UC MERCED

Introduction to *categorical* Recurrence Quantification Analysis (RQA)

Rick Dale

Cognitive and Information Sciences

University of California, Merced

http://cognaction.org/rick/ati

Nonlinear Methods for Psychological Science

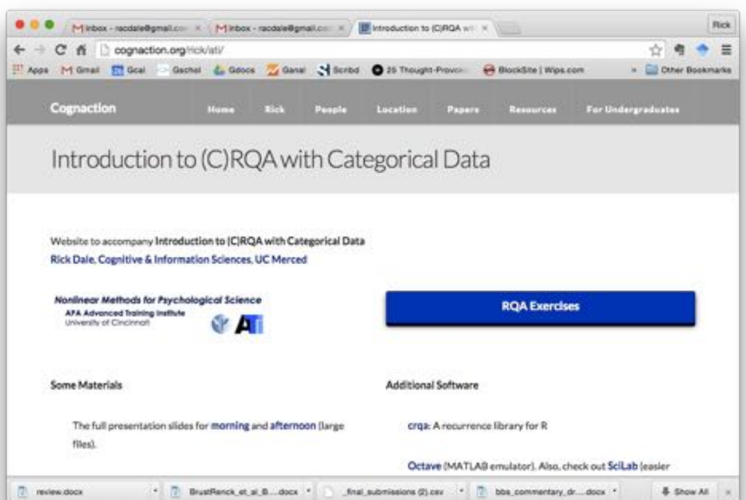
APA Advanced Training Institute

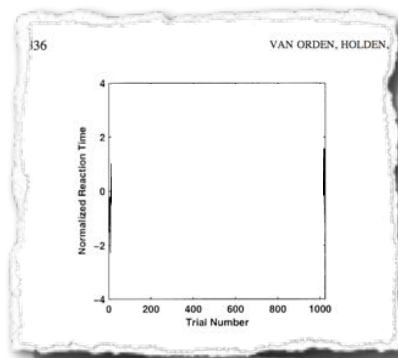
University of Cincinnati

June 15-19, 2015

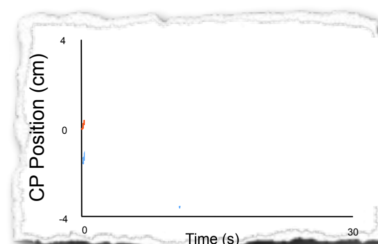
cogsci.ucmerced.edu

http://cognaction.org/rick/ati





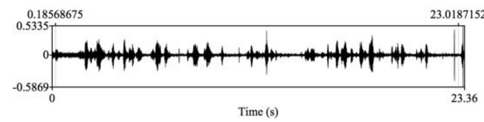
Van Orden, Holden, & Turvey, 2003



Schmit et al. (2006)

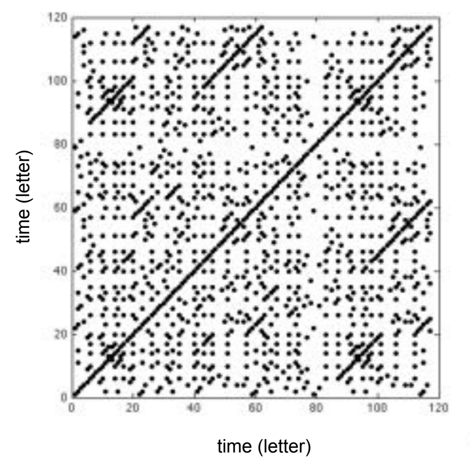
Linguistic Levels...

- *conversational topic, pragmatics, ...*
- *phrase structure, syntactic frames, ...*
- *syllables, words, ...*
- *speech rate, pause duration, ...*
- *amplitude, frequenc(ies), pitch, ...*



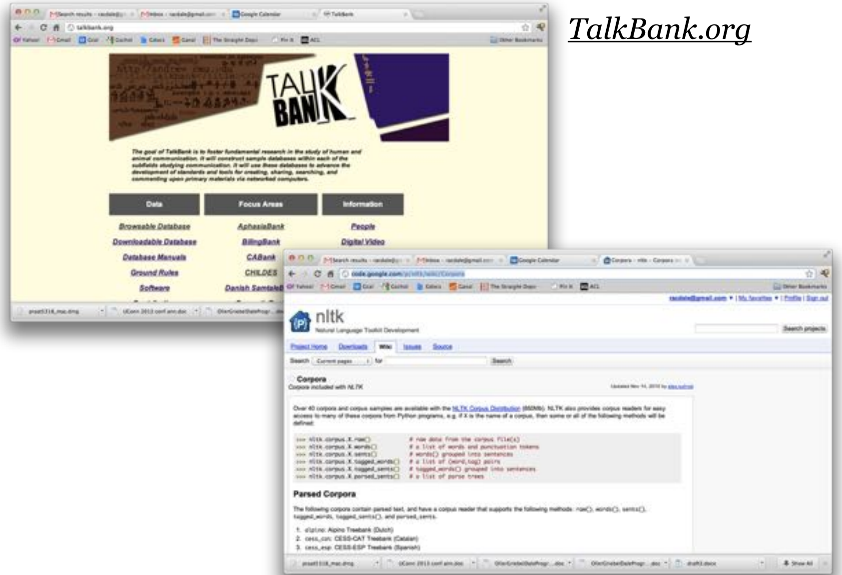
Higher-Order States

- *Interaction-dominant dynamics of the cognitive system produce higher-order collective properties that can be subjected to nonlinear methods.*
- E.g.: behavior sequences (eye-movement fixations, actions, etc.)
- E.g.: linguistic sequences (words, sentences, etc.)



Outline

- Time series of higher-order states
  - Analysis of series of behavioral *categories*
- The recurrence plot (RP) and "textures"
- Quantifying the plot (RQA)
- Examples and exercises



TalkBank.org

http://code.google.com/p/nltk/wiki/Corpora

Example

Quote

- *Rock and roll music, if you like it, if you feel it, you can't help but move to it. That's what happens to me. I can't help it.*

Lyric

- *Honey, lay off of my shoes. Don't you step on my blue suede shoes. You can do anything but lay off of my blue suede shoes*

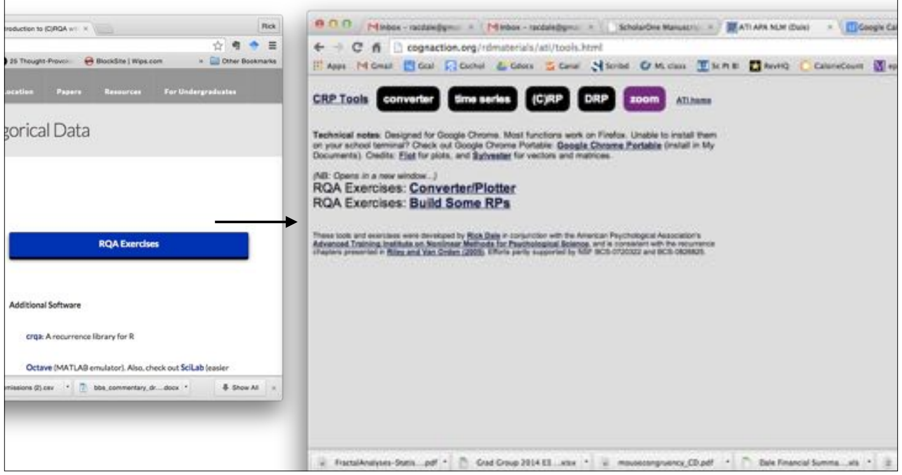
Conversion...

