Psychology of Language and Communication 2012, Vol. 16, No. 2

DOI: 10.2478/v10057-012-0012-x

GIUSEPPE LEONARDI
Nicolaus Copernicus University, Toruń

THE STUDY OF LANGUAGE AND CONVERSATION WITH RECURRENCE ANALYSIS METHODS

In the last decade we witness an increase in approaching issues in language, and more generally, cognition, from a dynamical standpoint. This theoretical shift necessitates new research methods and statistical / analytical tools. Some of these tools gain popularity and are being applied to language in many of its multifaceted perspectives. Recurrence analysis is one of those methods. Its relative simplicity of application and quite unconstrained statistical assumptions give researchers an insight into the dynamical nature of the phenomena under scrutiny. The aim of this paper is an introduction to this method, a review of its convincing applications in the language research on several levels of language analysis and finally, a reflection on its possible further uses.

Key words: recurrence analysis, recurrence plots, conversation analysis, dynamical systems, methods of analysis in linguistic research

Introduction

In the latest years, a new theoretical framework seems to emerge in the behavioral sciences, fueled by the difficulties that the information processing approach (based on the computer metaphor of the brain) manifests when modeling the ongoing activity of human agents in their natural environment. In this new framework human organism is seen as a complex system, continuously interacting and dynamically coupled with the environment (see e.g. Kelso, 1995; Van Orden et al., 2003). This kind of systems in physics and the natural sciences have been approached by means of concepts and methods from the dynamical system theory, a mathematical tool designed to describe the behavior of systems composed of many non-linearly interacting parts. Concepts relating to stability and emergence of behavior and in-

Address for correspondence: Giuseppe Leonardi, Institute of Philosophy, Nicolaus Copernicus University, Fosa Staromiejska 1a, 87-100 Toruń, Poland. E-mail: giuseppe.leonardi@umk.pl

teraction of elements at different time scales allow to model and better understand or explain physical and biological phenomena (Haken, 1983; Bassingthwaighte, Liebovitch & West, 1994), promising to be equally useful in the behavioral and brain sciences (Kelso, 1995; Thelen & Smith, 1994; Port & Van Gelder, 1995).

One thing that is central to this view is that such phenomena unfold in time, and hence time dependent models have to be built and time dependent data have to be recorded and appropriately analyzed. In the case of language this time dimension is certainly a vital component. The flow of a conversation with its sequence of turn taking, the series of words in a text/speech or the order of the different syntactic structures are only few examples of time dependent characteristics of linguistic phenomena.

In psychology it is often the case that researchers try to capture human behaviors by averaging measures across trials and across subjects. The assumption is that the mean value is the best estimator of behavior to be measured, while variability indicates noise due to uncontrolled factors and to the measuring process itself. These data are usually analyzed through statistical methods which make several assumptions such as linearity, stationarity, normality of distribution, homogeneity of variance etc. While this type of research strategy certainly has produced many important results, still it is important to recognize how behaviors unfold in time and that the different ways in which they do it cannot be simply discarded as variability and noise (Van Orden et al., 2003; Van Orden et al., 2005). Especially for the behaviors such as language, where the time dimension is a crucial component, understanding the underlying mechanisms requires the analysis of the ordered, real time unfolding of behavior, not just their outcome. Paradoxically, studying just outcomes of behavior puts modern cognitive psychology in line with behaviorism, which it tried to oppose (Costall, 2004).

It may be fruitful and it is actually an already widely tested approach (Tschacher & Dauwalder, 2003) to conceive human behavior as the product of a complex system obeying dynamical laws to be identified and modeled. And, since from this perspective the measurement of the dynamical time series of real behavior becomes crucial, we also need new tools which appropriately capture the complexity and time dependent nature of these behaviors. One of these tools is the Recurrence Quantification Analysis (RQA), first introduced by Zbilut and Webber (1992; Webber & Zbilut, 1994) who built upon the work on recurrence plots of Eckman et al. (1987). The aim of this paper is to introduce the main concepts of RQA, and to illustrate a few examples of how this technique can be fruitfully applied to the study of language and communication.

Recurrence in nature

One of the crucial characteristics of a dynamical system is a patterning of its behavior in time:

> "Insofar as natural patterns are found in all dynamical systems, the degree to which those systems exhibit recurrent patterns speaks volumes regarding their underlying dynamics."
> Webber & Zbilut, 2005, p. 27

In physics and physiology recurrences, i.e. a periodical coming back to the same or very similar states of behavior, may involve simple patterns, such as waves and tides going up and down regularly, the planets turning around themselves, the heart muscles continuously contracting and relaxing or the neurons firing at certain frequencies and so on. In language, repetition of words, particular grammatical structures or concepts in a text or discourse, the pauses and the turns in a conversation can also be seen as manifestations of recurrent behavior. Simple systems may have a quite regular and straightforward pattern of recurrence while with increasing complexity recurrences become less frequent and not easily predictable, but can still be registered and ascribed to regions of stability in a dynamical system, technically called the attractors of the system.

