to a list of 88 ordered propositions, split into a macro-level (57 macro-propositions) and a micro-level (31 micro-propositions), as illustrated in Appendix B.

2.2.3. Symbolic sequence's generation

Elementary propositions of the text and of the subjects' recall were encoded according to their hierarchical importance: micro-propositions were encoded by symbol *m* and macro-propositions by symbol *M*. In the case of the subjects' recall, the hierarchical structure was only assessed on the basis of the proper structure of the recall, regardless of whether the propositions correctly reflected the text's content or not.

This procedure can be illustrated by the following sentences excerpted from the story:

S1: *Peter is a 25-year-old man.*
S2: *Peter works in a bank.*

According to the pre-test, S_1 belong to the micro-level and S_2 to the macro-level. Using the description based on predicate/argument(s) decomposition, we obtain:

$$S1 = BE(\text{Peter},\ 25\ \text{years}) + BE(\text{Peter},\ \text{man})$$

leading to two micro-propositions and

$$S2 = WORKING(\text{Peter},\ \text{bank})$$

leading to a single macro-proposition. So, the S_1–S_2 sequence is encoded as: "*m m M*".

Each recall was encoded, according to this procedure, into a symbolic sequence composed by successive *M*'s and *m*'s. A recall of *N* propositions thus formed a sequence $S=\{s_i\}$ with $s_i \in \{m,M\}$ for $i=1,...,N$. The symbolic sequence obtained from the text formed our reference data set, and the symbolic sequences obtained from the subjects' recalls formed our experimental data sets.

2.3. Data analysis

From these symbolic sequences, data analysis followed three successive steps:

(1) The number and frequency of each symbol were computed to describe patients' and controls' recalls.
(2) The presence of a temporal structure in the data was tested using surrogate data methods (Kantz and Schreiber, 1997; Schreiber and Schmitz, 2000).
(3) Dynamical processes were studied both at a global level using entropy indices and then at a finer level using transition probabilities between contiguous symbols.

2.3.1. Counting

For each subject, the total number of propositions and the number and frequency of both macro- and micro-propositions were computed.

2.3.2. Surrogate data testing

In order to assess the importance of temporal structure in the data, surrogate data testing was used. The simple hypothesis:

H_0. "data do not depict any temporal structure"

was tested according to the following procedure:

1st step: Compute a complexity index *C* from the original data (i.e., symbolic sequence encoded either from the story or from a subject recall).

2nd step: Generate a set of "surrogate" data according to the tested hypothesis (i.e., surrogate and original data only differ on the tested property).

3rd step: Compute the probability that the complexity index (*C*) of the surrogate data is higher than that of the original data, namely, compute:

$$P_C = \frac{\#\left(C_{sur} > C_{orig}\right)}{\#\text{data}}$$

where # stands for "number of", C_{sur} complexity of surrogate, C_{orig} complexity of original and #data=100 (i.e., 1 original data plus 99 surrogates).

4th step: For a given θ threshold, if $P_C > \theta$, then H_0 can be rejected.

In this study, we tested two complexity indices: metric (*H*) and Lempel–Ziv (*L*) entropies (to be described below). A set of 99 surrogate data was generated according to H_0 by rearranging the original symbolic sequence. Thus, surrogate data share the same number of micro- and macro-propositions with the original data, but they depict a random temporal structure. Counting the number of micro- and macro-propositions thus cannot differentiate between original and surrogate data. The hypothesis of random sequence was tested with a statistical threshold:

$\theta = 0.95$.

2.3.3. Entropy indices

The dynamical complexity of the symbolic sequences was quantified using two entropy indices:

2.3.3.1. Shannon entropy. In the case of a finite sequence of length N, an estimator of Shannon (or metric) entropy is given by:

$$H_n = \frac{-1}{n} \sum_{w_n} Pr_N(w_n) \log_k Pr_N(w_n) \quad (1)$$

where $\mathrm{Pr}_N(w_n)$ denotes the frequency of occurrence of any finite subsequence ('word') of length n (a fully worked example is given in Appendix C).

In order to minimize the finite-length effect in the estimate of $\mathrm{Pr}_N(w_n)$, we used a word length n=3, which fulfills the criterion: $N \geq n \times k^n$ (Xu et al., 1997) for typical sequence length observed in our data ($24 \leq N \leq 88$).

2.3.3.2. Lempel–Ziv entropy. Algorithmic complexity is defined as the length of the minimal program used to reproduce a symbolic sequence. Nevertheless, this value cannot be determined since there is no certainty, in general, that the shortest description has indeed been found (Badii and Politi, 1997). We therefore used a specific description of the symbolic sequence, namely, the Lempel–Ziv encoding algorithm (Kaspar and Schuster, 1987). The general principle of the Lempel–Ziv algorithm is to enumerate new substrings discovered as the sequence evolves from left to right. The number of substrings gives an estimate of the sequence complexity (see Appendix C for an example).

In the case of short symbolic sequences, L is normalized by the maximal value obtained for random sequences with the same length (N) and number of symbols (here k=2) (Rapp and Schmah, 1996).