- Converting transcripts to sequence of numeric identifiers:
  - *blue suede shoes …*

    b = 1, l = 2, u = 3, e = 4, …
  - *1 2 3 4 5 6 3 4 7 4 5 6 8 9 4 6 …*
- Level of analysis here: letters

http://cognaction.org/rick/ati

Exercise: Converter / Plotter

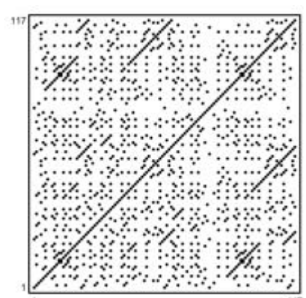


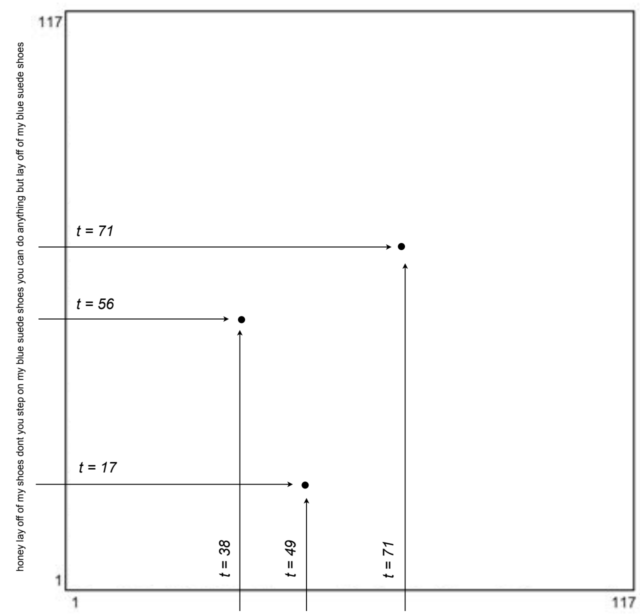
Trajectories

- Think of the behavior sequences in geometric terms: trajectories in “category space.”
- How repetitive is a sequence? If repetitive, how *long* are the repetitions? Lots of short bursts, or fewer but lengthy repeated trajectories?

Recurrence Plot (RP)

- Bird’s-eye view of the system’s trajectories through its behavior space.





Under the Hood

- If time series is:

  $\mathbf{x} = 1, 3, 2, 3, \ldots, x_N$
- RP = set of points (i, j) such that:

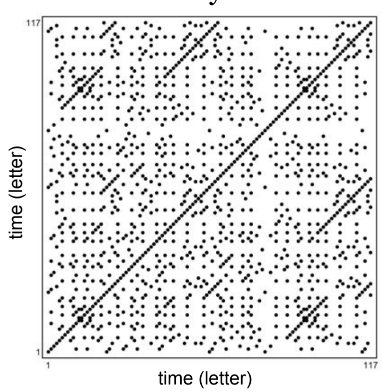
  $x_i - x_j = 0$
- In other words, set of points such that the numeric identifiers *have a distance of zero from each other*.

Under the Hood

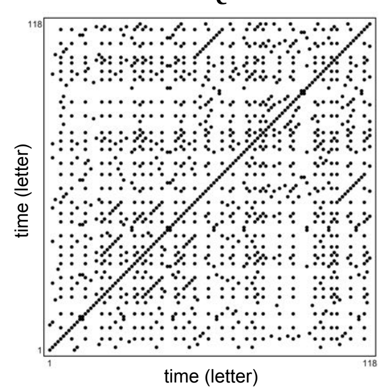
- In other words, the system's trajectory is revisiting the same state, or *recurring* in time.

Textures (Eckmann et al., 1987)

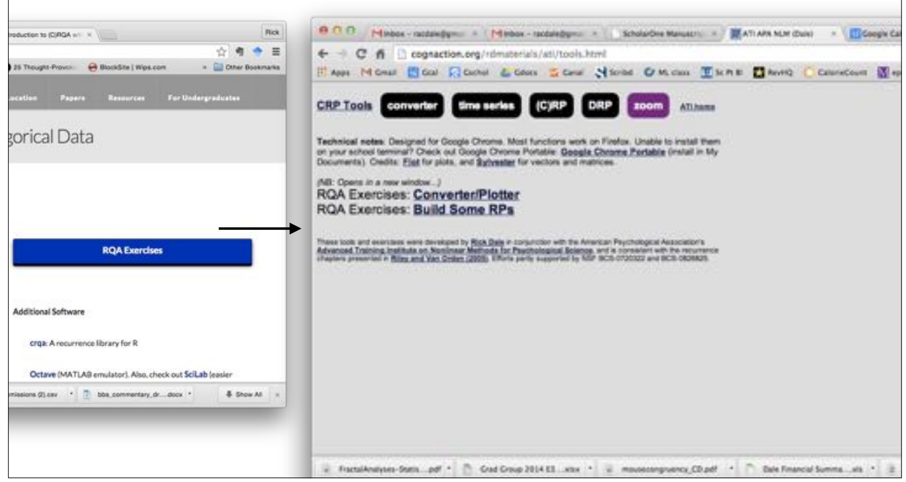
Elvis Lyrics



Elvis Quote



Exercise: Build Some RPs



Quantification

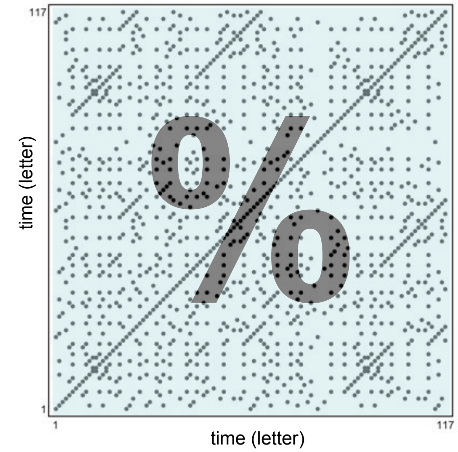
- You can eyeball a plot, but in real contexts we want some quantification so that plots (or, e.g., conditions) can be compared.
- Enter: *recurrence quantification analysis* (RQA).
- Quantities capturing the amount and distribution of points on the plot.

RQA Measures

Recurrence rate (*%REC*): Total percentage of the plot occupied by points.

$$RR = \frac{1}{N^2} \sum_{i,j=1}^{N} \mathbf{R}_{i,j}$$

http://www.recurrence-plot.tk

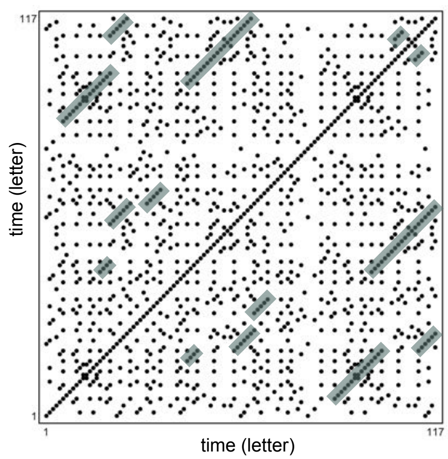


RQA Measures

Percent determinism (*%DET*): Percentage of the points on the plot that fall on diagonal lines (length > 1).

$$DET = \frac{\sum_{l=l_{min}}^{N} l\,P(l)}{\sum_{i,j}^{N} \mathbf{R}_{i,j}}$$

http://www.recurrence-plot.tk

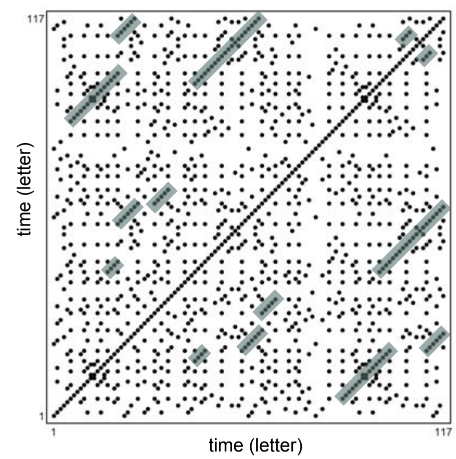


RQA Measures

Average diagonal line length (*MEANLINE*): Average length of diagonal lines on the plot excluding the line of incidence (length > 1).

