TOWARD A LIFESPAN METRIC OF READING FLUENCY

SEBASTIAN WALLOT
Department of Psychology, University of Cincinnati,
ML 0376, Cincinnati, OH 45221-0376, USA
wallotsn@mail.uc.edu

GUY VAN ORDEN
Center for Cognition, Action & Perception,
University of Cincinnati, ML 0376,
Cincinnati, OH 45221-0376, USA
guy.van.orden@uc.edu

Received February 22, 2010; Revised May 26, 2010

Much evidence suggests complexity in cognitive and motor task performances [Gilden, 2001]. The present study builds upon this work, treating reading of text as a kind of complex coordination or coupling between reader and reading conditions. Three self-paced reading conditions presented connect text in units of different sizes: word, phrase, or sentence units, and repeatedly measure times between spacebar presses to advance the text. The three conditions reveal different patterns across the data. These patterns were evaluated using fractal analyses and Recurrent Quantification Analyses to distinguish highly fluent readers, PhD candidates in English literature, from competent but less fluent undergraduate readers.

Keywords: Reading fluency; text reading; recurrence quantification analysis; fractal analysis.

“...to completely analyze what we do when we read would almost be the acme of a psychologist’s achievements... to unravel the tangled story of the most specific performance that civilization has learned in all its history.”

[Huey, 1908]

1. Introduction

Whether reading a newspaper, a ballot, or the fine print of a mortgage policy — competent reading is essential for success and well being in modern societies. Reading is the foundation of elementary and higher education, and the gateway to most kinds of employment, all the professions, and to a meaningful and independent life. Yet, despite its central place in modern society, more than a century of research on reading has failed to yield a metric for gauging reading fluency that can generally contrast different socioeconomic groups, different languages, or readers across different ages and at different levels of education.

On the one hand, to be called a fluent reader, a person must be well able to understand written text, to readily comprehend connected text and grasp its meaning. On the other hand, fluency also implies an effortlessness in the act of reading — written text is understood by the reader fairly easily, as a fluent reader can progress through a text quickly and flexibly. So, a minimal definition...
of reading fluency might encompass the terms comprehension, effortlessness, flexibility, and — superficially — reading speed.

The best predictors of reading fluency are rapid naming or reading practice, yet reading fluency is typically studied after it is achieved, so it is unclear how acquisition of fluency is related to its predictors. In this regard, reading fluency is not sufficiently gauged by reading speed, nor summary comprehension scores, or even vocabulary because these measures do not reliably pick out fluent readers. For example, hyperlexic reading savants can read aloud quickly yet understand nothing, whereas compensated developmental dyslexics exist who cannot accurately read a text aloud but have good text comprehension (e.g. [Van Orden et al., 1990, 2001], are reviews).

Word identification in reading is crucial for reading acquisition, but it is not wholly predictive of reading fluency. Also, while oral reading fluency is predictive of reading comprehension of elementary and junior high school readers, the relationship is weak for adults [Fuchs et al., 2001]. The nature of adult reading fluency remains largely a mystery. A critical gap in this respect is the lack of a metric that can capture the pattern of flow in silent reading. It is this concern that inspired the present study. Here, we begin to examine whether time-series measures of Recurrence Quantification Analysis (RQA), together with fractal analyses, may characterize adult fluency and close this gap.

The present study departs somewhat from previous efforts to study reading and is in that regard admittedly exploratory, but also innovative in the use of RQA to examine reading of connected text. RQA allows a quantitative analysis of time ordered repeated measures, and we use RQA to analyze repeated measures of self-paced reading times, comparing performance between PhD candidates in English Literature and freshman undergraduates. The RQA quantities are estimated from a recurrence plot of the data-series. The recurrence plot is obtained through a method of time-delayed phase-space reconstruction and projection into higher dimensional space [Mané, 1981; Takens, 1981]. Several quantities based on the recurrence plot can be estimated, revealing information about the dynamics of the data series, including stability, complexity, and transitions between epochs of chaos and order (e.g. [Marwan et al., 2002; Webber & Zbilut, 2005]).