2.3.4. Transition matrix

Symbolic sequences were also described as a Markov process on the basis of matrix of probabilities quantifying transitions between symbols. In our case, a 2×2 matrix (A) of transition probabilities, defined as:

$$A = \begin{bmatrix} Pr(m \to m) & Pr(m \to M) \\ Pr(M \to m) & Pr(M \to M) \end{bmatrix}$$

(where $Pr(s \to s')$ denotes the conditional transition probabilities from s to s' with s and $s' \in \{m, M\}$) was built from the original data (see Appendix C for a complete example).

3. Results

3.1. Counting

We recall the characteristics of the story: 88 propositions including 51 (i.e., 57.96%) macro-propositions and 37 (i.e., 42.04%) micro-propositions.

Schizophrenic patients and controls did differed neither for the overall number of recalled propositions (patients: mean=39, S.D.=8.53; controls: mean=40.01, S.D.=11.91; Wilcoxon test, W=24, P=0.39) nor for the frequency of macro-propositions (patients: mean=0.75, S.D.=0.09; controls: mean=0.76, S.D.=0.08; Wilcoxon test, W=25, P=0.42. See: Table 1 and Fig. 1).

The frequency of macro-propositions increases for all the subjects (but one) compared with the story. The macro-proposition production is thus facilitated during the recall.

3.2. Presence of temporal structure

For the story, Lempel–Ziv entropy and metric entropy, respectively, were L=0.53 and H=0.68. These values are significantly (P=1%) lower than the entropy obtained for a set of 99 shuffled surrogate data (Fig. 2) and thus point to a significant non-random temporal structure in the original data.

Surrogate data testing was computed for each subject in each group (the individual results are given in Table 2). For the statistical threshold θ=0.95, the

Table 1
Results for schizophrenic and matched control subjects: sequence length (N), number and percentage of macro-propositions (resp. $\#M$ and $\%M$), number and percentage of micro-propositions (resp. $\#m$ and $\%m$)

	Schizophrenic patients						Matched controls				
	$\#M$	$\#m$	N	$\%M$	$\%m$		$\#M$	$\#m$	N	$\%M$	$\%m$
S1	37	7	44	84.1	15.9	T1	31	14	45	68.9	31.1
S2	31	6	37	83.8	16.2	T2	30	7	37	81.1	18.9
S3	27	7	34	79.4	20.6	T3	17	7	24	70.8	29.2
S4	30	14	44	68.2	31.8	T4	36	10	46	78.3	21.7
S5	37	17	54	68.5	31.5	T5	23	16	39	59.0	41.0
S6	24	8	32	75.0	25.0	T6	31	4	35	88.6	11.4
S7	25	9	34	73.5	26.5	T7	45	17	62	72.6	27.4
S8	2	10	38	73.7	26.3	T8	26	10	36	72.2	27.8
S9	14	11	25	56.0	44.0	T9	20	4	24	83.3	16.7
S10	41	7	48	85.4	14.6	T10	45	8	53	84.9	15.1
Mean	29.4	9.6	39.0	74.8	25.2	Mean	30.4	9.7	40.1	75.9	24.1
S.D.	7.8	3.5	8.53	9.0	9.0	S.D.	9.6	4.6	11.9	8.9	8.9

S.D.: standard deviation.

null hypothesis (i.e., absence of temporal structure) was significantly rejected in 8/10 schizophrenics and 8/10 controls when metric entropy was used as a complexity index and in 5/10 schizophrenics and 4/10 controls for Lempel–Ziv entropy. These results point to the presence of an underlying temporal structure in the data set. Thus, counting occurrences of propositions should be completed with the dynamical properties in further computation.

3.3. Dynamical characterization

3.3.1. Entropy indices: global complexity

The sequences generated by schizophrenic patients and controls did not differ for both metric entropy (patients: mean=0.61, S.D.=0.16; controls: mean=0.64, S.D.=0.14; Wilcoxon test, W=22, P=0.313) and Lempel–Ziv entropy (patients: mean=0.54, S.D.=0.14; controls: mean=0.59, S.D.=0.08; Wilcoxon test, W=22, P=0.313). See Fig. 3 and Table 2.

Since raw number of macro- or micro-propositions could covariate with entropy indices, an analysis of covariance (ANCOVA) was performed. Normality and homogeneity of group variances were assessed for the distribution of Lempel–Ziv entropy (Lilliefors test: L=0.1051, P>0.20; Fisher test: $F_{9,9}$=2.794, P=0.07) and for Shannon entropy (Lilliefors test: L=0.0957, P>0.20; Fisher test: $F_{9,9}$=1.236, P=038). Only one significant negative correlation was observed between Lempel–Ziv entropy and the raw number of macro-propositions (r=−0.68, T_8=2.66, P=0.03). Nevertheless, no significant differences appeared between patients and controls when the number of macro-propositions was used as a covariate variable ($F_{1,17}$=2.62, P=0.12).

These results thus show that the global complexity of recall did not differ between the schizophrenic and control groups.