$$L = \frac{\sum_{l=l_{min}}^{N} l\,P(l)}{\sum_{l=l_{min}}^{N} P(l)}$$

http://www.recurrence-plot.tk

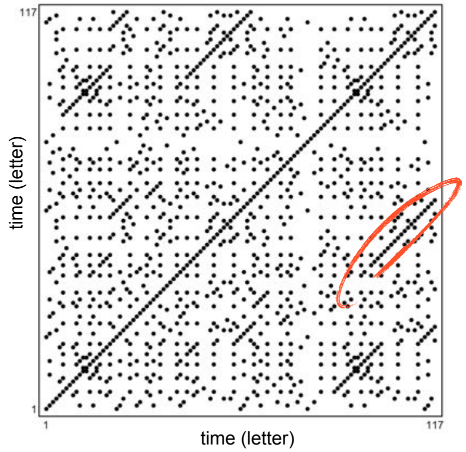


RQA Measures

Maximum line length (*MAXLINE*): The longest diagonal line on the plot (excluding the line of incidence).

$$L_{max} = \max(\{l_i;\ i = 1 \ldots N_l\})$$

http://www.recurrence-plot.tk

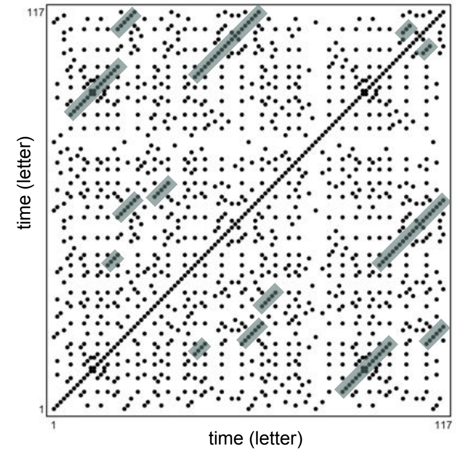


RQA Measures

Entropy (*ENTROPY*): The entropy of the distribution of diagonal lines on the plot (how much "disorder" is there in the sequences).

$$ENTR = -\sum_{l=l_{min}}^{N} p(l)\ln p(l)$$

http://www.recurrence-plot.tk

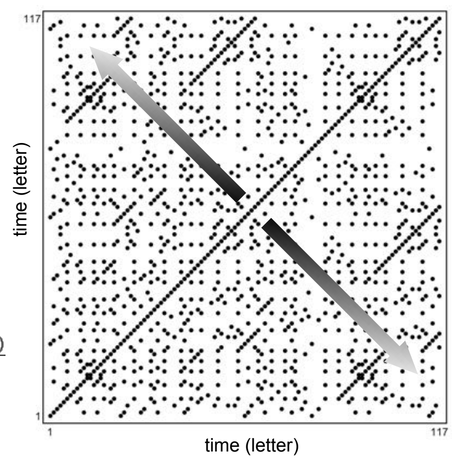


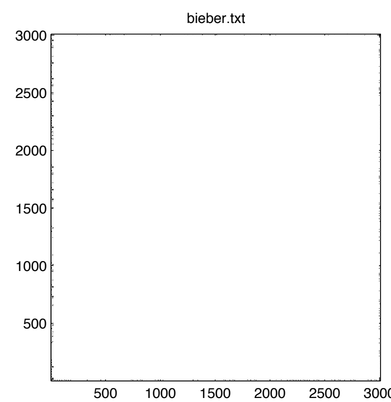
RQA Measures

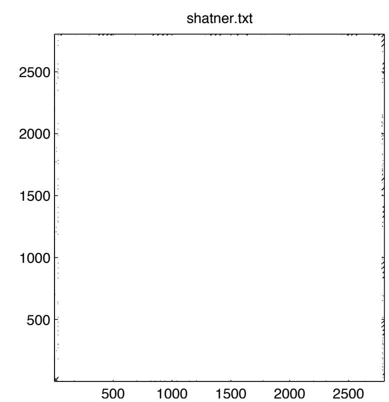
Trend (*TREND*): How the density of points changes as you move away from the line of incidence.

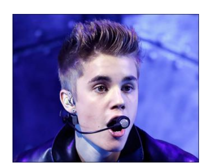
$$TREND = \frac{\sum_{i=1}^{\tilde{N}}(i - \tilde{N}/2)(RR_i - \langle RR_i \rangle)}{\sum_{i=1}^{\tilde{N}}(i - \tilde{N}/2)^2}$$

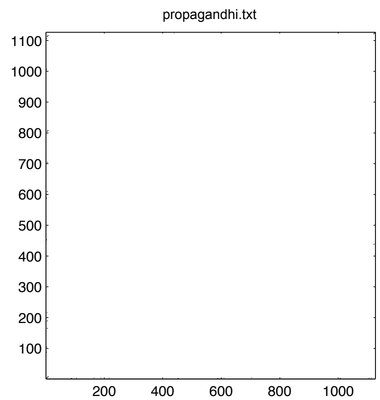
http://www.recurrence-plot.tk

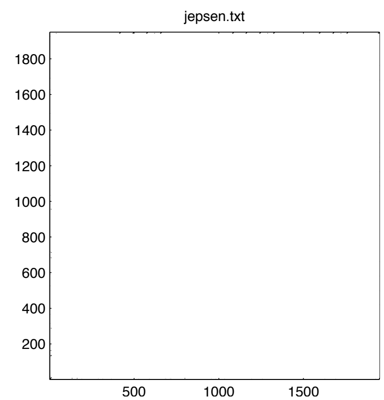


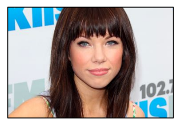










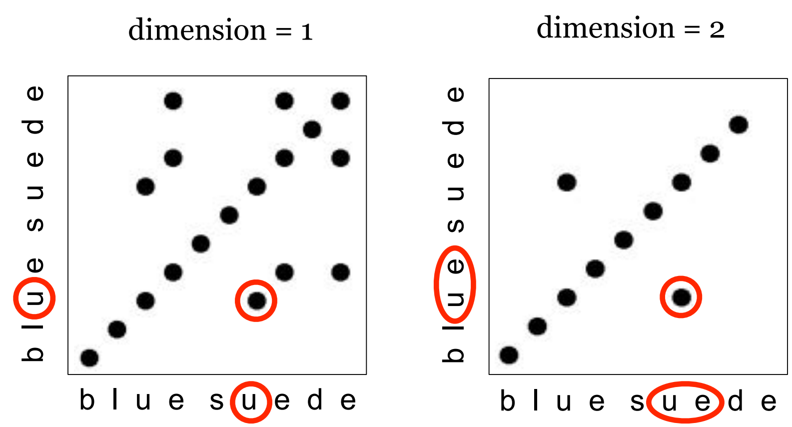


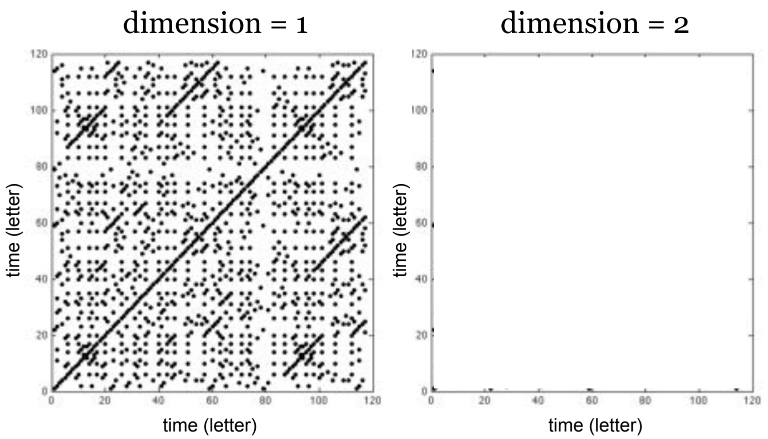
Three Important Concepts

- Parameters important for later…
  - embedding dimension,
  - delay,
  - radius
- These parameters will be *crucial* when we apply these RQA methods to *continuous* data (e.g., posture).