To complement the analysis of text reading that follows, we also employ methods of fractal estimation in time ordered reading data — many standard experimental procedures used to examine cognitive and motor performance, including reading performance, have produced data that exhibit fractal signatures, as statistically self-similar fractal time (e.g. [Gilden, 2001; Holden, 2002]). Additional evidence exists that the fractal dimensions of cognitive and motor performance, estimated with scaling exponents, change with learning, development, skill acquisition and task difficulty ([Van Orden et al., 2011] is a review). Thus fractal analyses are of potential interest to examine reading fluency as well.

In the next section we discuss a problem that concerns the measurement of reading fluency on a characteristic scale, such as reading speed. Then we describe the present method of a self-paced reading task, the data that were gathered, and the results of traditional reading time analyses. Following that, we apply fractal estimation and recurrence quantification analysis to distinguish the different modes of text presentation in simple or multiple readings of the same text by adult readers with different levels of college education.

2. The Measurement Problem

All previous metrics to evaluate reading performance have assumed that a characteristic scale of reading performance exists. The more abstract logic that requires this assumption is a logic of concatenated effects [Michell, 1999]. Concatenated effects are those that share a common timescale and that unfold in a time ordered fashion on that timescale. Thus, to read aloud a printed word, the components and subcomponents of sensation, perception, reading, and articulation are assumed to concatenate their effects like a row of time-ordered falling dominos, each affecting the next in its turn.

The abstract logic of concatenated effects justifies ordinary contrasts between the average times to read aloud, for example. Yet reading aloud is done by a brain and body in which ubiquitous feedback processes compose the nervous system, and in which simultaneous afferent and efferent processes control the eyes and mouth at the sensorimotor periphery of the reader. Moreover, feedback dynamics self-organize rapid and even instantaneous changes across the nervous system [Kelso, 1995], which puts the nervous system on an entirely different range of timescales than the time course of repeated measurements of behavior [Van Orden et al., 2011].
If this organization may dictate assumptions, then the measurement problem may be less one of timing dominoes, and more one of characterizing loops and levels of interdependence in entangled cognitive phenomena [Bell, 1999].

The linguistic structure of a written text may also motivate loops and levels of interdependence in reading behavior. When words are read in succession they influence each other in complicated ways. Conventionally, exacting laboratory controls are used to strip away the complication, and simple laboratory tasks and measurement procedures are chosen to minimize task effects. Nonetheless, even the simplest reading tasks discover complicated interactions in which the interacting factors may either amplify, reduce, or even eliminate each other’s effects. Merely presenting two words in succession can be enough to create complicated interactions, due to their relation in meaning, spelling, pronunciation, or in more than one of these.

For example, a key-press response in a simple reading task, to indicate that the word pepper is indeed a word (with respect to English spelling) will be about 48 msec faster (on average) if pepper is preceded by the word salt (compared to a control condition that precedes pepper by an unrelated word such as bon [Neely et al., 1998]). This is a large effect, given that a single word is easily read within about 200 msec from first sight. However, if salt is presented twice in succession, in the same task, just before pepper appears, the large facilitation effect vanishes [Balota & Paul, 1996; Neely et al., 1998]. If this was merely an isolated oddball finding then it might be of little consequence, but all simple reading tasks reveal such complicated nonlinear patterns of interaction among the factors that reading scientists study (see [Bargh, 2006; Pickering & Ferreira, 2008], for reviews and discussions).

A slight variation in a laboratory reading context — the introduction of a new factor into an experimental paradigm — and the entire landscape of reading effects can change [Van Orden & Kloo, 2005]. Nonetheless, the default assumption in all this research has been that the impact of a contributing factor to reading, whether the property of a text or of the reader, will be additive and proportional to its magnitude. A less skilled reader should require proportionally more time and effort on the same text, and a more difficult text should require proportionally more time and effort to read. But reading may comprise a heterarchy of overlapping and interacting capacities, such that different combinations may even compensate for deficiencies, insofar as reading speed or comprehension are concerned.

The measurement problem as we see it may stem from the absence of a characteristic scale of reading. Related concerns have arisen in other areas of cognitive research where compensatory data procedures have been proposed (e.g. [Faust et al., 1999; Fisher & Glaser, 1996]). The procedures are patches or workarounds within the common practice, however, that skirt narrow problems, none proving sufficiently reliable or general to replace the common practice. An alternative would be to use scale-free measures [Gottlob, 2007]. Scale-free measures are invariant under scale transformation. For example, normative measures, such as z-scores, become scale-free after one subtracts from each score the population mean and divides by the standard deviation. But normative scales are not inherently scale-free and the same assumption about concatenated effects must yet be true for conventional analyses.