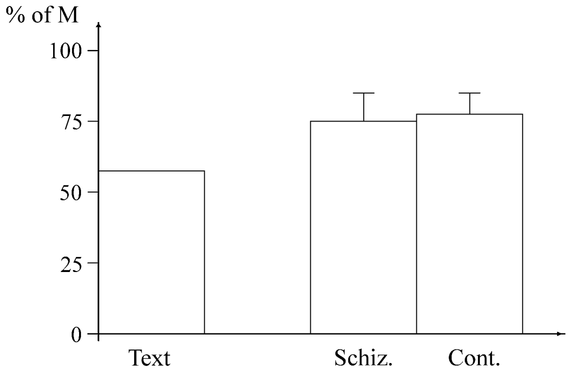

Fig. 1. Percentage of macro-propositions (% of M) for the story (Text) and for the recall in control (Cont.) and schizophrenic (Schiz.) groups. Vertical bars depict standard deviations.

3.3.2. Transition matrix

The averaged results are given in Table 3. Transitions from micro- to micro-propositions ($m \rightarrow m$ transitions) were significantly more frequent for patients than for controls (Wilcoxon test, W=8,

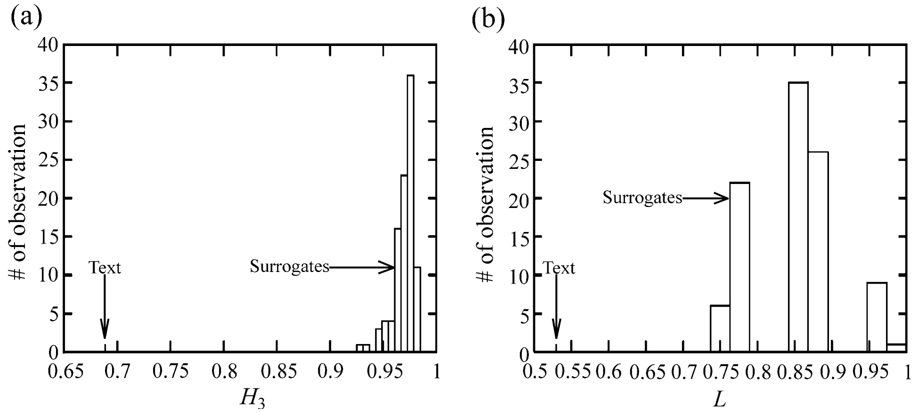

Fig. 2. Entropy for the story (Text) compared with the distribution obtained from a set of surrogate data. (a) Metric entropy, (b) Lempel–Ziv entropy.

P=0.024) and thus[1] $m \rightarrow M$ transitions were less frequent for schizophrenics. Probabilities of transitions starting from M did not differ between the two groups (Wilcoxon test, W=19, P=0.216).

Since assumptions of normality and homogeneity of group variances were not all ascertained for the transition's distribution,[2] no ANCOVA was performed for the transition matrix.

4. Discussion

Our analysis of dynamical properties of speech shows that significant temporal structure is observed in the data. Taking this structure into account, we observed no difference between patients and controls for global complexity, but a specific organization in the transition between propositions: schizophrenic patients connect micro-propositions more often than control subjects.

4.1. Methodological issues

The central concern of our study is the temporal organization of speech. The results obtained using surrogate data testing justify its validity. Indeed, they demonstrate the presence of a temporal organization in the succession of micro- and macro-propositions that is significantly different from that of random sequences. This observation was obtained for the story and for the experimental

Table 2
Results for schizophrenic and matched control subjects: metric entropy (H), Lempel–Ziv entropy (L) and the probability for the entropy of the surrogate data to be strictly greater than the entropy of the experimental data (PH for metric and PL for Lempel–Ziv entropy)

	Schizophrenics					Matched controls			
	H	P_H	L	P_L		H	P_H	L	P_L
S1	0.41	**0.99**	0.42	**0.99**	T1	0.75	**0.99**	0.58	**0.97**
S2	0.38	**0.99**	0.45	**0.95**	T2	0.53	**0.99**	0.67	0.30
S3	0.59	**0.97**	0.64	0.22	T3	0.78	0.77	0.78	0.11
S4	0.69	**0.99**	0.50	**0.98**	T4	0.59	0.99	0.50	**0.98**
S5	0.66	**0.99**	0.33	**0.99**	T5	0.89	**0.97**	0.67	0.92
S6	0.72	0.90	0.60	0.80	T6	0.51	0.29	0.54	0.17
S7	0.72	**0.99**	0.70	0.47	T7	0.59	**0.99**	0.53	**0.98**
S8	0.62	**0.99**	0.55	**0.96**	T8	0.76	**0.95**	0.55	**0.97**
S9	0.88	0.88	0.78	0.32	T9	0.49	**0.97**	0.56	0.30
S10	0.44	**0.99**	0.46	0.76	T10	0.50	**0.99**	0.57	0.19
Mean	0.61	0.97	0.54	0.74	Mean	0.64	0.89	0.59	0.59
S.D.	0.14	0.04	0.14	0.29	S.D.	0.13	0.20	0.08	0.40

Bold-face font corresponds to significant rejection of H_0.

[1] Since, $\sum_{i \in \{m,M\}} P(m \rightarrow i)=1$.

[2] Normality was checked for transitions (Lilliefors test: L=0.1574, P>0.20), but homogeneity of group variances was violated (Fisher test: $F_{9,9}$=3.1409, P=0.0522). Homogeneity of group variances was confirmed for $M \rightarrow m$ transitions (Fisher test: $F_{9,9}$=1.3, P=0.3514), but normality was not verified (Lilliefors test: L=0.1946, P<0.05).