Embedding Dimension

- Interpretation in categorical data: how many states must match in order to count it as a recurrence.
- In previous analysis, dimension = 1
- What about 2? 3?
- *Window, vector, sequence, etc.*





It is possible to use dimension with categorical data, and the window size can be determined by theoretical or practical concerns (e.g., avoiding single-letter recurrence, as above).

Delay

- With most categorical data (behavior sequences, linguistic sequences, etc.) temporal ordering should probably be preserved (delay = 1).
- NB: Situation more complex with continuous data (tomorrow).

Radius

- The distance between units (or windows if dimension > 1) required in order to count (*i, j*) as a *recurrent point*.
- With category data: radius = 0.
- NB: Situation more complex with continuous data.

But, with categories…

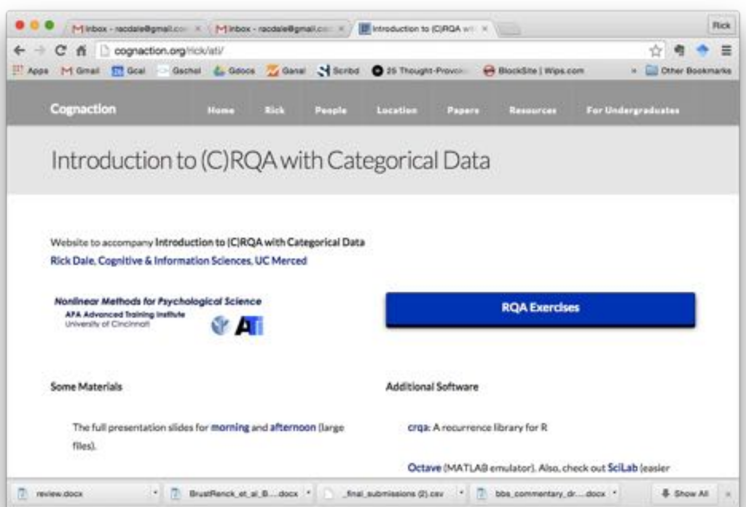
- With most data of nominal codes (e.g., letters, words, etc.), the following parameters often suffice:
  - Embedding dimension: 1
  - Delay: 1
  - Radius: .0001

Exercise: Quantify Plots!

See instructions on paper…

Coco & Dale, 2014

- With Dr. Moreno Coco
  University of Edinburgh
- **R** package for categorical recurrence
  (adaptable for continuous recurrence)
  - Basic (C)RQA measures
  - Diagonalwise recurrence (this afternoon)
  - Windowed recurrence measures



Nonuniformity in
behavioral dynamics

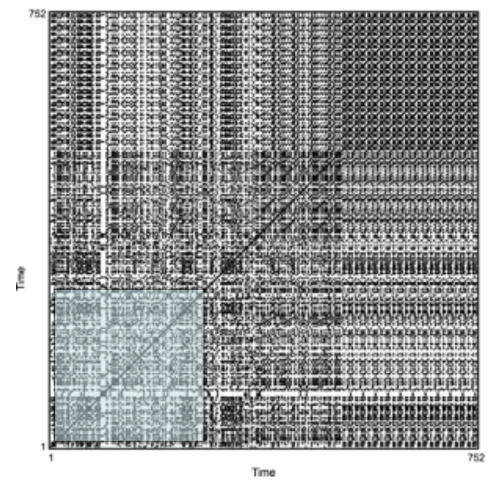
Behavioral "modes"
Stable, but temporary,
functional structures

Cattail Down

by MeWithoutYou

headed east out of st. paul,
we stopped for water.
rested in the cemetery,
watched the mississippi.
running out of food stamps,
found a bag along the footpath
off highway 61 filled with
what looked like marijuana.
(don't worry mom, we left it there)
hopped a grainrail out of pig's eye
toward milwaukee, a deer
between the tower and the tracks,
saw right through us.
said, "you don't know where you came from,
you don't know where you're going,
you think you're you-
you don't know who you are,
you're not you.

Cattail Down by MeWithoutYou

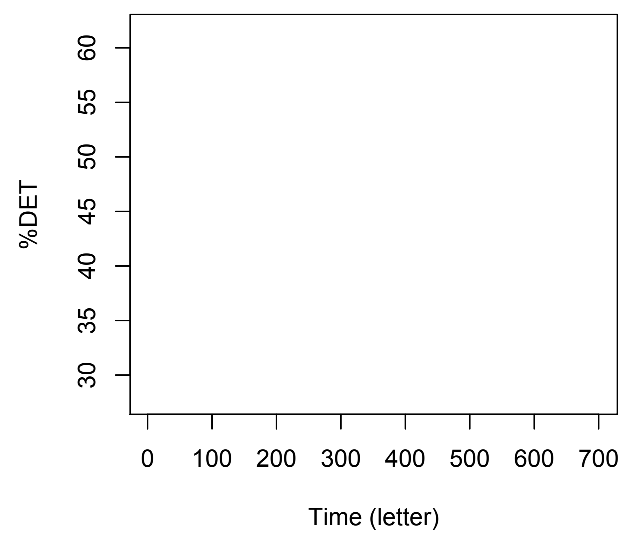


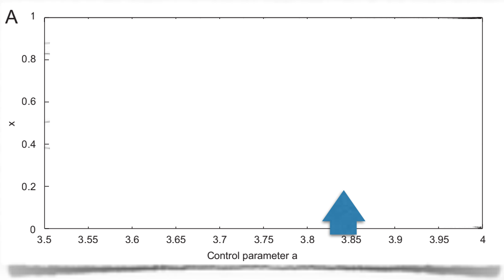
"Windowed recurrence"

Obtain several new time series of recurrence measures as they change across windows

Coco & Dale, 2014

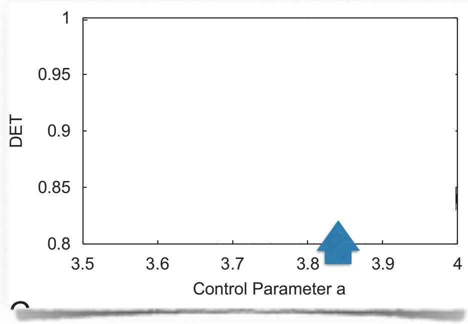
```
runcrqa(  ts1,
          ts2,
          par  )

       # par = list of parameters
       par = list(  type = 2, step =  1,
                    windowsize =  50,
                    lagwidth = 40,
                    method = "window",
                    datatype = "categorical",
                    thrshd = 8  )
```





$x(t+1) = ax(t)(1-x(t))$

each value computed with 800 data points!