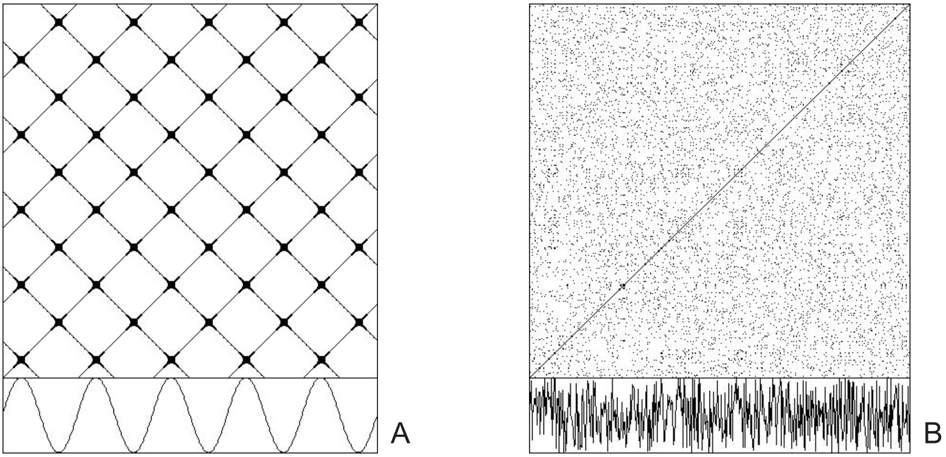
Although there are already statistical methods for analyzing this recurrent behavior, by modeling it in linear terms (e.g. the additive based Fourier transform which considers a time series as the sum of many independent processes at different frequencies, see e.g. Burrus, 2012) the assumption of linearity (i.e., that the series interact additively) is not always justified when dealing with complex animate behavior. RQA is a sufficiently robust method, which does not need any additional assumptions about the analyzed time series. This means that even highly non-linear processes may be analyzed with this method and some of their properties can be captured and studied.

Technically speaking, RQA consists in extracting quantification measures from a so-called recurrence plot. So the first thing we will do in the next section is to introduce recurrence plots and to explain how they are built.

Recurrence matrixes and recurrence plots

A recurrence plot (RP) is the visualization of a special kind of a distance matrix – the recurrence matrix. Starting from a time series (e.g. a series of events) we can compare the value of each point in the series (e.g., each event) with that of every other point and say how similar they are on some metric – i.e. we compute the distance between the two points. We can now fill a square matrix, the same time series being both in the columns and in the rows, with the values of all these point to point comparisons (or distances) on the metric chosen. This matrix is called the distance matrix of the time series, but since we are interested in recurrences (i.e. appearance of the same, or very similar, events), we will ignore all the distances bigger than a predefined (small) value and consider just the remaining values as recurrences. The sparse matrix so obtained (i.e. just a few of the total cells of the

Figure 1. A. Sample time series of a sine function and its corresponding recurrence plot. B. Time series and corresponding recurrence plot of a white noise process. The lack of structure in the recurrence plot in B compared to the geometrical order of the recurrence plot of a deterministic time series (the sine function) in A is readily seen. Plots generated with the freely available program *RQA Software v. 14.1* (by C. Webber Jr.)



matrix will contain a value) is the recurrence matrix of the system. In other words, a recurrence in such a situation means that two points in a time series take on the same or very close values, and that we can represent the whole pattern of such recurrences in a square matrix where the time series is compared with a copy of itself.

The next step involves visualizing the pattern of recurrences obtained. It is easily achieved by changing every recurrence of the matrix into a single dot, in this way we transform a numerical matrix in a graphic plot, which we call the recurrence plot of the time series. It will be evident that the main diagonal of this matrix (and of the plot) is composed by the distance of every point from itself, and so it will be just a series of zeroes, i.e. an exact recurrence. Also, quite evidently, the plot will be symmetrical around this diagonal line. A couple of examples of an extremely regular (A) and a random (B) time series are given in Figure 1, together with their corresponding recurrence plot.

In all the previous discussion and examples we compared a time series with a copy of itself hence obtaining a square recurrence matrix and recurrence plot. It should be noted here however, that the same strategy may be applied to two different time series which for some reason we think may be related or generated by common processes, e.g. the language production of two persons conversing with each other – see below Dale & Spivey, 2006). In this case we may obtain a rectangular recurrence plot (i.e. the time series are not of the same length), where the main diagonal is not anymore the place of self-repetition of the series. We call this plot a cross recurrence plot (CRP), and analogously the quantification analysis of this plot will be called a Cross Recurrence Quantification Analysis (CRQA). Such plots are good tools for detecting mirroring of patterns in two distinct data series.

In all the previous cases we built recurrence plots where every single point in the time series is compared with each and every other point of the same series (RP) or of another series (CRP). In these cases we are then dealing with a one-dimensional embedding. But we may apply the same strategy to systems whose states follow trajectories in an *n*-dimensional state space and are defined by an *n*-dimensional position vector at every point in time. The computation of the distance of these vectors from each other is a straightforward extension of the one dimensional case and ends up in a similarly looking recurrence plot. A trivial exemplification may be the computing of the recurrence plot of every single letter (one dimensional embedding) in a text, as compared to the recurrence plot of three consecutive letters in the same text (three-dimensional embedding; similar strategies had been applied and will be dealt with in a later section, see Orsucci et al., 2006).

Usually in a complex system we deal with multiple factors and variables. The usefulness of recurrence plots comes from the fact, that in order to characterize a complex system of this sort, we don't need to have all the measures of all the variables defining it. As Taken has shown about three decades ago (Taken, 1981), a recorded time series may be thought of as just one measure of an underlying multidimensional dynamical system, and its behavior can be reconstructed from this single measure thanks to the method of delays (see Appendix 1).

Recurrence plots can then be used to capture visually and in a straightforward way the behavior of potentially very complex systems, whose dimensionality does not allow to represent their behavior in a state space (the space of all the possible states of the system – i.e. *n*-dimensional spaces are impossible to visualize when $n > 3$). In a bi-dimensional recurrence plot as described above, the distance between every two vectors, even in very high dimensional systems, reduces to a single number and, if it qualifies as a recurrence, can be easily represented with a single point.

Recurrence parameters

It is important to note that in order to perform a recurrence analysis, we need to specify several parameters in a non-trivial and non-automatic way. In their treatment of the RQA, Webber & Zbilut (2005) identified 7 parameters which have to be specified or estimated before a recurrence analysis can be run and a recurrence plot can be built. These parameters are: *embedding dimension* (M or EMBED), *delay* (τ or tau), *range*, *norm, rescaling*, *radius* and *line length (Line)*. Although there are some general guidelines in choosing the appropriate parameter values for particular types of systems (Abarbanel, 1996), the "golden standard" is still missing. A good knowledge of the phenomena under study should then guide the decisions about the values of these parameters. There is a growing literature on RQA applied to several study fields, from physiology to biology and behavioral sciences, where the interested reader can find some hints on the best strategies to adopt in this respect (Riley et al. 1999; Webber et al., 1995; Zbilut et al., 2002).