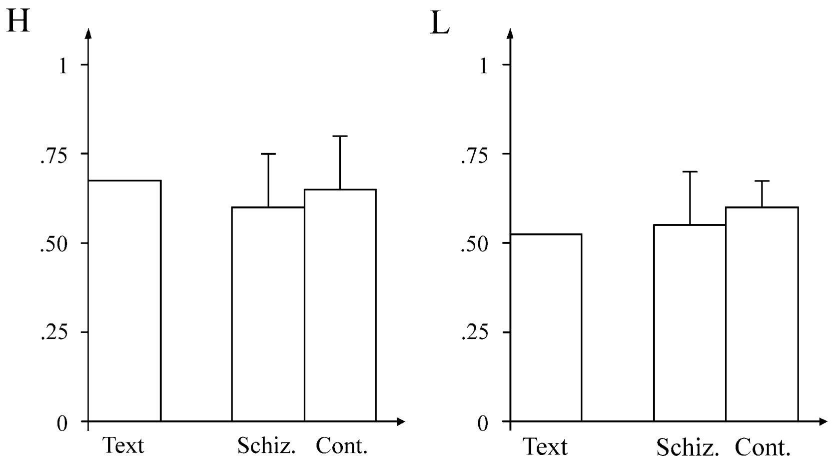

Fig. 3. Averaged values of metric (H) and Lempel–Ziv (L) entropies for the story (Text), schizophrenic (Schiz.) and control (Cont.) groups. Vertical bars depict standard deviations.

data sets, whatever the group of subjects. These results thus emphasize that:

(1) temporal organization is a significant feature of speech,
(2) counting is not sufficient for an adequate characterization of language, and
(3) symbolic dynamical methods are needed for the sake of completeness.

We thus used indices of dynamical complexity that are different from those based on the hierarchical structure of sentences. The definition of language complexity on the basis of the number of embedded clauses (e.g., DeLisi et al., 1997) clearly overlooks the temporal organization of speech.

The complexity values obtained for the story and for the recalls are intermediate between those obtained for periodic sequences (H=0) and for random sequences (H=1). They are thus in accordance with the intuitive understanding of a story or discourse complexity: between periodic and random. Nevertheless, those measurements were not able to differentiate between schizophrenic and control subjects. Such a negative result may be due to the excessively short length of the symbolic sequences.[3] Different encoding procedures or other experimental situations that can generate longer symbolic sequences are thus clearly needed to overcome this drawback.

4.2. Schizophrenic speech

Our results show that patients with schizophrenia hardly access the highest level of the hierarchical structure (macro-propositions) of discourse once they have accessed the lowest level (micro-propositions). They thus display a dynamical trend to connect micro-propositions one after the other. Nevertheless, the connections between macro-propositions and either micro- or macro-propositions are similar in controls' and patients' discourse. Moreover, the same numbers of macro- and micro-propositions were observed in both groups.

Macro-propositions correspond to the macro-structure of discourse and can reveal the discourse plan hypothesized in language-production models (Levelt, 1989). Since the number of macro-propositions and

Table 3
Transition probabilities for the story and averaged transition probabilities for schizophrenic patients and control subjects

	Text	Patients	Controls
$Pr(m \rightarrow m)$	0.84	**0.64** (0.08)	**0.53** (0.14)
$Pr(m \rightarrow M)$	0.16	**0.36** (0.08)	**0.47** (0.14)
$Pr(M \rightarrow M)$	0.90	0.91 (0.06)	0.88 (0.06)
$Pr(M \rightarrow m)$	0.10	0.09 (0.06)	0.12 (0.06)

Bold-face fonts represent significant differences between groups. Standard deviations are given in parentheses.

[3] The low number of rejections of H_0 for Lempel–Ziv entropy is also due to such a limitation.

the probability of $M \rightarrow M$ transitions are similar in both groups, schizophrenic patients did not display any deficit in generating the overall structure of a discourse plan. Nevertheless, the high probability of $m \rightarrow m$ transitions in schizophrenic patients can be associated with an impairment in the ability to inhibit nonessential responses (Servan-Schreiber et al., 1996) and thus with an inability to maintain the discourse plan (Barch and Berenbaum, 1997). This overactivation of the discourse micro-structure is in accordance with clinical observations of formal thought disorders and discourse-coherence disturbance in schizophrenia (Berenbaum and Barch, 1995).

The macro-structure of discourse acts as a constraint on its micro-structure, by means of macro-rules such as deletion, generalization and construction (Kintsch and Van Dijk, 1978). In order to control discourse at its lowest level, macro-structure has to be actively held in working memory. Moreover, an appropriate response can only be produced if context information such as task instruction or response content is also maintained in working memory (Servan-Schreiber et al., 1996). Since all our subjects correctly performed the overall task, our results only deal with the content aspect of context information. The impairments observed here in maintaining the discourse plan in schizophrenic patients are thus in accordance with the schizophrenic deficits in attention, working and episodic memory, and executive functions (Braver et al., 1999).