Trulla et al., 1996, Physics Letters A; Marwan et al., 2007, Physics Reports

8 *Dale, Warlaumont, & Richardson*

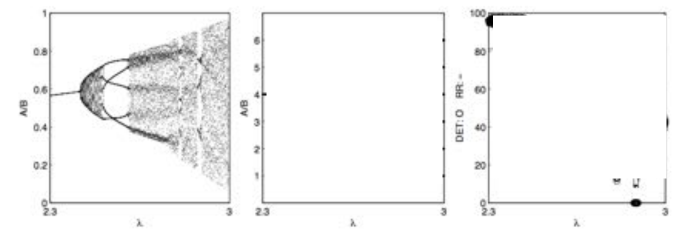


Fig. 5. Left panel shows the A/B states in the final 10 values of the 200 iterations at each λ; middle panel shows the coarse-graining of the system into 6 bins; right panel shows recurrence measures *DET* and *RR* changing relatively across bifurcation

Dale, Warlaumont, & Richardson, 2011, IJBC

Example Analyses

- Cross-linguistic...
  - ...stylistic complexity
  - ...encoding differences
- Within-language
  - ...text analysis
  - ...topic dynamics

Orthographic Structuring of Human Speech and Texts: Linguistic Application of Recurrence Quantification Analysis

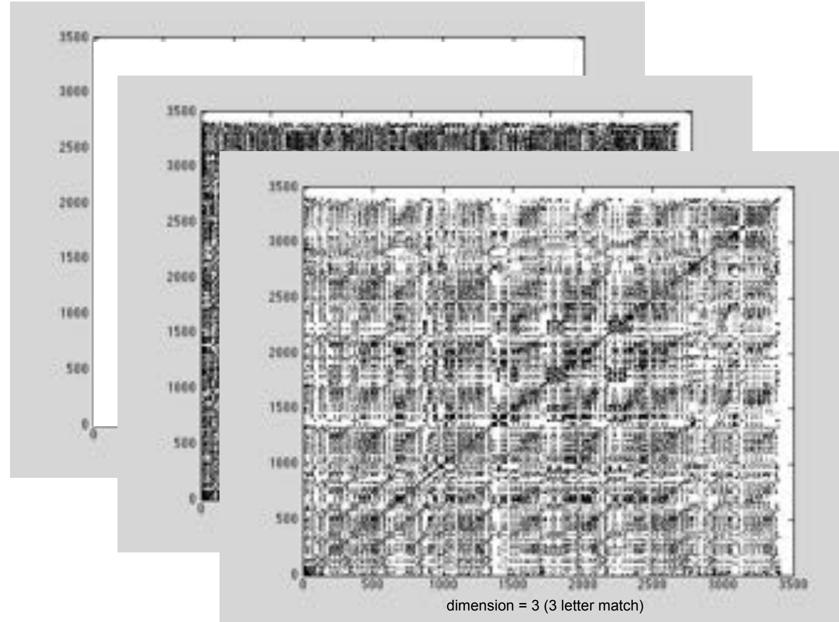
Franco Orsucci, Kimberly Walter* Alessandro Giuliani†
Charles L. Webber Jr.‡ Joseph P. Zbilut‡

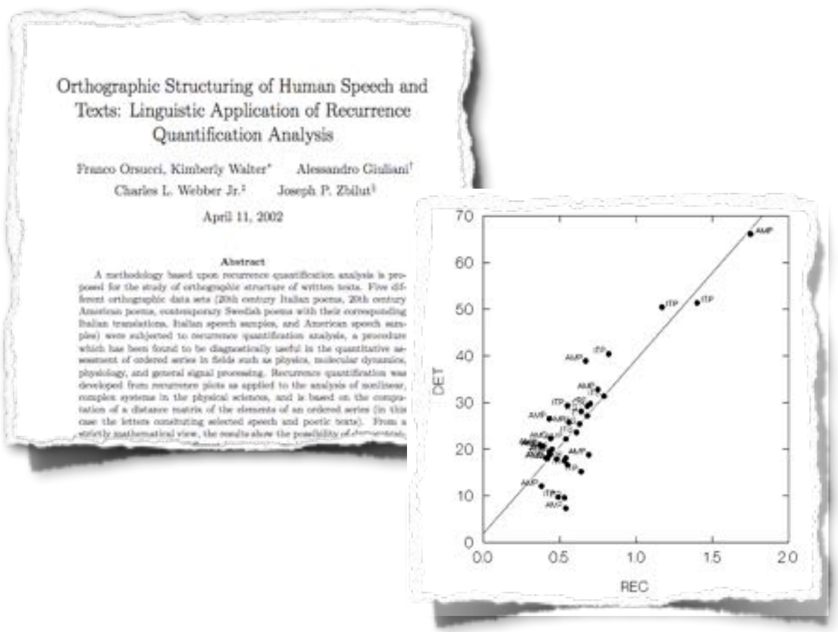
April 11, 2002

Abstract

A methodology based upon recurrence quantification analysis is proposed for the study of orthographic structure of written texts. Five different orthographic data sets (20th century Italian poems, 20th century American poems, contemporary Swedish poems with their corresponding Italian translations, Italian speech samples, and American speech samples) were subjected to recurrence quantification analysis, a procedure which has been found to be diagnostically useful in the quantitative assessment of ordered series in fields such as physics, molecular dynamics, physiology, and general signal processing. Recurrence quantification was developed from recurrence plots as applied to the analysis of nonlinear, complex systems in the physical sciences, and is based on the computation of a distance matrix of the elements of an ordered series (in this case the letters consituting selected speech and poetic texts). From a strictly mathematical view, the results show the possibility of

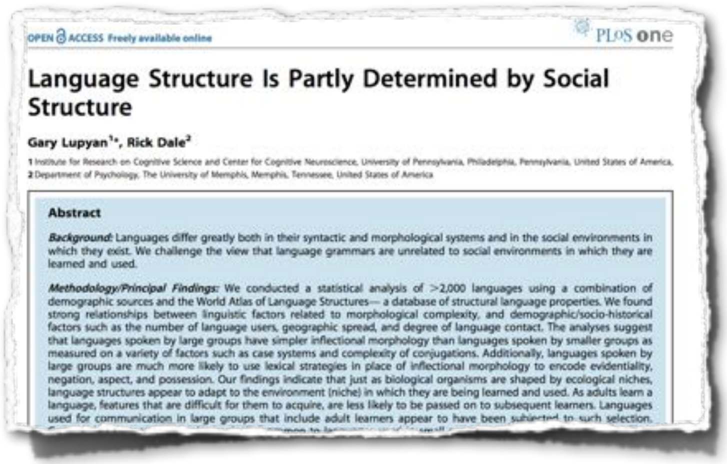
I am Sam I am Sam Sam I am That Sam-I-am That Sam-I-am! I do not like that Sam-I-am Do you like green eggs and ham I do not like them, Sam-I-am. I do not like green eggs and ham. Would you like them Here or there? I would not like them here or there. I would not like them anywhere. I do not like green eggs and ham. I do not like them, Sam-I-am Would you like them in a house? Would you like them with a mouse? I do not like them in a house. I do not like them with a mouse. I do not like them here or there. I do not like them anywhere. I do not like green eggs and ham. I do not like them, Sam-I-am. Would you eat them in a box? Would you eat them with a fox? Not in a box. Not with a fox. Not in a house. Not with a mouse. I would not eat them here or there. I would not eat them anywhere. I would not eat green eggs and ham. I do not like them, Sam-I-am. Would you? Could you? in a car? Eat them! Eat them! Here they are. I would not , could not, in a car You may like them. You will see. You may like them in a tree? d not in a tree. I would not, could not in a tree. Not in a car! You let me be. I do not like them in a box. I do not like them with a fox I do not like them in a house I do mot like them with a mouse I do not like them here or there. I do not like them anywhere. I do not like green eggs and ham. I do not like them, Sam-I-am. A train! A train! A train! A train! Could you, would you on a train? Not on a train! Not in a tree! Not in a car! Sam! Let me be! I would not, could not, in a box. I could not, would not, with a fox. I will not eat them with a mouse I will not eat them in a house. I will not eat them here or there. I will not eat them anywhere. I do not like them, Sam-I-am. Say! In the dark? Here in the dark! Would you, could you, in the dark? I would not, could not, in the dark. Would you, could you, in the rain? I would not, could not, in the rain. Not in the dark. Not on a train, Not in a car, Not in a tree. I do not like them, Sam, you see. Not in a house. Not in a box. Not with a mouse. Not with a fox. I will not eat them here or there. I do not like them anywhere! You do not like green eggs and ham? I do not like them, Sam-I-am. Could you, would you, with a goat? I would not, could not. with a goat! Would you, could you, on a boat? I could not, would not, on a boat. I will not, will not, with a goat. I will not eat them in the rain. I will not eat them on a train. Not in the dark! Not in a tree! Not in a car! You let me be! I do not like them in a box. I do not like them with a fox. I will not eat them in a house. I do not like them with a mouse. I do not like them here or there. I do not like them ANYWHERE! I do not like green eggs and ham! I do not like them, Sam-I-am. You do not like them. SO you say. Try them! Try them! And you may. Try them and you may I say. Sam! If you will let me be, I will try them. You will see. Say! I like green eggs and ham! I do!! I like them, Sam-I-am! And I would eat them in a boat! And I would eat them