For the purposes of our discussion we will focus here on just the most important parameters to be set in most research situations, which are *embedding dimension*, *delay*, *radius* and *line length*. Others, mentioned above, may be required only for particular data and need a more technical discussion which is out of the scope of this paper (see Webber & Zbilut, 2005 for an exhaustive discussion).

Embedding dimension. This parameter postulates the underlying dimensionality of the system under study, and in practical terms defines the number of points (extracted from our data series) constituting the *n*-dimensional vectors on which recurrences are then computed. For example, as we mentioned above, we could perform a recurrence analysis on a text and use three consecutive letters to compute recurrences. In such a case embedding dimension will be equal to 3. If the text is composed of *k* total letters, letters $l_1l_2l_3$ will then be compared with letters $l_2l_3l_4$, $l_3l_4l_5$, ..., $l_{k-2}l_{k-1}l_k$, and so on with all the other possible comparisons. Real world data are often very noisy which is a source of dimension inflation. This means the real dimensionality of the system is typically lower than the dimensionality we should use in RQA i.e. thanks to these additional dimensions it is possible to completely unfold the dynamics of the system and capture relevant patterns therein – up to a point where no additional dimensions will be informative anymore. Webber & Zbilut (2005), for example, suggest using embedding dimension between 10 and 20 when studying biological systems (for an example of the techniques of the dimension estimation in the case of real measures of a biological system see Pellecchia and Schokley, 2005).

Delay (τ**)** specifies the time-lag used to reconstruct the multidimensional phase space. In the preceding example of the three letters used to compute recurrence, the delay used to sample the vector of letters was $\tau = 1$, i.e. the letters were consecutive with each other. But that does not have to be always the case[1]. Some authors (Grassberger et al, 1991) propose the delay being a non-critical parameter, and hence advice not to be overly concerned with finding the 'optimal' τ. Yet a conservative strategy, adopted by researchers in the behavioral sciences (see Pellecchia & Shockley, 2005; Riley et al, 1999), would be to observe the changes in value of quantitative measures (e.g. %REC – see below) as a function of a range of values of this parameter. If these changes are smooth within the range considered we may be safe that no artifact due to parameter selection will be present.

Radius indicates the value of the distance within which two points in phase space are considered recurrent. In other words, this is the greatest dissimilarity

[1] It has been proposed to select proper delay by finding the first minimum in the linear autocorrelation function or mutual information function (Fraser & Swinney, 1986) but this method is only generally valid when the system is stationary (Zbilut & Webber, 2006), which may be not often the case in physiological and behavioral phenomena.

between two events which still allows us to consider them as equivalent. Thus radius is the cut-off value separating those distances, which are to be counted as recurrences from those which are not. When the radius approaches the maximal distance, more and more cells in the matrix will fall within this cut-off value and more and more points will fill up the recurrence plot. Radius hence cannot be set to high values, since this would let in points quite dissimilar to each other, i.e., events, which should not be considered recurrent anymore. Again, heuristic methods can help us set this parameter. As discussed above in the case of the delay parameter, here too we may track changes on some quantification variable (e.g. %REC) modulated by a range of values of the radius parameter. This can lead to the choice of the most appropriate radius, the rationale being to keep the value of %REC relatively low (i.e. 1% to 5%) (Zbilut & Webber, 2006; Pellecchia & Shockley, 2005).

Line length (Line), at last, is a parameter whose setting has not an effect on how the recurrence plot will look like, but rather on some of the measures extracted during RQA. Consecutive points in a recurrence plot indicate trajectories of the system in the phase space which get close and 'go together' for some time. Long lines are a sign of increasingly deterministic, exactly recurrent behavior. The Line parameter, then, just specifies the number of consecutive recurrent points which we take to define a line segment. The minimum number of points to make a line is 2, and this is also the typical value assigned to this parameter.

Recurrence quantification

Recurrence parameters help building a RP of the system under study, giving a first appreciation of the nature of the system itself through a qualitative observation of the patterns forming on the plot. A natural extension of this observation would be to obtain some quantitative measures from the plot, which could automatically and objectively uncover some of the intrinsic features of the system. This extension is what has been called the Recurrence Quantification Analysis of the time series (Zbilut and Webber,1992, Webber & Zbilut, 1994). Once computed, quantitative recurrence measures can be used as data to be further analyzed, since they capture some of the dynamic aspects of the phenomenon under study. Quantitative recurrence measures of a RP include: *recurrence rate* (%REC or RR), *determinism* (%DET), *linemax* (LMAX), *entropy* (ENT), *trend* (TND), *laminarity* (%LAM) and *trapping time* (TT).

The first measure, **recurrency rate (%REC)**, is a simple proportion of recurrence points present in the plot

%REC = [number of recurrence points in RP] / [number of possible points in RP].

Since we tend to maintain the recurrence matrix sparse, %REC will also tend to be relatively low (~ 1-6%).

Structures of particular interest in a recurrence plot are the diagonal lines which indicate the fact that trajectories in the phase space come close and run parallel for some time together (or, in other words, that consecutive events in a time series are similar to other consecutive events in a different moment in time of the series). If we measure the percentage of recurrent points forming these lines (as defined by the parameter *Line*) with respect to all recurrence points, we obtain the second quantification measure, **deteminism (%DET)**:

> %DET = [number of recurrent points being part of a diagonal line] / [number of recurrence point in RP]

This quantity is a measure of deterministic behavior, since regular and repeating patterns of the dynamics will produce (more or less) long lines in the recurrence plot. In Figure 1 A for example we can see long, uninterrupted lines since the system visits regularly the same set of values and in this case %DET will tend to 100% (more precisely 82% – almost all recurrence points in the plot fall on a line, since we consider just those lines parallel to the diagonal). The plot in Figure 1 B, on the contrary, shows almost no lines at all. The process generating the signal is completely random and the measure %DET will be close to zero.