Nonetheless, the suggestion to use scale-free measures appears to be the right suggestion, so long as the measured phenomena are inherently scale-free. This rationale stands behind our use of scale-free fractal analyses and Recurrence Quantification Analysis (RQA) and allows us to relax the assumptions of linearity and proportionality of effects. The nonlinear measures can be highly robust [Riley & Turvey, 2002; Webber & Zbilut, 1994, 1996, 2005; Zbilut & Webber, 1992] and have been used successfully to evaluate other cognitive and behavioral phenomena where multiple factors interact in a presumably nonlinear fashion (e.g. [Shockley et al., 2007; Riley & Clark, 2003]), as we hypothesize about reading.

3. The Reading Data
We collected data in a self-paced reading task, much used in psycholinguistic studies of language and reading. In this self-paced reading task, the reader advances through a text by pressing the spacebar of a keyboard. The times between spacebar presses estimate reading times of the text units. The reading times we collected were sampled from a single story divided into text units that were either word units, phrase units, or sentence units. Each time a participant reader pressed the spacebar, the next unit of text in sequence appeared on the computer screen, so text accumulated on the
screen either word by word, phrase by phrase, or sentence by sentence.

The participant readers were 24 (PhD candidate) graduate students of English Literature, and 24 freshman undergraduates, attending the University of Cincinnati. All reported being unfamiliar with the story they read and had normal or corrected to normal vision. We assumed that the graduate students were more fluent readers overall, but the specific expectation with respect to the present exploratory study was that they may exhibit reliable differences in reading performance from undergraduates. Additionally, 12 of the graduate students and 12 of the undergraduates read the story more than once. Text repetition effects are among the most reliable and pronounced effects in laboratory reading tasks (e.g. [Gilden, 2001; Holden, 2002]) and reports exist of changes in fractal dimension with task repetition (e.g. [Wijnants et al., 2009]). Differences in reading skill between graduate students and undergraduates, as well as the repetition manipulation supply a kind of backdrop of ordinary reading effects within which to situate the results of the new analyses. Again, we are mostly concerned with producing standard reading performances that can be examined further using estimates of scaling exponents (fractal dimensions) and also Recurrence Quantification Analysis.

The story read by participants in our investigation was *The Arelis Complex* by DeGrado [2003]. This story is about the fictive intergalactic politics of the Arelians. The Arelians were nearly wiped out previously in a conflict with another civilization. Star ship officer Drakh Norh, charged with the safety of the Arelian’s home planet, oversees the interception of an unmanned alien space vessel. The vessel is traced back to a “blue planet” whose inhabitants had been deemed incapable of such technology. Officer Drakh concludes eventually that the alien race may develop into a threat and leads a war fleet to destroy the inhabitants of the “blue planet”. The attempt fails, however, leaving the Arelians facing an uncertain future.

The Arelis Complex consists of 13,930 words, 1,696 phrases, and 1,042 sentences. Each phrase was a sequence of words demarcated by any punctuation (comma, period, colon, semicolon, parenthesis, question mark, exclamation mark, or dash). Text units were presented in Times New Roman font (13.5 pt.) on a standard computer monitor, paced by the reader’s spacebar presses. The graduate and undergraduate participants read the story while being seated before the computer monitor, pressing the spacebar to bring text to the screen. Figure 1 illustrates the self-paced reading procedure.

Reading the story took between 30 min and 80 min, depending on the individual. Participants were compensated for their participation in class credit and cash. After reading the story each reader completed an exam, of which they had been forewarned, to assess story comprehension and memory. The exam required a written summary of the story plot, indicating characters and their roles in the story, plus a multiple-choice sentence completion

---

**Fig. 1.** Illustration of the self-paced reading task for sentence-units.
4. Average Reading Times

We expected that graduate students could be faster readers on average than undergraduates (239 msec versus 271 msec for word units, 1767 msec versus 1915 msec for phrase units, and 2534 msec versus 3655 msec for sentence units). Also readers who read the story only once would read more slowly on average than those who read it multiple times (267 msec versus 243 msec for word units, 2137 msec versus 1535 msec for phrase units, and 3682 msec versus 2507 msec for sentence units). And we expected that the average reading times for word units would be much shorter than the average reading times for multi-word phrase or sentence units ($M = 255$ msec; $SD = 50$ msec, for words, $M = 1835$ msec; $SD = 591$ msec, for phrases, and $M = 3095$ msec; $SD = 1402$ msec, for sentences, $F(2,45) = 13.62, p < 0.001$). These expectations were all borne out.