Example Application

- Difference in the morphosyntactic dynamics of languages around the world.
- E.g., languages with highly complex morphology (e.g., Yagua, ...) vs. very simple morphology (e.g., English, ...).
- Encode information differently
- %DET ∝ population (Lupyan & Dale, 2010)



OPEN ACCESS Freely available online PLoS one

Language Structure Is Partly Determined by Social Structure

Gary Lupyan[1*], Rick Dale[2]

1 Institute for Research on Cognitive Science and Center for Cognitive Neuroscience, University of Pennsylvania, Philadelphia, Pennsylvania, United States of America,
2 Department of Psychology, The University of Memphis, Memphis, Tennessee, United States of America

Abstract

Background: Languages differ greatly both in their syntactic and morphological systems and in the social environments in which they exist. We challenge the view that language grammars are unrelated to social environments in which they are learned and used.

Methodology/Principal Findings: We conducted a statistical analysis of >2,000 languages using a combination of demographic sources and the World Atlas of Language Structures— a database of structural language properties. We found strong relationships between linguistic factors related to morphological complexity, and demographic/socio-historical factors such as the number of language users, geographic spread, and degree of language contact. The analyses suggest that languages spoken by large groups have simpler inflectional morphology than languages spoken by smaller groups as measured on a variety of factors such as case systems and complexity of conjugations. Additionally, languages spoken by large groups are much more likely to use lexical strategies in place of inflectional morphology to encode evidentiality, negation, aspect, and possession. Our findings indicate that just as biological organisms are shaped by ecological niches, language structures appear to adapt to the environment (niche) in which they are being learned and used. As adults learn a language, features that are difficult for them to acquire, are less likely to be passed on to subsequent learners. Languages used for communication in large groups that include adult learners appear to have been subjected to such selection.

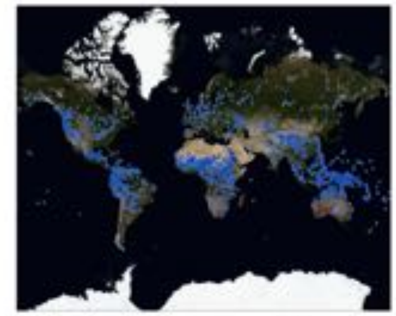
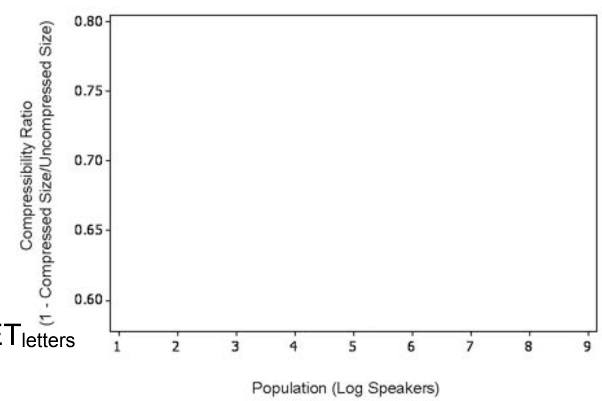


Figure 1. Geographic distribution of the 2,236 languages included in the present study.
doi:10.1371/journal.pone.0008559.g001

Analysis: 103 languages that use Roman alphabet and have translation in UN-UDHR

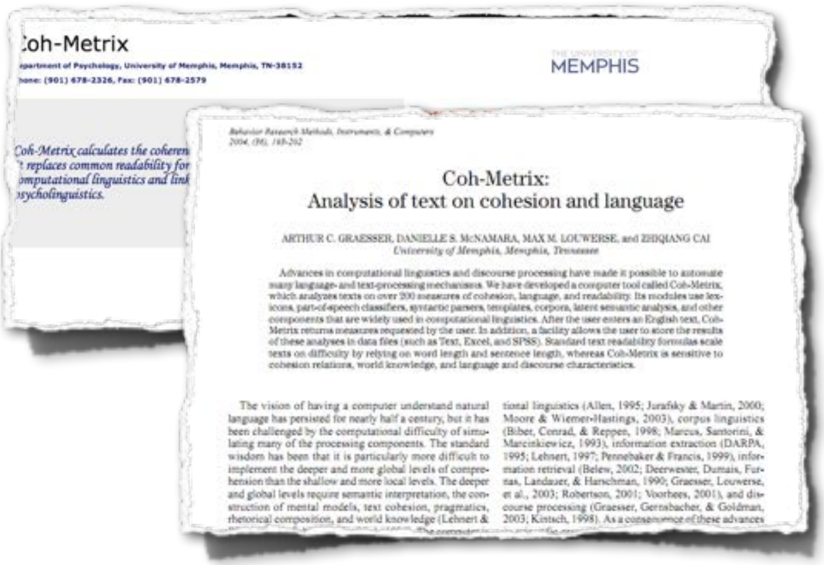


$\propto \%DET_{letters}$

Example Application

- Genre identification in educational data mining contexts.
- Do “textual dynamics” differ across history, science, etc. texts?
- Do these dynamic patterns correlate with accessibility, learning gains, etc.?

“cohesion”



Coh-Metrix

Department of Psychology, University of Memphis, Memphis, TN-38152
Phone: (901) 678-2326, Fax: (901) 678-2579

MEMPHIS

Coh-Metrix calculates the coheren... replaces common readability for... computational linguistics and link... psycholinguistics.

Behavior Research Methods, Instruments, & Computers
2004, 36 (2), 193-202

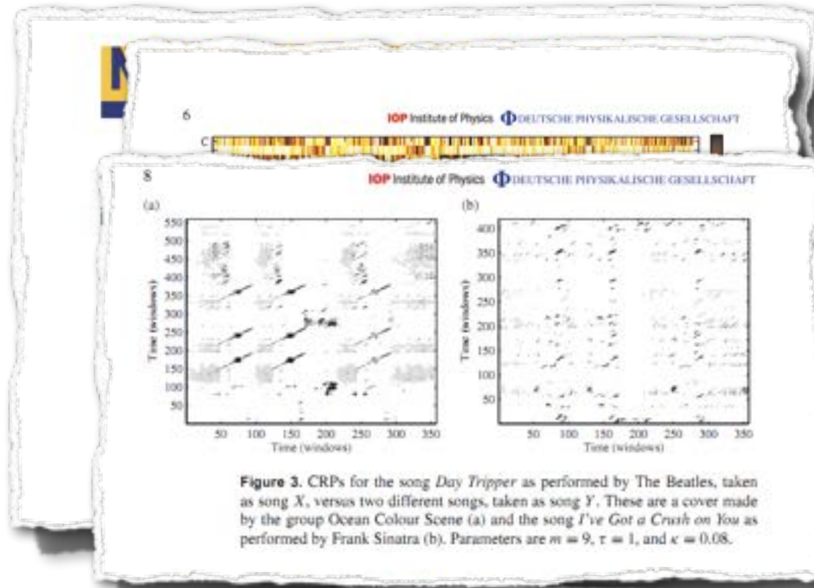
Coh-Metrix:
Analysis of text on cohesion and language

ARTHUR C. GRAESSER, DANIELLE S. McNAMARA, MAX M. LOUWERSE, and ZHIQIANG CAI
University of Memphis, Memphis, Tennessee

Advances in computational linguistics and discourse processing have made it possible to automate many language- and text-processing mechanisms. We have developed a computer tool called Coh-Metrix, which analyzes texts on over 200 measures of cohesion, language, and readability. Its modules use lexicons, part-of-speech classifiers, syntactic parsers, templates, corpora, latent semantic analysis, and other components that are widely used in computational linguistics. After the user enters an English text, Coh-Metrix returns measures requested by the user. In addition, a facility allows the user to store the results of these analyses in data files (such as Text, Excel, and SPSS). Standard text readability formulas scale texts on difficulty by relying on word length and sentence length, whereas Coh-Metrix is sensitive to cohesion relations, world knowledge, and language and discourse characteristics.