Finally, low verbosity has been associated with a deficit in generating a discourse plan and is correlated with negative thought disorders in schizophrenia (Barch and Berenbaum, 1997). Since, schizophrenic subjects were selected on the basis of their ability to produce the minimal number of propositions compatible with dynamical tools efficiency, our inclusion criteria represented a selection bias towards linguistically skilled or disinhibited subjects. The present study needs to be extended to avoid such selection bias. In fact, the selection of patients according to the PANSS criteria could permit a specific test of the difference between generating and maintaining a discourse plan in schizophrenia. Further studies should be conducted in an experimental setting where spontaneous verbal production occurs more frequently to avoid biases due to negative symptoms and short symbolic sequences.

4.3. Conclusion

Using symbolic methods to analyze the dynamics of speech in schizophrenia, we showed that schizophrenic patients persist in the production of micro-propositions and thus show a deficit in maintaining rather than in generating a discourse plan. From a methodological point of view, this study emphasizes the validity of quantifying the dynamical properties of speech with symbolic methods. From a clinical point of view, it can be proposed that communication with schizophrenic patients could be improved by structuring the discourse interaction (Barch and Berenbaum, 1997). For example, the production of a macro-proposition or a simple reformulation of the patient's essential idea could allow one to avoid tangential responses or loss of goal.

Acknowledgments

We thank G. Quehen for sharing the data with us and for her help in the encoding process. The careful readings by D. Papo and A. Lesne of a previous version of this manuscript are greatly appreciated. We are obliged to three anonymous reviewers for the significant improvement of our initial manuscript.

Appendix A. The story: "Peter and John"

Peter is a young man, 25 years of age and of Italian extraction. He works in an important Parisian bank and lives in the suburbs. Every morning, Peter takes the train to go to his office and has his breakfast at the pub in the railway-station.

Once, Peter met John there. They have not met for 18 years. They talked about childhood memories. They used to live in the Dordogne, in the same village. Peter went to Paris for his studies and found a job. Afterward, they didn't meet again.

Suddenly, Peter realizes that he is late. He apologizes and proposes another appointment. When standing up, Peter bumps against the waiter who drops his tray. Peter's suit is then stained with chocolate. There is no possibility of his going to work with such dirty clothes.

Table B1
Arguments included in the propositions from the first paragraph

	Px1	*Px2*	*Px3*	*Px4*	*Px5*	*Px6*	*Px7*	*Px8*	*Px9*	*Px10*
Argument	Man	Years	Extraction	Bank	Suburbs	Morning	Train	Office	Pub	Station

Peter goes back home to change clothes. The manager of the bank, worrying about his lateness, phones and asks him to hurry. Peter needs to meet with an important client for a loan. He takes a taxi to be there faster. Despite his delay, Peter receives the client and grants the credit.

Appendix B. Examples of macro- and micro-propositions

We present here the propositions from the first paragraph as an illustrative example. Table B1 gives the list of arguments that appear in the propositions. Table B2 depicts the complete list of propositions and their decomposition into predicate and arguments.

Table B2
List of propositions from the first paragraph

	Proposition	Status
P1	BE CALLED(*Px1,Px'1=Peter*)	M
P2	YOUNG(*Px1*)	M
P3	NUMBER: TWENTY-FIVE(*Px2*)	M
P4	AGE(*P2,P3*)	M
P5	ITALIAN(*Px3*)	M
P6	HAVE(*Px1,P5*)	M
P7	WORK(*Px1*)	M
P8	LOCATION: IN(*P7,Px4*)	m
P9	IMPORTANT(*Px4*)	m
P10	PARISIAN(*Px4*)	M
P11	AND(*P7,P12*)	M
P12	LIVE(*Px1*)	m
P13	LOCATION: IN(*P12, Px5*)	m
P14	TAKE(*Px1,Px7*)	m
P15	TIME: EVERY(*P15,Px6*)	m
P16	GO TO(*Px1,Px8*)	m
P17	AND(*P15,P19*)	m
P18	BREAKFAST(*Px1*)	m
P19	LOCATION: AT(*P19, Px9*)	m
P20	OF(*Px9,Px10*)	M

M denotes a macro-proposition and *m* a micro-proposition.

Appendix C. Computation of entropies and transition matrix

According to the encoding of the propositions from the first paragraph presented in Appendix B, we can extract the following symbolic sequence:

$$S = \mathrm{M\,M\,M\,M\,M\,M\,M\,m\,m\,M\,M\,m\,m\,m\,m\,m\,m\,m\,m\,M}$$

This sequence, S, written with an alphabet of k=2 symbols and of length N=20, will be considered as a data set to illustrate the procedures for computing Shannon and Lempel–Ziv entropies and the matrix of transition probabilities.

C.1. Shannon entropy

We repeat the definition of Shannon (or metric) entropy (Eq. (1)):

$$H_n = \frac{-1}{n} \sum_{w_n} Pr_N(w_n) \log_k Pr_N(w_n) \qquad \text{(C1)}$$

where $\Pr_N(w_n)$ denotes the frequency of occurrence of any 'word' of length n in the sequence. We used n=3 in our data analysis to fulfill the criterion: $N \geq nk^n$.