A more informative effect appears in Fig. 2 as a three-way interaction among graduate versus undergraduate X phrase versus sentence units X one versus multiple readings ($F(1,24) = 4.79, p < 0.05$). The benefit to graduate students’ average reading time of reading the story more than once is most pronounced when they are presented with phrase units ($F(1,12) = 2.20, p = 0.16$). In contrast, undergraduate readers benefit from multiple readings in their average reading times to sentences ($F(1,12) = 2.65, p = 0.13$).

These simple main effects of one versus multiple readings, in the phrase- and sentence-unit conditions, would likely have been statistically reliable with a larger sample size of readers. At present, however, these contrasts between one and multiple readings are between different participants reading the story with different text units than their first reading, and any statistical contrast comparing different participants in different conditions will have less statistical power than a contrast between the same individuals in the different conditions.

The graduate versus undergraduate distinction, our speculative fluency distinction, appears to be most salient in differences between the phrase-unit and sentence-unit conditions. The salient differences between these larger text unit conditions is replicated and clarified by the nonlinear analyses that follow.

5. Power Spectral Analysis

Figure 3 illustrates a spectral analysis of one participant’s repeatedly measured word-unit reading times. The outcome of the analysis is a log-log spectral plot of the decomposed data series. The slope...
Fig. 3. Example of the estimation of scaling in word unit response times. (a) 8192 word reading times (normalized). (b) Log-log spectral plot with least square regression line.

of the fitted, least square, regression line estimates a scaling relation that can be used to quantify the fractal dimension of the pattern of variation across the data. The negative slope \(-\alpha = -0.66\) in Fig. 3 corresponds to a scaling exponent of \(\alpha = 0.66\), which lies between the scaling exponents of idealized white noise (slope = 0, scaling exponent of \(\alpha = 0\), fractal dimension = 1.5) and idealized pink noise (slope = -1, scaling exponent of \(\alpha = 1\), fractal dimension = 1.2).

Scaling exponents of human performance have been widely studied, providing a background in which to situate new outcomes. For example, self-paced tapping of a key alone, without reading, yields scaling exponents close to that of Fig. 3 on average, though possibly a bit closer to \(\alpha = 1\). Repeatedly measured simple reaction times yield scaling exponents very close, on average, to \(\alpha = 0.66\) (Mean \(\alpha = 0.66, SD = 0.22\), see [Van Orden et al., 2003]). Were we to reorder the words from the story, present them individually in a random order, and record the naming times word by word, generating a trial series of naming times, the average scaling exponent would be closer to the \(\alpha = 0\) of white noise (Mean \(\alpha = 0.29, SD = 0.04\), see also [Van Orden et al., 2003]). These absolute scaling exponents tend to be relatively uninformative by themselves, however. Human performances produce a wide range of scaling exponent values that depend upon tradeoffs among the integrity of system dynamics, the task demands, the participant skills, and the availability of other constraints. Tradeoffs make it difficult to interpret absolute values of exponents, and encourage manipulations to induce changes in scaling exponents that are more revealing in their direction of change [Van Orden et al., 2011].

Spectral analyses like that portrayed in Fig. 3, together with standardized dispersion analyses and detrended fluctuation analysis, were used to estimate scaling exponents and fractal dimension for each time series. The methods all entail workarounds because they are linear methods, at heart, that are used here to analyze nonlinear phenomena. Each method has weaknesses that another method compensates for. When all goes well, the outcomes of the three analyses are redundant, as they were in the present data. Because they were redundant, we report the detailed results of the scaling exponents from the spectral analyses exclusively. Data were preprocessed only as necessary, as suggested in [Holden, 2005].

5.1. Scaling exponents across text units

Figure 4 presents summary results of the spectral analyses. As is apparent in Fig. 4, when the story was presented in word units, the corresponding data series exhibited \(1/f^\alpha\) scaling exponents much closer to \(\alpha = 1\) than in the other conditions. The \(\alpha = 1\) scaling relation was less present in sentence reading times and even less so in the phrase reading times, and a reliable trend was observed across these conditions, away from \(\alpha = 1\) and toward \(\alpha = 0\) (\(F(3,36) = 24.60, p < 0.001\)).