The third recurrence measure (variable) is **linemax** (**LMAX**), which simply measures the length of the longest diagonal line segment in the plot (other than the main diagonal of self identity).

Another informative measure that can be computed for each recurrence plot is the **entropy** (**ENT**), which is a measure of the signal's complexity. The higher is the value of this variable, the more complex the system generating the signals.

A fifth recurrence variable is the **trend** (**TND**). This quantity measures the degree of stationarity of the system. Stationarity of a dynamical system indicates the fact that the time series fluctuates around a roughly stable mean value and does not drift away from it. The more recurrence points are homogeneously distributed across the RP, the more the system seems to be stationary in the time window analyzed, and the value of TND will tend to zero. On the contrary concentration of the recurrence points along the diagonal for example or generally in a non-homogeneous way across the recurrence plot will be a sign of non-stationarity of the system.

Important kinds of structures in recurrence plots, other than diagonal lines, are vertical/horizontal lines, meaning that a state in the phase space doesn't change or changes very slowly for some time. It's worth noticing that vertical lines in one half of the RP will look horizontal in the other half so we have to consider both. Here again the relevant defining parameter is *Line*. Quantitative variables capturing these laminar states are **laminarity** (**%LAM**) and **trapping time** (**TT**). %LAM corresponds to %DET with the difference that vertical structures are considered instead of diagonal. TT, on the other end, is just the average length of vertical line structures.

As should be clear by now, recurrence plots and recurrence variables strictly depend in their structure and values upon the sequential organization of the data in a time series. If we were to shuffle the string of data, we would compute a different set of quantification values while, for example, the same mean value and standard deviation would be obtained. This is to stress the fact that the way a phenomenon unfold in time has specific and unique features which are often overlooked by standard methods of data analysis and can be captured by RQA instead. RQA has also the advantage not to require any additional statistical assumption (stationarity, normality). Additionally, the application of the RQA methods does not require an excessively long series of data points.

With this basic knowledge of the tool, we can now turn to present some problems in the behavioral sciences where RQA has been applied, giving a fresh look to the tackled problems.

Applications to the psychology of language

RQA is recently having a significant impact on research in the human sciences and psychology. Language and communication in particular is an extremely complex and multidimensional phenomenon to approach and study, thus, as mentioned in the introduction one that may benefit from the application of dynamically sensitive analysis techniques such as RQA. Complexity and recurrence in language can be studied at different levels of analysis. There have been attempts to analyze language data at the lexical and semiotic level (Orsucci et al. 1999, 2005; Dale & Spivey, 2005), at the level of categorical and syntactic structures in a discourse (Dale & Spivey, 2006), at the level of turn-taking times in a conversation (Ashenfelder, 2007; Rączaszek-Leonardi et al., in preparation) or, more recently, there has been a very interesting attempt to analyze the level of concepts and semantics in human discourse (Angus, Smith & Wiles, 2012; Angus et al., 2012). Moreover conversation and language comprehension may result in forms of coordinative (dynamical) structures which are clearly apparent when complementary behavioral measures like postural sway (Shockley et. al 2003) or eye movements (Richardson et al. 2005) are analyzed by means of RQA.

Lexical level

One of the most straightforward and simple ways to approach the study of dynamical properties of a text, is to consider it as a flow of discrete symbols like the letters forming words and phrases. Orsucci and collaborators, in a series of studies, used the recurrence quantification strategy as a tool to obtain quantitative and reliable measures of the structuring of texts, intended as flows of symbols in which to look for recurrences of motifs (Orsucci et. al., 1999, 2004, 2005). In one of their studies (Orsucci et al. 1999) they analyzed different sets of texts, such as Italian and American poems, Swedish poems and their corresponding Italian translations. The recurrence variables computed on those series (different texts) were then compared and appropriately interpreted. By defining a delay τ of 1 and an embedding dimension M of 3, they were able to detect recurrences at the word level. In this way, in fact the recurrent vectors are sequences of three consecutive letters, which, as the authors claim, is a good approximation of word stems. They computed recurrence measures and plotted the most important two (%REC and %DET) against each other. In this way they were able to determine an invariant relation between these two variables, that is a regression line along which all the texts closely aligned. According to the authors the high correlation between the two variables reflects a common structural design of the texts. Moreover it was possible to interpret the position in this plot in terms of the 'complexity' of the prosodic structure of the poem, in that the more deterministic side of the plot (high %REC to %DET ratio) correspond to the more recurrent poems, the ones with highly repetitive motifs, while low %REC to %DET ratios indicate more complex poems, with weak or no rhythmic structure. Orsucci et al. (2004; 2006) also experimented a recurrence quantification strategy on conversations. They compared a clinical speech sample (an interview of a psychopathological patient with his doctor) with a normal speech sample (a loose conversation between friends). Using the same parameters and computing the same measures as seen above, they plotted for every turn in the conversation of each of the persons involved in it the %DET measure relative to the single turn considered. Hence two lines (one for each participant) of %DET resulted as a dependent function of the serial order of the conversational turns. In this way it was possible to evaluate visually the degree of coordination of the variable considered between the participants. Orsucci et al. (2004; 2006) interpreted in terms of coupling/decoupling of the actors at a conversational level the convergence/divergence of the different values of %DET in the course of it, although it is not quite clear through which mechanism the extracted measures should in fact be related.