The vision of having a computer understand natural language has persisted for nearly half a century, but it has been challenged by the computational difficulty of simulating many of the processing components. The standard wisdom has been that it is particularly more difficult to implement the deeper and more global levels of comprehension than the shallow and more local levels. The deeper and global levels require semantic interpretation, the construction of mental models, text cohesion, pragmatics, rhetorical composition, and world knowledge (Lehnert & ...

tional linguistics (Allen, 1995; Jurafsky & Martin, 2000; Moore & Wiemer-Hastings, 2003), corpus linguistics (Biber, Conrad, & Reppen, 1998; Marcus, Santorini, & Marcinkiewicz, 1993), information extraction (DARPA, 1995; Lehnert, 1997; Pennebaker & Francis, 1999), information retrieval (Belew, 2002; Deerwester, Dumais, Furnas, Landauer, & Harshman, 1990; Graesser, Louwerse, et al., 2003; Robertson, 2001; Voorhees, 2001), and discourse processing (Graesser, Gernsbacher, & Goldman, 2003; Kintsch, 1998). As a consequence of these advances ...

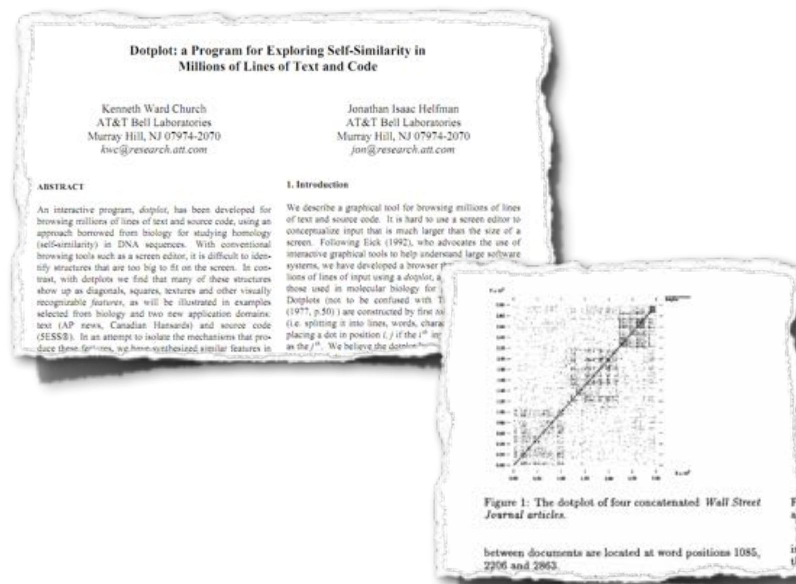


IOP Institute of Physics DEUTSCHE PHYSIKALISCHE GESELLSCHAFT

Figure 3. CRPs for the song *Day Tripper* as performed by The Beatles, taken as song *X*, versus two different songs, taken as song *Y*. These are a cover made by the group Ocean Colour Scene (a) and the song *I've Got a Crush on You* as performed by Frank Sinatra (b). Parameters are $m = 9$, $\tau = 1$, and $\kappa = 0.08$.

Example Application

- Identifying topic changes in very large texts.
- E.g., the pattern of points will change dramatically over a “topic” boundary.
- In natural language processing, related to methods using “dotplot.”



Dotplot: a Program for Exploring Self-Similarity in Millions of Lines of Text and Code

Kenneth Ward Church
AT&T Bell Laboratories
Murray Hill, NJ 07974-2070
kwc@research.att.com

Jonathan Isaac Helfman
AT&T Bell Laboratories
Murray Hill, NJ 07974-2070
jon@research.att.com

ABSTRACT

An interactive program, *dotplot*, has been developed for browsing millions of lines of text and source code, using an approach borrowed from biology for studying homology (self-similarity) in DNA sequences. With conventional browsing tools such as a screen editor, it is difficult to identify structures that are too big to fit on the screen. In contrast, with dotplots we find that many of these structures show up as diagonals, squares, textures and other visually recognizable *features*, as will be illustrated in examples selected from biology and two new application domains: text (AP news, Canadian Hansards) and source code (5ESS®). In an attempt to isolate the mechanisms that produce these features, we have synthesized similar features in ...

1. Introduction

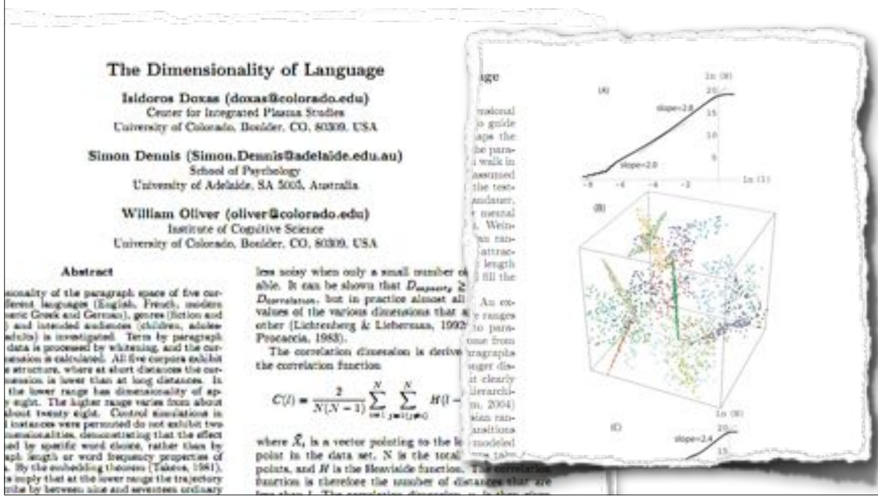
We describe a graphical tool for browsing millions of lines of text and source code. It is hard to use a screen editor to conceptualize input that is much larger than the size of a screen. Following Eick (1992), who advocates the use of interactive graphical tools to help understand large software systems, we have developed a browser ...

Figure 1: The dotplot of four concatenated *Wall Street Journal articles*.

between documents are located at word positions 1085, 2206 and 2863.