We conducted surrogate data analyses to test whether the scaling exponents we obtained from the reading time data differed reliably from \(\alpha = 0\). Surrogate data were created by shuffling the order of
the data-series, thus destroying the temporal structure of the data while preserving mean values and standard deviations. We shuffled each participant’s data once to create yoked pairs of surrogate versus original data series to be contrasted with each other. A statistical test using the scaling exponents of the original data, compared to their surrogates, revealed an interaction effect between reading units (i.e. word, phrase, sentence) and ordered versus surrogate data series ($F(2, 36) = 21.05$, $p < 0.001$).

Planned contrasts found that reading times in the word unit condition produced relatively higher scaling exponents (Mean $\alpha = 0.52; SD = 0.20$), compared to other conditions. The substantially higher scaling exponents indicate a substantial difference in how the word unit pacing affected reading. One possible source of different scaling exponents would be inherent differences among the different text units. Although the word and sentence lengths from *Arelis Complex* lie well in the range of standard English prose [Fengxiang, 2007; Sigurd et al., 2004] they may yet differ in variety from each other. The various words and sentences differ in length, for example, and sentences vary more unsystematically than words. Word units vary by number of letters, and words in the story ranged from one letter in length to 16 letters ($M = 4.56; SD = 2.38$). Sentence units vary by their number of words, and sentences in the story varied from one word to 50 words in length ($M = 13.36; SD = 7.89$).

Reading times are partly due to the physical length of a given text unit. If the wider variation across sentence lengths is also more unsystematic, it would yield less systematic reading times and scaling exponents closer to $\alpha = 0$ [Holden et al., 2010]. Effectively, unsystematic changes in reading times are perturbations to the spectral analyses that bias the estimates of scaling exponents in the direction of random white noise. We tested for differential bias effects of variation in length for word, phrase, or sentence lengths by contrasting the text-ordered counts of number of letters in words, and number of words in phrases and sentences, conducting spectral analyses on these data-series. The scaling exponent for the number of letters in each word was $\alpha = -0.03$, the number of words in phrases yielded a scaling exponent of $\alpha = 0.15$, and sentences yielded $\alpha = 0.18$. None of these values could be distinguished reliably from $\alpha = 0$ in surrogate analyses conducted using 20 surrogates of each data series. In contrast, the word and sentence reading times yielded scaling exponents reliably different from their surrogates and $\alpha = 0$ (all $t(0) > -3.89$, $p < 0.05$). Thus, the pattern of scaling exponents across these text units is not due to differences in surface variation of length.

There are surely differences in addition to length among the three text units and those differences could also affect the pattern of reading times. As we already noted, the absolute estimates of scaling exponents in human performance are not usually informative by themselves due to confounded sources of bias in the estimate. We next turn to the more informative patterns of change in scaling exponents associated with our manipulations.

5.2. Differences among scaling exponents

In the present study, readers entrain to the structure of a story. Story structure includes any instance or source of systematic or unsystematic change in
the progression of the story, as well as in the presentation conditions of the story. In the latter case, the previous variation in the unit lengths of words, phrases and sentences becomes a powerful source of unsystematic change. Unit lengths could not be distinguished from random noise in the previous surrogate analyses, for example, and sources of random noise will greatly reduce scaling exponents toward $\alpha = 0$ [Holden et al., 2010].

When the present analyses discover scaling exponents different from $\alpha = 0$ (and soon the outcomes supporting recurrence, determinism, and attractor strength) they do so against a background of random perturbations that might have overwhelmed any structure in the performance. In this light, the scaling exponents for phrase unit reading times only differed reliably from $\alpha = 0$ for graduate readers, reading the story more than once ($t(6) = -4.02, p < 0.05$). Otherwise the variation across reading times to phrase units did not differ from white noise, so perhaps phrases are suboptimal linguistic units for reading, producing unsystematic reading times and perturbing the spectral analysis. Among other things, phrases are relatively more ambiguous in meaning, for instance, compared to sentences.