Another example of an exploratory study on the use of RQA at the lexical level was given by Webber and Zbilut (2005). They analyzed texts of the transcripts of the talk of a schizophrenic patient as compared to the transcripts of an academic lecture. In this example the word level was also taken into account to codify the time series data and compared to the letter level method as used in the Orsucci et al., (1999) study. By arbitrarily indexing every new word in the text with a different number a time series was built and RQA was performed on these. The authors chose an embedding dimension M = 1 and radius = 0, meaning that only recurrences of exact words were considered as points in the recurrence matrix (in fact by choosing a radius > 0 two nearby numbers would have counted as recurrences, which is inappropriate considering the random assignment of numbers to words). The authors found some differences in the pattern of variables derived with RQA, and a tentatively greater difference between normal and schizophrenic texts in the word level coding compared to the letter level coding. Even though the aim of the study was not to draw far reaching conclusions about the difference between the normal and the schizophrenic speech, it indicated the best level of analyses and pointed to the possibilities for carefully designed experimental studies, aimed at exploring dynamical characteristics of different kinds of texts by means of quantities derived from RQA.

Similarly to Webber & Zbilut (2005) also Dale & Spivey (2005) codified every word in a series of texts as different numerals in an arbitrary way. Every lexical token corresponded to a different number, and only recurrences of pairs of words (that can be thought of as collocations – i.e. $M = 2$, $\tau = 1$ and radius = 0) were taken into account. In this case the texts were extracted from the English database CHILDES (MacWhinney, 2000) and represented dialogues between children at different ages and their caretakers. The aim of the researchers was to evaluate the hypothesis that language learning comes about during a process of coordination or alignment of language use, so that inputs and contingent responses in a conversation, if effective, shape the language of both the child and the caregiver (Dale & Spivey, 2005). An important innovation in the Dale & Spivey study was the use of a cross recurrence plot (CRQ) for their analyses. Since synchronization was sought between the language use of a child and its caregiver, the two time series (child and caregiver utterances) were plotted against each other in a recurrence plot where every dot corresponded to the recurrence of the words used by one of them in the language use of the other. The basic parameter and quantification measures are the same as for RQA but, as explained above, we may have to deal with rectangular matrices (the length of the compared time series may vary) and the main diagonal is not the place of self-similarity anymore. In their treatment, Dale & Spivey used the standard %REC measure but also introduced an *ad hoc* quantification measure which is meant to probe the synchronization of the conversation. The reasoning was that what matters the most for language learning is the on-line coordination and alignment, which would produce recurrences in nearby exchanges rather than along the whole conversation. In terms of the structure of the recurrence plot this means that recurrences should concentrate along the line of incidence (i.e. main diagonal of the rectangular matrix) and in a band around it. This band, constituted by +/- 50 points (words), is taken as an approximation of the on-line conversation of a child and a caregiver and the real time coordination of the utterances pronounced in consecutive or near-consecutive turns. By computing the proportion of recurrence points in the face of the total possible ones inside this band (i.e., if utterances were repeated by the interlocutors within 50 words), they obtained the $\%DREC_{50}$ quantification measure. These two measures (%REC and $\%DREC_{50}$) were obtained for the original sample (real child-parent conversations) and for a sample used as a control, constructed by pairing the utterances of the child in one with the utterances of the caregiver in a successive conversation. Results reveal a significant difference in each corpus of data between original and control samples, confirming a non-casual structuring of the recurrences in the conversation. Low values of %REC in general and relatively higher values of %DREC$_{50}$ are also registered pointing to a non trivial lexical coordination between the interacting actors. Moreover and interestingly enough, this lexical coordination (i.e. recurrence of lexical elements) appears to be dependent on the age of the children. If the values of %REC and %DREC$_{50}$ are regressed over time a significant decrease of these measures is evident. Children seem to be more prone to 'imitative' lexical behavior early on in their language learning process, probably due to their still incomplete/poor lexical competence, and it may be that this very imitative behavior, as revealed in the on-line coordination during conversations with the caretakers, leads to an improvement of their lexical competence and moves the alignment with the converser at a deeper syntactic level.

Syntactic level

Another way to analyze alignment in conversation, which is thought to be an important factor in structuring linguistic interactions (Pickering & Garrod, 2004), is to consider the syntactic structures making up the text instead of the concrete words used. Dale and Spivey (2005, 2006) performed their analysis of children's conversations with their caregivers taking into account the syntactic structuring of the conversation. The broad hypothesis is that language use in children and caregivers follow a pattern of alignment or coordination, within which the process of language learning is accomplished. The same corpus of data that was previously analyzed at the lexical level was now transformed to consider the syntactic category of each word (e.g. verb, determiner, noun etc.) in a sequence, coded in a series of arbitrarily assigned numerals. Cross-recurrences of chunks of different lengths ($M = 2$, 3 and 4; $\tau = 1$; radius = 0) in the syntactic sequence were then plotted and quantified using the same quantification measures as in the lexical study (%REC and %DREC$_{50}$), and finally compared with corresponding quantification of control conditions (cross recurrence with another conversation plus shuffled samples of same conversation). The significant increase in the crucial measure %DREC$_{50}$ indicates that indeed there is a coordination of the syntactic structures in the ongoing conversation, which, again, seems to decrease in the course of development. This means that syntactic structure alignment during conversation is stronger at earlier stages of development. Coordination of this type may then provide the kind of context in which children learn word class sequences by being guided and guiding at the same time the linguistic exchanges with their caregivers.