Reynar, 1994

Document Dynamics


The Dimensionality of Language

Isidoros Doxas (doxas@colorado.edu)
Center for Integrated Plasma Studies
University of Colorado, Boulder, CO, 80309, USA

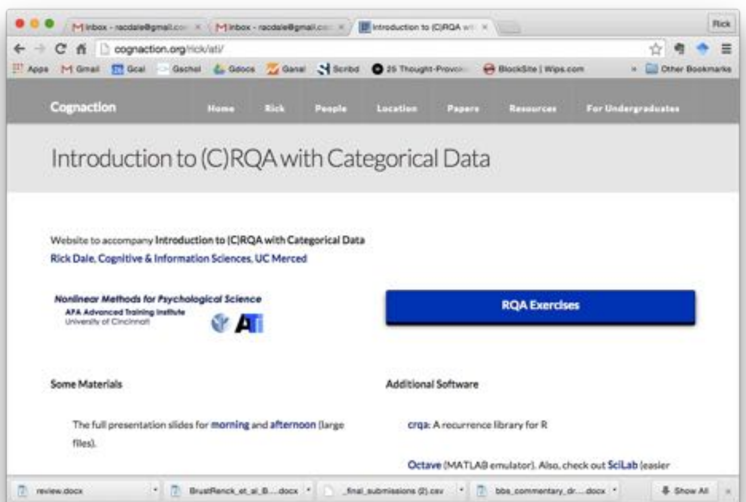
Simon Dennis (Simon.Dennis@adelaide.edu.au)
School of Psychology
University of Adelaide, SA 5005, Australia

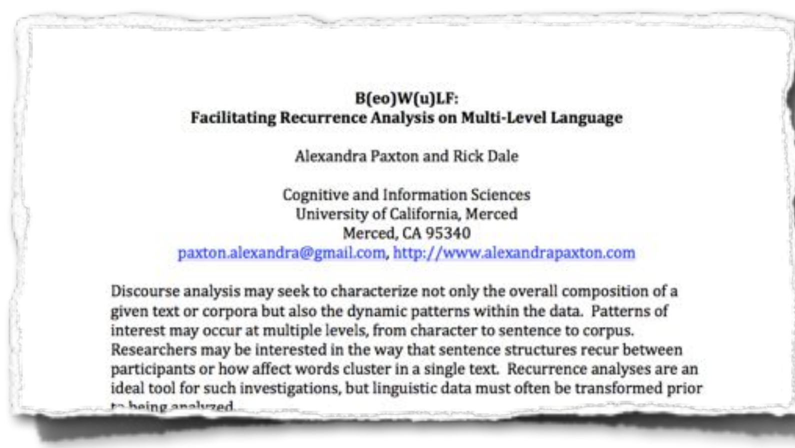
William Oliver (oliver@colorado.edu)
Institute of Cognitive Science
University of Colorado, Boulder, CO, 80309, USA

Abstract

Dynamic Discourse Analysis

- Collection of methods, models, etc. related to dynamic analysis of transcripts.
- Standard data format for dynamic discourse analysis: by-word long form (BWLF).
- **B(eo)W(u)LF**




B(eo)W(u)LF:
Facilitating Recurrence Analysis on Multi-Level Language

Alexandra Paxton and Rick Dale

Cognitive and Information Sciences
University of California, Merced
Merced, CA 95340
paxton.alexandra@gmail.com, http://www.alexandrapaxton.com

Discourse analysis may seek to characterize not only the overall composition of a given text or corpora but also the dynamic patterns within the data. Patterns of interest may occur at multiple levels, from character to sentence to corpus. Researchers may be interested in the way that sentence structures recur between participants or how affect words cluster in a single text. Recurrence analyses are an ideal tool for such investigations, but linguistic data must often be transformed prior to being analyzed

Appendix 1:
***Beowulf* Sample Text**

Modern English translation of *Beowulf*. Taken from an e-text version by Robin Katsuya-Corbet (released into the public domain July 1993).

LO, praise of the prowess of people-kings
of spear-armed Danes, in days long sped,
we have heard, and what honor the athelings
won!
Oft Scyld the Scefing from squadroned foes,
from many a tribe, the mead-bench tore,
awing the earls. Since erst he lay
friendless, a foundling, fate repaid him:
for he waxed under welkin, in wealth he
throve,
till before him the folk, both far and near,
who house by the whale-path, heard his
mandate,

who in former time forth had sent him
sole on the seas, a suckling child.
High o'er his head they hoist the standard,
a gold-wove banner; let billows take him,
gave him to ocean. Grave were their spirits,
mournful their mood. No man is able
to say in sooth, no son of the halls,
no hero 'neath heaven, - who harbored that
freight!
Now Beowulf bode in the burg of the
Scyldings,
leader beloved, and long he ruled
in fame with all folk, since his father had gone

canto	line	word	charnum	social	affect	relativ	time
1	1	lo	2	0	0	0	0
1	1	praise	6	100	100	0	0
1	1	of	2	0	0	0	0
1	1	the	3	0	0	0	0
1	1	prowess	7	0	0	0	0
1	1	of	2	0	0	0	0
1	1	people	6	100	0	0	0
1	1	kings	5	0	0	0	0
1	2	of	2	0	0	0	0
1	2	spear	5	0	0	0	0
1	2	armed	5	0	0	0	0
1	2	danes	5	0	0	0	0
1	2	in	2	0	0	100	0
1	2	days	4	0	0	100	100
1	2	long	4	0	0	100	100
1	2	sped	4	0	0	100	100
1	3	we	2	100	0	0	0
1	3	have	4	0	0	0	0
1	3	heard	5	100	0	0	0
1	3	and	3	0	0	0	0
1	3	what	4	0	0	0	0
1	3	honor	5	0	100	0	0
1	3	the	3	0	0	0	0
1	3	athelings	9	0	0	0	0
1	3	won!	3	0	100	0	0
1	4	oft	3	0	0	0	0
1	4	scyld	5	0	0	0	0
1	4	the	3	0	0	0	0
1	4	scefing	7	0	0	0	0

.

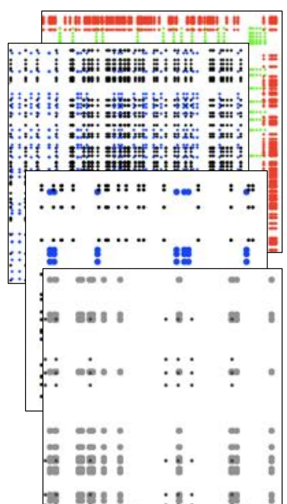
.

Visualizing Beowulf with B(eo)W(u)LF

RP
=
Cartesian product of time indices of noted events

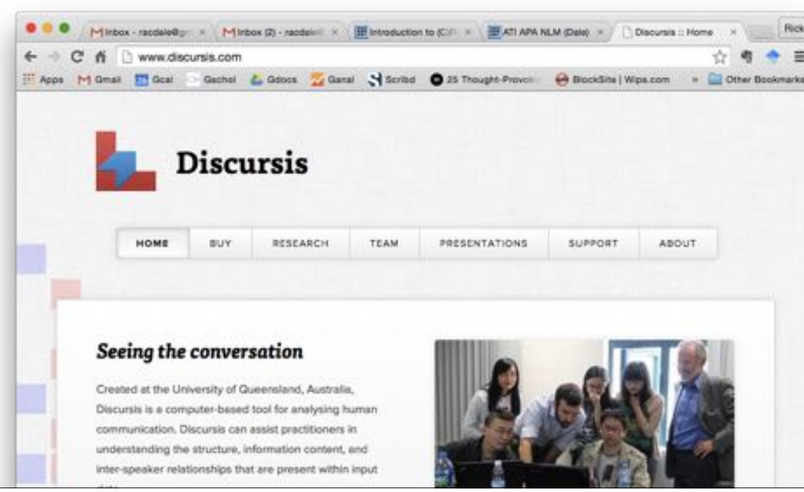
affect term indices = (1, 4, 423)

RP = affect terms × affect terms
= [(1,1),(1,4),(1,423),(4,1),(4,4),...]



akin to Hasselman, 2009

Discursis (Angus et al.)



Outline: Part 1

- Time series of higher-order states
  - Analysis of series of behavioral categories
- The recurrence plot (RP) and "textures"
- Quantifying the plot (RQA)
- Examples

Additional Exercises…

Afternoon: Cross RQA

fixations to objects

facial expressions

musical notes

gestural acts

words

musical phrases

conversation topics

discourse moves

sentence structures