Language in Time - Introduction

This short manual is designed primarily for researchers in the social sciences who work with text, transcripts, or related data, and wish to think about text in terms of dynamical systems. The manual is focused on the quantification and visualization of the temporal organization of these data sources. I will also describe the relationships between the approaches of this manual to techniques in computational linguistics, natural language processing, and other statistical domains. Recurrence quantification has conceptual and computational overlap with many techniques in those domains.

In this first section, we introduce the rudiments. Starting with some very simple text data, we will showcase how to import, convert, and plot some simple aspects of raw text. We then introduce the recurrent plot (RP).

In the second section, we introduce quantification of recurrence plots, called recurrence quantification analysis (CRQA). We will describe various measures to you, and identify how the measures are related to each other.

In the third section, we cover cross-recurrence quantification analysis (CRQA) and describe the corresponding quantities that accompany it.

In the fourth section, we describe windowed (C)RQA, which allows us to explore how these measures change over time in a text or related data set.

In the last section, we describe how (i) many of the measures of recurrence relate to quantities in NLP and related domains (such as n-gram modesl) and (ii) we show how (C)RQA can be adapted for the purpose of getting other NLP measures to describe a text in time. The overall goal is not to argue that recurrence is better or worse – but rather it is a descriptive-statistic framework that allows researchers to think in terms of time.

This manual uses the `crqa` library written in R primarily by Dr. Moreno Coco with a modicum of help from me, and with considerable inspiration from the impressive dynamics group at the University of Cincinnati (Jay Holden, Mike Richardson, Kevin Shockley, Mike Riley, and Guy Van Orden).

Why bother thinking in terms of dynamical systems? From one quite reasonable perspective, this is a surprising if not silly question. Cognitive systems are dynamical systems (Simon, 1990), so why not avail ourselves of this incredibly successful framework that has been applied all over the place, from physics to economics! Arguably, dynamical systems theory (DST) has already been widely successful in cognitive science, in particularly in motor control, including in speech. Extending it to higher levels of analysis, such as lexical or discourse levels, is a natural next step in understanding complex cognition in terms of dynamical systems. Indeed, there are plenty of folks doing so already.

Recurrence analysis is a useful tool to add to this conceptual framework. It provides a kind of lense through which we can talk about how a system tends to change in time: Is it highly regular, or disordered? How often does a cognitive system revisit sequences of behaviors? How do these sequences of behaviors relate to sequences in other cognitive agents?

We can think of texts in this dynamic manner. An obvious illustration of this is the distinction between diverse genres. Poetry is quite different from prose, which itself is quite different from, say, technical writing in manuals. The way that a writer opts to transition her texts through technical terms or jargon, or engage in playful repetition, is a kind of dynamic – ordering patterns that are ripe grounds for thinking dynamically.

Subtler possibilities are available though. Perhaps the dynamic descriptives that recurrence analysis supplies could distinguish among types of poetry (Orsucci et al., 2000), or perhaps it could shed light on readability of text, which is a particularly attractive topic at the time of this writing, in applied natural language processing. The dynamic lense of recurrence is a way to think about these systems and how they exert cognitive effects. I hope to convince readers of this to some extent in this short manual.