Semantic level

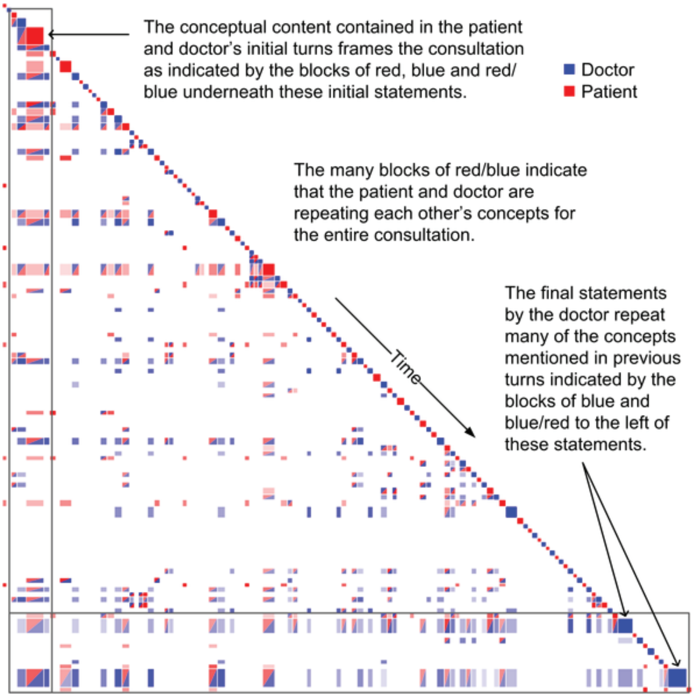
Conversation analysis can obviously gain considerable depth if we find a way to represent in a visual and informative way what is the most important element of every conversation, i.e. its semantic content. Recurrence Analysis is one of the techniques, which can help also in this respect. Angus and collaborators (Angus, Smith & Wiles 2012; Angus et al., 2012) proposed a recurrence plot that highlights the recurrences of conceptual chunks in a discourse or a conversation – structures at a more abstract level than the lexical or syntactic ones. They argue that at a semantic level words which are not equal (e.g. 'cat' ≠ 'dog') and would never be counted as recurrences in a recurrence analysis of the kind presented above, may in fact be connected in a deeper way and hence considered as similar or equivalent (e.g. they are both 'pets'), i.e. a recurrent concept.

The key of their proposal is to join a conceptual similarity algorithm (a semantic net, able to compute the 'semantic' distance of every word with the others) with a recurrence plot technique applied to text data and augmented in its graphical properties (i.e. colors and dimensions of the recurrence dots). In short, given a semantic model of the text (built from the text itself or adopted among available ones) it is possible to compute the degree of conceptual similarity of every word in the text/conversation with the others. In their work, Angus, Smith & Wiles (2012) adopted the method proposed by Salton (1989), which was built on a probabilistic model of co-occurrences of words in the same sentence (here a 3 terms long window). On the basis of these probability values, the similarity of every two terms in the text can be computed which, in the end, allows for the computation of conceptual similarity of two utterances. Utterances are some appropriately chosen meaningful chunks (e.g. the turns in a conversation; the paragraphs in a written text etc.) of the text under study, Once divided in this way, the text assumes the form of a time series of events, on which RQA is performed – i.e. utterances are compared with each other in terms of similarity of the words within them, and recurrence of semantic content is measured (see Angus, Smith & Wiles, 2012, for a step by step explanation of the similarity computational algorithm and probabilistic model used).

As an application of this approach Angus et al (2012) took sample conversations – a doctor/patient consultation – and divided them into utterances corresponding to the conversational turns within them. Each turn is compared with every other in the conversation and a similarity score is obtained (from the semantic model adopted). The main diagonal of the recurrence plot is the place where each utterance is compared with itself (and the similarity is maximal), while the rest of the plot will be filled by recurrence dots (i.e. squares in this particular case) in the places where semantically similar utterances cross. Since recurrences are symmetrical with respect to the diagonal, only the lower half can be plotted and analyzed (see Figure 2).

The pattern of recurrences in the plot is very informative of the course of the consultation, and Angus, Smith & Wiles (2012; Angus et al., 2012) pointed to several structures appearing in these conceptual recurrence plots as indicative of particular conversational situations. The presence of white spaces, recurrence lines, engagement blocks and other qualitative features – which may also be appropriately evidenced by a sensible use of colors and dimension of the blocks in

Figure 2. An example of conceptual recurrence plot of a conversation between doctor and patient. Recurrences are evident and some of its structure is indicative of significant synchronization / alignment in the interpersonal interaction. *From Angus et al. (2012)*



the plot – are all characteristics of particular conversational situations, such as "conceptual drifts", the relative contribution of the conversers to the subject matter, the moments in which subjects are strongly engaged in discussing a particular set of concepts and so on. It is even possible to evaluate a patient-doctor conversation in terms of effectiveness, if we are able to tie different typologies of exchanges apparent in these conceptual recurrence plots, to the final outcome of the therapy (Angus et al., 2012).

This research seems to be a particularly astonishing and promising direction for the future of RQA-type analyses in natural language research, as one can clearly sense how it allows for the reduction of a description of an immensely complex system of two conversing people without destroying important and relevant patterns. The observations are still qualitative in nature, but it's not hard to imagine that quantitative measures as the ones reviewed above (Zbilut & Webber, 1992; Webber & Zbilut, 1994) can be designed in the future and applied to these peculiar applications of recurrence analysis to semantic contents.

Conversation as coordination

A quite different way to approach the structuring of an on-going conversation has been taken by Shockley et al. (2003). Conversation in the view of these authors is a form of joint action, and as such requires a strict time dependent coordination between the actors involved in it. While the coordination certainly takes place at a level of semantic communication, it can be reflected at other levels of the individuals' behavior as well. For example, postural sway, defined as the slight shifts around the center of mass of a standing actor, is related to other activities he or she is involved in at a given moment (i.e. the actor must account for them to maintain a stable postural stance). Hence, it is supposed that if the activity an actor is involved in is a conversation with a second person, postural sway should be related to it as well.