While phrases may be suboptimal reading units, graduates students nevertheless produced statistically reliable fractal structure in multiple readings. Other work has shown that practice and improved performance on novel cognitive or motor tasks moves scaling exponents away from the $\alpha = 0$ of random patterns and toward the $\alpha = 1$ of long-range correlated, fractal behavior as practice appears to do for phrase units. However, only the elite PhD candidates in English Literature show the benefit of practice, which they also produced in their faster reading times to phrase units in multiple readings. Thus we may see in this result the greater capacity of the graduate student readers to entrain to story structure presented in an unusual form, phrase by phrase. It would be interesting to explore further the number of readings of an identical text before undergraduate readers show this same degree of entrainment to phrase unit presentations.

We asked one undergraduate to read the story four times, each time presenting phrase units of text. The consequent changes in scaling exponents were correlated with number of readings in the right direction ($r = 0.48$), and the maximum of $\alpha = 0.13$ was in the ball park of the graduate student scaling exponents in multiple readings, but surrogate analyses did not distinguish either of her four readings reading times from random noise.

The sentence and word unit conditions did not produce reliable differences in scaling exponents, neither for graduate versus undergraduate

![Fig. 5. Average scaling exponents and standard errors for (a) word units, (b) phrase units and (c) sentence units. Scaling exponents are grouped by readers (PhD candidates and undergraduate students) and number of readings.](image)
readers nor for single versus repeated readings (all $F(1, 12) < 2.64$ — see Fig. 5). Nonetheless the average scaling exponents of word unit reading times change in the same direction, toward $\alpha = 1$, for both graduate students and undergraduates. Scaling exponents to sentence units, by contrast, trend away from $\alpha = 1$ toward $\alpha = 0$ (although undergraduates vary widely in the scaling exponents of multiple readings).

Scaling exponents appear to gauge the coordination of participants with the temporal demands of tasks, and coordinations more in sync with a new task are also closer to $\alpha = 1$ [Holden et al., 2010; Kloos & Van Orden, 2010]. In the present self-paced reading task, enhanced coordination could mean at least two different things, an enhanced capacity of participants to entrain to the linguistic structure of the story, or an enhanced strategy of spacebar pressing to advance through the text.

Entrainment may apply to graduate students who produce a more reliable fractal pattern in multiple readings. Word units on the other hand may encourage strategic spacebar pressing, a hypothesis that will become more salient across the remaining analyses. One possibility is that participants can usually read individual words faster than they can advance the text by pressing the spacebar. If so, then it is spacebar presses, not words, that are rate-limiting as participants advanced the story word by word.

6. Recurrence Quantification Analysis (RQA)

The unprocessed data were submitted to RQA. RQA parameters were estimated for each data set individually. Mutual-average information was used to estimate the embedding dimension parameter, which ranged from three to 11, using recommended procedures from Webber and Zbilut [2005].

The radius parameters for each individual readers’ data set was adjusted to produce 2% recurrence, which makes comparisons between readers possible.

Figures 6-8 portray results from each of the reading units for the RQA quantities of $%\text{DET}$ and MAXL. $%\text{DET}$ is computed from the number of diagonally adjacent points in the recurrence plot divided by the total number of recurrent points, and estimates the degree of order in the data, the extent to which readers fall into rhythm of sorts across text units. MAXL is the length of the longest chain of adjacent, diagonal, recurrent points and estimates the strength of the strongest attractor. MAXL estimates the degree of stability in the reading times and in the coordination between the participant and the self-paced reading task.

Reliable differences were found between reading times to the word units, phrase units, and sentence units, respectively, in the overall level of $%\text{DET}$ ($F(2, 45) = 147.35, p < 0.001$) and MAXL ($F(2, 45) = 49.31, p < 0.001$). Reading times to word units exhibited a much higher percentage of deterministic structure and stronger attraction to stable reading times compared to phrases and sentences, which may be further evidence that word units induce a different reading strategy than phrases or sentences.

6.1. Word units

Despite the fact that word properties change from word to word, and that the story context changes as each new word appears, the estimated reading times of word units appear relatively uniform in their pattern over time. Figure 6 displays the results for word units of $%\text{DET}$ and MAXL and the high values for $%\text{DET}$ and MAXL in word reading suggest a preponderance of orderly, repeated, equivalent, reading times in the word unit condition. Participants appear to discover a strategy of space bar pressing to advance the text, that is relatively unreflective of reading itself.