Table C1 presents the 'words' w_n of length n=3 in S (the words *mMm* and *MmM* do not appear in S), their number of occurrences ($\#(w_n)$) and their

Table C1
Words w_n of length 3 in S, their number of occurrences ($\#(w_n)$) and their frequencies ($\Pr_N(w_n)$)

w_n	$\#(w_n)$	$\Pr_N(w_n)$
mmm	6	0.333
mmM	2	0.222
mMM	1	0.056
MMm	2	0.111
Mmm	2	0.111
MMM	5	0.278

frequency ($Pr_N(w_n)$) in S. The frequency is defined by:

$$Pr_N(w_n) = \frac{\#(w_n)}{\sum \#(w_n)} \quad \text{(C2)}$$

where $\sum\#(w_n)$ is the total number of 'words' of length n in S, i.e., $\sum\#(w_n)=N-(n-1)$; here $\sum\#(w_n)=18$.

We then obtain Shannon entropy of S on the basis of Eq. (D.1) (using $\log_2(x)$):

$$H_3(S) = -\frac{1}{3} \times (-2.33) \approx 0.78 \quad \text{(C3)}$$

C.2. Lempel–Ziv entropy

The general principle of the Lempel–Ziv algorithm is to enumerate new substrings discovered as the sequence evolves from left to right (Kaspar and Schuster, 1987). Using sequence S, new substrings can be delimited by dots, as follows:

$$S = \mathtt{M{\cdot}MM{\cdot}MMM{\cdot}Mm{\cdot}m{\cdot}MMm{\cdot}mm{\cdot}mmm{\cdot}mmM{\cdot}}$$

The number of substrings gives the value of the Lempel–Ziv complexity L; here, $L(S)=9$.

C.3. Transition matrix

In the present study, the transition matrix A is defined as:

$$A = \begin{bmatrix} Pr(m{\rightarrow}m) & Pr(m{\rightarrow}M) \\ Pr(M{\rightarrow}m) & Pr(M{\rightarrow}M) \end{bmatrix}$$

We compute here A for the sequence S given above.

$Pr(m{\rightarrow}m)$ is the probability to obtain an m at time $t+1$ knowing that an m was produced at time t. Its value is computed first by counting the total number of couples starting with an m, i.e., either mm or mM. We obtain here 10 such pairs. Then, the number of mm is observed; here this pair appears eight times. So, we can conclude that:

$$Pr(m{\rightarrow}m) = \frac{8}{10} \approx 0.8$$

Then, mM appears twice, thus:

$$Pr(m{\rightarrow}M) = \frac{2}{10} \approx 0.2$$

Thus, by definition: $Pr(m{\rightarrow}m)+Pr(m{\rightarrow}M)=1$.

It is important to notice that the probability to obtain mm in the whole sequence S is equal to 0.42 (=8/19), which is generally different from the transition probability (here 0.8=8/10).

The same procedure applied to M gives: $Pr(M{\rightarrow}m)=2/9\approx0.23$ and $Pr(M{\rightarrow}M)=7/9\approx0.77$.

References

Andreasen, N., 1979a. Thought, language and communication disorders: 1. Clinical assessment, definition of terms, and evaluation of their reliability. Archives of General Psychiatry 36, 1315–1321.

Andreasen, N., 1979b. Thought, language and communication disorders: 2. Diagnostic significance. Archives of General Psychiatry 36, 1322–1334.

Badii, R., Politi, A., 1997. Complexity. Hierarchical Structures and Scaling in Physics. Cambridge University Press, Cambridge, UK.

Barch, D.M., Berenbaum, H., 1996. Language production and thought disorders in schizophrenia. Journal of Abnormal Psychology 105 (1), 81–88.

Barch, D., Berenbaum, H., 1997. The effect of language production manipulations on negative thought disorder and discourse coherence disturbances in schizophrenia. Psychiatry Research 71, 115–127.

Bazin, N., Perruchet, P., Hardy-Bayle, M.-C., Feline, A., 2000. Contextdependent information processing in patients with schizophrenia. Schizophrenia Research 45, 93–101.

Berenbaum, H., Barch, D., 1995. The cattegorization of thought disorder. Journal of Psycholinguistic Research 24, 349–376.

Braver, T., Barch, D., Cohen, J., 1999. Cognition and control in schizophrenia: a computational model of dopamine and prefrontal function. Biological Psychiatry 46 (3), 312–328.

Chapman, L., Chapman, J., Dant, R., 1976. Schizophrenic inability to disattend from strong aspects of meaning. Journal of Abnormal Psychology 85, 35–40.

Cohen, J., Servan-Schreiber, D., 1992. Context, cortex, and dopamine: a connectionist approach to behavior and biology in schizophrenia. Psychological Review 99 (1), 45–77.

Cohen, J.D., Barch, D., Carter, C., Servan-Schreiber, D., 1999. Context-processing deficits in schizophrenia: converging evidence from three theoretically motivated cognitive tasks. Journal of Abnormal Psychology 108 (1), 120–133.

Condray, R., Steinhauer, S.R., van Kammen, D.P., Kasparek, A., 1996. Working memory capacity predicts language comprehension in schizophrenic patients. Schizophrenia Research 20, 1–13.