Given these premises, Shockley et al. (2003) measured the postural sway of couples of individuals engaging in a conversation between each other or each with a third party confederate whose posture was not recorded (first factor), and in a situation of visual contact with each other or lack of it (second factor). The general goal of the conversation was to communicate in order to discover a few differences in pictures they could just see separately from each other. Measures of coordination were reflected in the %REC and LMAX quantities extracted from the cross recurrence plot of standardized postural sway of the two subjects, who in one condition were engaged in a reciprocal conversation and in a second condition, were not (i.e. they were conversing with a third party). Computing a CRQA of the postural data in this factorial design resulted in discovering significant differences in the task partner condition (i.e. reciprocal conversation vs. conversation with confederates), and was not affected by visual contact. In other words subjects showed coordination in postural sway recurrence measures if they were conversing with each other (as opposed if they were conversing with someone else), and they showed coordination not only if they could see each other but even if they conversed out of each other sight.

Although it is a far leap to connect in a detailed explanation the signs of coordination in the postural sway of communicating individuals with the signs of coordination on the semantic, syntactic and other levels of their superordinate on-going conversation, these results undoubtedly point to the fact that the former must be related to the latter. We can think of language as a coordination device producing some sort of entrainment among initially autonomous individuals. This basic entrainment facilitates forming an organized ensemble, allowing to achieve shared goals (Rączaszek-Leonardi & Cowley, 2012; Fusaroli, Tylen, Rączaszek-Leonardi, in press). More research is needed to specifically delineate the nature of such entrainment, but recurrence strategies in postural constraints, studied together with recurrence on other levels of linguistic behaviors, seem to show a promising way where to look when studying coordinative behavior.

Conclusion

The study of psychological and behavioral phenomena is increasingly focusing on the dynamics of cognition and action, i.e. there is a feeling that these phenomena could be better understood if studied in a framework where they are seen as happening in time, as a product of specific dynamics, which we need to unveil and characterize. New methods of observation and experimental designs are of course needed as well as new methods of the analysis of data collected in this way. RQA is definitely one of the most promising of such methods. It is an already quite established and useful tool, which is able to capture different dynamical signatures in data sets of great complexity.

In this paper we gave an introduction to the main concepts and steps involved in running RQA. Application to the field of interest will have to refer to these concepts and steps, but at the same time a creative approach and sound theoretical reflections are of primary importance when using it as an analytical tool to describe the collected data. Accurate considerations and informed parameters choice are a core step in RQA and not a trivial / automatic one. An increasing body of literature (see e.g. www.recurrence-plot.tk/bibliography.php), both on the theoretical side, and on the applicative one is also of great help in our methodological decisions. It should be emphasized that RQA methods are continuously developing under the pressures of new needs from the experimental sciences. Thus we hope that this work will not only encourage qualitatively-oriented behavioral scientists to use these quantitative methods but also that this will contribute to a further development and refinement of the methods themselves.

References

Abarbanel, H.D.I (1996). *Analysis of observed chaotic data.* New-York: Springer.

Angus, D., Smith, A.E., & Wiles, J. (2012). Conceptual recurrence plots: Revealing patterns in human discourse. *IEEE Transactions on Visualization and Computer Graphics*, 18 (6), 988-997.

Angus, D., Watson, B., Smith, A., Gallois, C., & Wiles, J. (2012). Visualising conversation structure across time: Insights into effective doctor-patient consultations. *PLoS ONE*, 7 (6): e38014.

Bassingthwaighte, J., Liebovitch, L., & West, B. (1994). *Fractal physiology*. New York: Oxford University Press.

Burrus, C. (2012). Fast fourier transforms. Retrieved from the Connexions Web site: http://cnx.org/content/col10550/1.22/

Costall, A. (2004). From Darwin to Watson (and cognitivism) and back again: The principle of animal-environment mutuality. *Behavior and Philosophy*, 32, 179-195

Dale, R. & Spivey, M.J. (2006). Unraveling the dyad: Using recurrence analysis to explore patterns of syntactic coordination between children and caregivers in conversation. *Language Learning*, 56 (3), 391-430.

Dale, R. & Spivey, M.J. (2005). Categorical recurrence analysis of child language. In B. Bara, L. Barsalou, & M. Bucciarelli (Eds.), *Proceedings of the 27th Conference of the Cognitive Science Society* (pp. 530-535). Mahwah, NJ: Lawrence Erlbaum.

Eckmann, J.-P., Kamphorst, S.O., & Ruelle, D. (1987). Recurrence plots of dynamical systems. *Europhysics Letters*, 4, 973-977.

Fraser A.M. & Swinney H.L. (1986). Independent coordinates for strange attractors from mutual information. *Physical Review A*, 33, 1134-1140.

Fusaroli, R., Tylen, C., & Rączaszek-Leonardi, J. (under review). Dialogue as synergy.

Haken, H. (1983). *Synergetics, an introduction: Nonequilibrium phase transitions and self-organization in physics, chemistry, and biology*. New York: Springer.

Kelso, J.A.S. (1995). *Dynamic patterns: The self-organization of brain and behavior*. Cambridge: MIT Press.

MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk*. Mahwah, NJ: Erlbaum.

Orsucci, F., Walter, K., Giuliani, A., Webber, C.L.Jr., & Zbilut J.P. (1999). Orthographic structuring of human speech and texts: Linguistic application of recurrence quantification analysis. *International Journal of Chaos Theory and Applications*, 4 (2-3), 21-28.

Orsucci, F., Giuliani, A., Zbilut, J.P. (2004). Structure & coupling of semiotic sets, In: *Experimental Chaos Conference 8, Florence, AIP Conference Proceedings*, 742 (1), 83-93.

Orsucci F., Giuliani, A., Webber, C.L.Jr., Zbilut, J.P., Fonagy, P., & Mazza, M. (2006). Combinatorics and synchronization in natural semiotics. *Physica A*, 361 (2), 665-676.

Pellecchia, G.L. & Shockley, K. (2005). Application of recurrence quantification analysis: Influence of cognitive activity on postural fluctuations. In M.A. Riley & G.C. Van Orden (Eds.), *Tutorials in contemporary nonlinear methods for the behavioral sciences* (pp. 95-141). Retrieved June 1, 2012, from http://www.nsf.gov/sbe/bcs/pac/nmbs/nmbs.jsp

Pickering, M. & Garrod, S. 2004. Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, 27 (2), 169-190.