Word reading times are expected to be rate limiting of spacebar presses, as the logic of self-paced reading experiments dictates, but the opposite appears to be true. Apparently the execution of a spacebar press usually takes more time than the time it takes to read a single word, so spacebar presses are the effective rate limiting behavior. This is a concern because word unit presentation is the typical mode of presentation in self-paced reading experiments. This discovery also presents an opportunity because the strategy distinguishes graduate students from undergraduates.

Although $%\text{DET}$ did not differ reliably between groups of readers or repeated readings of the text (all $F(1, 12) < 1.71$), MAXL reveals a statistically reliable 2 x 2 interaction ($F(1, 12) = 10.85, p < 0.01$), as can be seen in Fig. 6(b). Graduate students’ space bar presses became less strongly attracted in multiple readings to the
strategic rhythm of space bar presses to advance the
text. Undergraduate readers change in the oppo-
site direction to become more strongly attracted to
rhythmic pressing to advance the text. These oppo-
site effects of multiple readings for graduate versus
undergraduate readers, are generally comparable to
other findings concerning less skilled versus more
skilled, fluent behavior.

For instance, perhaps the more fluent graduate
student readers relax their strategic voluntary con-
trol of spacebar pressing into a least effort, more
flexible, but less systematic form of control. Having
read a story more than once, they do not require the
same degree of voluntary effort to read the story
again or to advance the text while reading. Stud-
ies comparing more or less skilled performance of
other tasks have discovered a similar outcome using
scaling exponents as the outcome variable ([Van
Orden et al., 2011], is a review). Initially, in skill
acquisition, there is a strong correlation between
skill and scaling exponents. Greater skill equals scal-
ing exponents that converge toward $\alpha = 1$ (e.g.,
[Wijnants et al., 2009]). In the most highly skilled
performance, however, the performance of so-called
automatic behaviors, scaling exponents appear to
find a place of “least effort”, retreating from $\alpha = 1$
to a value in the direction of $\alpha = 0$. Scaling expo-

nents may back off from $\alpha = 1$ to increase flexibility,
for instance, as skill reduces the need for voluntary
control [Kloos & Van Orden, 2010].

This fluency hypothesis explains the perfor-
mance of undergraduates as well. Undergradu-
ates show an effect opposite to graduate students,
becoming more stably attracted to their strategy
of advancing the text in multiple readings. They
have not yet reached the pinnacle of flexible read-
ing fluency that graduate students have achieved. Consequently, undergraduate readers require more
voluntary control of reading and advancing the text in all conditions and actually benefit from multiple readings toward a more orderly, controlled, strategy of spacebar presses to advance the text.

6.2. Phrase units

In contrast to the results from word units, phrase units are a sufficiently unnatural mode of text presentation that all readers have an opportunity to increase their task skills in self-paced reading. Consequently, both graduate and undergraduate readers exhibit more order in the pattern of reading times in multiple readings [i.e. increased %DET in multiple readings of the text, \( F(1, 12) = 15.03, p < 0.01 \), and no reliable interaction effect, as portrayed in Fig. 7(a)], and both graduate and undergraduate readers are more strongly attracted to the orderly pattern of reading times in multiple readings [i.e. increased MAXL in multiple readings of the text, \( F(1, 12) = 4.66, p = 0.052 \), and no reliable interaction effect, as portrayed in Fig. 7(b)]. This speculation agrees as well with the outcome of the analyses of scaling exponents, that phrase unit presentations were the most disruptive of the fractal pattern of coupling between reader and text in self-paced reading. Recall that the spectral analyses only found a fractal pattern of coupling in the reading times of the more fluent graduate student readers, and then only in multiple readings.

6.3. Sentence units

When presented with the more natural text unit of sentences, highly fluent PhD candidates in English Literature can read more flexibly and with less effort than undergraduates, and benefit little or not at all from multiple readings. This outcome for reading times to sentence units again reinforces our speculation that undergraduates are less fluent readers who enhance their skill in the self-paced reading task in multiple readings, which we see as a stronger more stable pattern in the self-paced reading times of sentence units. With sentence unit presentation, PhD candidates do not gain appreciably from multiple readings, while undergraduates produce more orderly and stable sequences of reading times in multiple readings (i.e. visible increases in average %DET and MAXL, see Fig. 8, that are statistically reliable in %DET, \( F(1, 12) = 11.99, p < 0.01 \)). This outcome is also the source of a reliable main effect of multiple readings on %DET, \( F(1, 12) = 11.94, p < 0.01 \), which appears to be localized (marginally) in undergraduate reading times (\( F(1, 12) = 3.36, p = 0.092 \)).