Copland, D., Chenery, H., Savage, G., McGrath, J., 2002. An online investigation of lexical ambiguity processing in schizophrenia. Brain and Cognition 48 (2–3), 324–327.

DeLisi, L., Sakuma, M., Kushner, M., Finer, D., Hoff, A., Crow, T., 1997. Anomalous cerebral asymmetry and language processing in schizophrenia. Schizophrenia Bulletin 23 (2), 255–271.

Docherty, N., Hawkins, K., Hoffman, R., Quinlan, D., Rakfeldt, J., Sledge, W., 1996. Working memory, attention, and communication disturbances in schizophrenia. Journal of Abnormal Psychology 105 (2), 212–219.

Docherty, N., Hall, M., Gordinier, S., Cutting, L., 2000. Conceptual sequencing and disordered speech in schizophrenia. Schizophrenia Bulletin 26 (3), 723–735.

Ehlers, C., 1995. Chaos and complexity. Can it help us to understnad mood and behavior? Archives of General Psychiatry 52 (11), 960–964.

Elbert, T., Lutzenberger, W., Rockstroh, B., Berg, P., Cohen, R., 1992. Physical aspects of the EEG in schizophrenics. Biological Psychiatry 32 (7), 595–606.

Elman, J.L., 1995. Language as a dynamical system. In: Port, R., Van Gelder, T. (Eds.), Mind as motion: Explorations in the Dynamics of Cognition. MIT Press, Cambridge, MA, pp. 195–225.

Frith, C., 1992. The Cognitive Neuropsychology of Schizophrenia. Lawrence Erlbaum Associates, Hillsdale, NJ.

Frith, C., Blakemore, S., Wolpert, D., 2000. Explaining the symptoms of schizophrenia: abnormalities in the awareness of action. Brain Research Reviews 31 (2–3), 357–363.

Globus, G.G., Arpaia, J.P., 1994. Psychiatry and the new dynamics. Biological Psychiatry 35 (5), 352–364.

Goldberg, T.E., Aloia, M.S., Gourovitch, M.L., Missar, D., Pickar, D., Weinberger, D.R., 1998. Cognitive substrates of thought disorder, i: the semantic system. American Journal of Psychiatry 155 (12), 1671–1676.

Harrow, M., Lanin-Kettering, I., Prosen, M., Miller, J., 1983. Disordered thinking in schizophrenia: intermingling and loss of set. Schizophrenia Bulletin 9 (3), 354–367.

Hoffman, R., Stopek, S., Andreasen, N., 1986. A comparative study of manic vs. schizophrenic speech disorganization. Archives of General Psychiatry 43 (9), 831–838.

Jeong, J., Kim, D.J., Chae, J.H., Kim, S.Y., Ko, H.J., Paik, I., 1998. Nonlinear analysis of the EEG of schizophrenics with optimal embedding dimension. Medical Engineering & Physics 20 (9), 669–676.

Kantz, H., Schreiber, T., 1997. Nonlinear Time Series Analysis. Cambridge University Press, Cambridge, UK.

Kaplan, D., Glass, L., 1995. Understanding Nonlinear Dynamics. Springer-Verlag, New York.

Kaspar, F., Schuster, H.G., 1987. Easily calculable measure for the complexity of spatiotemporal patterns. Physical Review A 36, 842–848.

Kay, S.R., Fiszbein, A., Opler, L.A., 1987. The Positive and Negative Syndrome Scale (PANSS) for schizophrenia. Schizophrenia Bulletin 13 (2), 261–276.

Kim, D.J., Jeong, J., Chae, J.H., Park, S., Yong Kim, S., Jin Go, H., Paik, I.H., Kim, K.S., Choi, B., 2000. An estimation of the first positive Lyapunov exponent of the EEG in patients with schizophrenia. Psychiatry Research 98 (3), 177–189.

King, K., Fraser, W.I., Thomas, P., Kendell, R.E., 1990. Re-examination of the language of psychotic subjetcs. British Journal of Psychiatry 156, 211–215.

Kintsch, W., Van Dijk, T.A., 1978. Toward a model of text comprehension and production. Psychological Review 85 (5), 363–394.

Kowalik, Z.J., Schiepek, G., Kumpf, K., Roberts, L.E., Elbert, T., 1997. Psychotherapy as a chaotic process. II. The application of nonlinear analysis methods on quasi time series of the client–therapist interaction: a nonstationary approach. Psychotherapy Research 7 (3), 197–218.

Kwapil, T., Hegley, D., Chapman, L., Chapman, J., 1990. Facilitation of word recognition by semantic priming in schizophrenia. Journal of Abnormal Psychology 99 (3), 215–221.

Levelt, W., 1989. Speaking: From Intention to Comprehension. MIT Press, Cambridge, MA.

McCarley, R., Niznikiewicz, M., Salisbury, D., Nestor, P., O'Donnell, B., Hirayasu, Y., Grunze, H., Greene, R., Shenton, M., 1999. Cognitive dysfunction in schizophrenia: unifying basic research and clinical aspects. European Archives of Psychiatry and Clinical Neuroscience 249 (Suppl. 4), IV69–IV82.