Port, R., & van Gelder, T. (Eds.) (1995). *Mind as motion: Explorations in the dynamics of cognition*. Cambridge: MIT Press.

Rączaszek-Leonardi, J. & Cowley, S. (2012). The evolution of language as controlled collectivity. *Interaction Studies*, 13 (1), 1-16.

Richardson, D.C., Dale R., & Kirkham N.Z. (2007). The art of conversation is coordination: Common ground and the coupling of eye movements during dialogue. *Psychological Science*, 18 (5), 407-413.

Riley, M.A., Balasubramaniam, R., & Turvey, M.T. (1999). Recurrence quantification analysis of postural fluctuations. *Gait and Posture*, 9, 65-78.

Salton, G. (1989). *Automatic text processing: The transformation, analysis, and retrieval of information by computer*. Boston, MA: Addison-Wesley.

Shockley, K., Santana, M.-V., & Fowler, C.A. (2003). Mutual interpersonal postural constraints are involved in cooperative conversation. *Journal of Experimental Psychology: Human Perception and Performance*, 29, 326-332.

Takens, F. (1981). Detecting strange attractors in turbulence. In D. Rand & L.-S. Young (Eds.), *Dynamical systems and turbulence, Lecture notes in mathematics. Vol. 898* (pp. 366-381). Berlin: Springer.

Thelen, E. & Smith, L.B. (1994). *A dynamic systems approach to the development of cognition and action.* Cambridge, MA: The MIT Press.

Tschacher, W. & Dauwalder, J.P. (Eds.) (2003). *The dynamical systems approach to cognition: Concepts and empirical paradigms based on self-organization, embodiment, and coordination dynamics.* Singapore: World Scientific Publishing.

Van Orden, G.C., Holden, J.G., & Turvey, M.T. (2003). Self-organization of cognitive performance. *Journal of Experimental Psychology: General*, 132, 331-350.

Webber, C.L.Jr., Schmidt, M.A., & Walsh, J.M. (1995). Influence of isometric loading on biceps EMG dynamics as assessed by linear and nonlinear tools. *Journal of Applied Physiology*, 78, 814-822.

Webber, C.L.Jr. & Zbilut, J.P. (1994). Dynamical assessment of physiological systems and states using recurrence plot strategies. *Journal of Applied Physiology*, 76, 965-973.

Webber, C.L.Jr. & Zbilut, J.P. (2005). Recurrence quantification analysis of nonlinear dynamical systems. In M.A. Riley & G.C. Van Orden (Eds.), *Tutorials in contemporary nonlinear methods for the behavioral sciences* (pp. 26-94). Retrieved June 1, 2012, from http://www.nsf.gov/sbe/bcs/pac/nmbs/nmbs.jsp

Zbilut, J.P. & Webber, C.L.Jr. (1992). Embeddings and delays as derived from quantification of recurrence plots. *Physics Letters A*, 171, 199-203.

Zbilut, J.P. & Webber, C.L.Jr. (2006). Recurrence quantification analysis. In M. Akay (Ed.), *Wiley encyclopedia of biomedical engineering*. Hoboken, NJ: John Wiley & Sons.

Zbilut, J.P., Sirabella, P., Giuliani, A., Manetti, C., Colosimo, A., & Webber, C.L.Jr. (2002). Review of nonlinear analysis of proteins through recurrence quantification. *Cell Biochemistry and Biophysics*, 36, 67-87.

Appendix 1

Takens (1981) demonstrated that it is possible to reconstruct from a single measure the underlying multidimensionality of a system thanks to the method of time delays. Takens' theorem (and method) states that starting with a single measured variable of the system, we could reconstruct and plot in the appropriate *n*-dimensional state space its behavior by appropriately choosing a time delay τ (tau). The delay is used to extract from the given time series *n* points at every step, each tau-delayed from the others. The *n*-dimensional vector so obtained can be thought as the position vector (the coordinates) of the system, at a given moment, in the *n*-dimensional state space. By plotting every vector we can then reconstruct the trajectory of the system. For example it can be shown that from one single lead ECG data (Figure 3. A) we may effectively model in a 3-dimensional space the real behavior of polarization/depolarization of the heart tissue during heartbeat. The single lead ECG registration lives in reality in a higher-order phase space, i.e. the 3-dimensional positioning in space and working of a real heart (Figure 3. B).

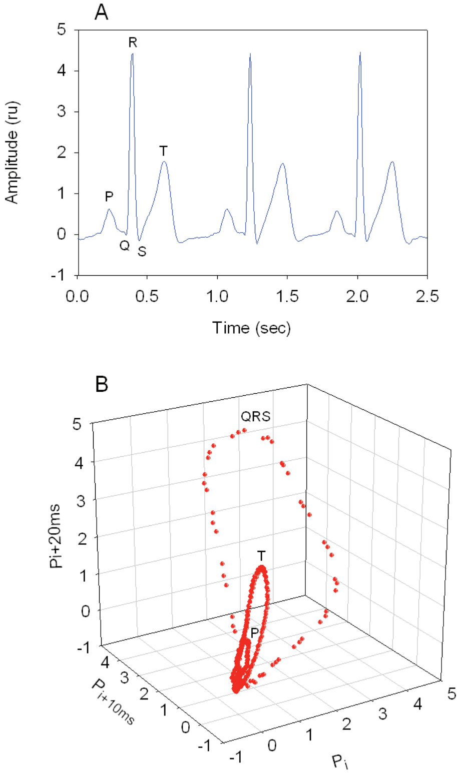


Figure 3. A. Registration of a single lead ECG data. B. Reconstruction of the three-dimensional activation of the /depolarization pattern/ as it could appear in a real system (heart). *From Webber and Zbilut (2005) reprinted with permission*