Mesure, G., Passerieux, C., Besche, C., Widlöcher, D., Hardy-Baylé, M.C., 1998. Impairment of semantic categorization processes among thought-disordered schizophrenic patients. Canadian Journal of Psychiatry 43 (3), 271–278.

Minzenberg, M., Ober, B., Vinogradov, S., 2002. Semantic priming in schizophrenia: a review and synthesis. Journal of the International Neuropsychological Society 8 (5), 699–720.

Morice, R.D., Ingram, J.C.L., 1982. Language analysis in schizophrenia: diagnostic implications. Australian and New Zealand Journal of Psychiatry 16, 11–21.

Nandrino, J.-L., Leroy, F., Pezard, L., in press. Dynamics as a heuristic framework for psychopathology.

Paulsen, J.S., Romero, R., Chan, A., Davis, A.V., Heaton, R.K., Jeste, D.V., 1996. Impairment of the semantic network in schizophrenia. Psychiatry Research 63 (2–3), 109–121.

Paulus, M.P., Geyer, M.A., Braff, D.L., 1996. Use of methods from chaos theory to quantify a fundamental dysfunction in the behavioral organization of schizophrenic patients. American Journal of Psychiatry 153, 714–717.

Plagnol, A., Pachoud, B., Claudel, B., Granger, B., 1996. Functional disorganization of representations in schizophrenia. Schizophrenia Bulletin 22, 383–398.

Port, R., Van Gelder, T., 1995. Mind as Motion: Explorations in the Dynamics of Cognition. MIT Press, Cambridge, MA.

Rapp, P.E., Schmah, T.I., 1996. Complexity measures in molecular psychiatry. Molecular Psychiatry 1, 408–416.

Rapp, P.E., Jimenez-Montano, M.A., Langs, R.J., Thomson, L., Mees, A.I., 1991. Toward a quantitative characterization of patient–therapist communication. Mathematical Biosciences 105 (2), 207–227.

Rochester, S.R., Martin, J., 1979. Crazy Talk: A Study of the Discourse of Schizophrenic Speakers. Plenum Press, New York.

Röschke, J., Aldenhoff, J., 1993. Estimation of the dimensionality of sleep EEG data in schizophrenics. European Archives of Psychiatry and Clinical Neuroscience 242 (4), 191–196.

Röschke, J., Mann, K., Fell, J., 1994. Nonlinear EEG dynamics during sleep in depression and schizophrenia. International Journal of Neuroscience 75, 271–284.

Röschke, J., Fell, J., Beckmann, P., 1995. Nonlinear analysis of sleep EEG data in schizophrenia: calculation of the principal Lyapunov exponent. Psychiatry Research 56, 257–269.

Salamé, P., Danion, J.M., Peretti, S., Cuervo, C., 1998. The state of functioning of working memory in schizophrenia. Schizophrenia Research 30, 11–29.

Salisbury, D., Shenton, M., Nestor, P., McCarley, R., 2002. Semantic bias, homograph comprehension, and event-related potentials in schizophrenia. Clinical Neurophysiology 113 (3), 383–395.

Schiepek, G., Kowalik, Z.J., Schütz, A., Köhler, M., Richter, K., Strunk, G., Mühlnickel, W., Elbert, T., 1997. Psychotherapy as a chaotic process I. Coding the client–therapist interaction by means of sequential plan analysis and the search for chaos: a stationary approach. Psychotherapy Research 7 (2), 173–194.

Schreiber, T., Schmitz, A., 2000. Surrogate time series. Physica. D 142, 346–382.

Servan-Schreiber, D., Cohen, J., Steingard, S., 1996. Schizophrenic deficits in the processing of context. Archives of General Psychiatry 53, 1105–1112.

Sitnikova, T., Salisbury, D., Kuperberg, G., Holcomb, P., 2002. Electrophysiological insights into language processing in schizophrenia. Psychophysiology 39 (6), 851–860.

Sullivan, E., Shear, P., Zipursky, R., Sagar, H., Pfefferbaum, A., 1997. Patterns of content, contextual, and working memory impairments in schizophrenia and nonamnesic alcoholism. Neuropsychology 11 (2), 195–206.

Thomas, P., King, K., Fraser, W.I., Kendel, R.E., 1990. Linguistic performance in schizophrenia: a comparison of acute and chronic patients. British Journal of Psychiatry 156, 204–210.

Titone, D., Levy, D.L., Holzman, P.S., 2000. Contextual insensitivity in schizopnrenic language processing: evidence from lexical ambiguity. Journal of Abnormal Psychology 109 (4), 761–767.

Tschacher, W., Scheier, C., Hashimoto, Y., 1997. Dynamical analysis of schizophrenia courses. Biological Psychiatry 41 (4), 428–437.

Xu, J., Liu, Z.R., Liu, R., Yang, Q.F., 1997. Information transmission in human cerebral cortex. Physica. D 107, 363–374.

Zellner Keller, B., Keller, E., 2000. The chaotic nature of speech rhythm: hints for fluency in the language acquisition process. In: Delcloque, P., Holland, V. (Eds.), Integrating Speech Technology in Language Learning. Swets & Zeitlinger, Amsterdam, The Netherlands.