7. Cross Recurrence Quantification Analysis (CRQ)

Cross-recurrence analysis was developed to investigate the shared dynamics of two systems that may be revealed in their respective repeated measurements [Shockley et al., 2002]. In the present study we used CRQ to examine the shared dynamical structure among different readers who read the same text under the same conditions. So, for instance, we can ask whether two undergraduates produce shared patterns of change in repeated measurements because they read the same text, advancing it identically, unit by unit in spacebar presses.

The abduction behind this analysis is that readers may entrain to a text in self-paced reading, and shared dynamics are detected in CRQ due to
the common text. The CRQ analysis of repeatedly measured reading times used %REC, the percentage of reading times located in the phase space (of z-scores) that are nearly identical, %DET, the percentage of recurrent points that fall on the same diagonal, and MAXL, the longest line of recurrent points on the same diagonal. All three quantities may gauge the strength of shared entrainment in self-paced reading. Each condition included four readers and we examine all possible pairings, which yield \( \frac{4!}{2! \times (4 - 2)!} = 6 \) possible pairings in each cell of identical reading conditions.

Every data series was z-transformed before submitting the data to CRQ to control for differences in average spacebar press times on the shared dynamic structure of each time-series. To explore differences in percent recurrence (%REC), the parameters for delay, dimension and radius were fixed to the average values derived from the individual fittings of RQA. To explore differences in %DET and MAXL, the radius parameter was adjusted to equate cross recurrence (again at 2%) across every pair of participants’ paired data series. Cross recurrence must be equated because %DET and MAXL are directly affected by differences in the percentage of recurrent points. As with RQA, the overall magnitudes of the measures %REC \((F(2, 69) = 27.02, p < 0.001)\), %DET \((F(2, 69) = 31.84, p < 0.001)\), and MAXL \((F(2, 69) = 84.26, p < 0.001)\) were greater for word unit conditions compared to phrase unit and sentence unit conditions (see Figs. 9–11).

7.1. Word units
Graduate students produce largely idiosyncratic dispersions of reading times to word units and thus share less recurrence overall, compared to undergraduates [i.e. less %REC of spacebar press times between paired graduate students than between paired undergraduates, \( F(1, 20) = 15.49, p < 0.001, \)]
Reading performance to phrase units of text in (a) average %RECurrence, (b) average %DETerminism and (c) average MAXLine (with standard errors), between reading times of pairs of graduate students versus pairs of undergraduate students, reading the same story either once or multiple times.

see Fig. 9(a)]. Nonetheless, despite less recurrence overall and fewer opportunities in which to discover order, paired graduate students share more orderly and entrained patterns of reading times the first time they read the text, drifting away from each other into strongly attractive, idiosyncratic patterns in multiple readings. Paired undergraduates produce reading times that recur (are equivalent) for fully 35% of word units, on average, in their first reading, but a shared order among these shared values only emerges in multiple readings, in which a strongly attractive pattern of shared performance between paired undergraduates is present (i.e. reliable 2 x 2 interaction effects were found between graduate versus undergraduate readers X single versus multiple readings in both %DET, $F(1, 20) = 10.25, p < 0.01$, and MAXL, $F(1, 20) = 6.02, p < 0.05$).

Either undergraduates are more strongly entrained to the text in multiple readings or they refine their spacebar tapping strategy for advancing the story. The latter possibility agrees with their slightly faster and more tightly dispersed spacebar pressing times [Fig. 2(a)] and the apparent change toward $\alpha = 1$ in undergraduate’s average scaling exponents [Fig. 5(a)]. The statistically reliable reduction in %REC [$F(1, 20) = 8.66, p < 0.01$, see Fig. 9(a)] may even imply that the strategy of spacebar pressing has become more independent from the structure of the text in multiple readings, reflecting more exclusively the shared stable strategic pattern of advancing the text. Thus, either the text entrains more of undergraduate readers’ behavior in the same way in multiple readings (compared to graduate students) or undergraduates hit on more similar strategies for pacing their spacebar presses. The latter hypothesis is in line with our previous results. The strategy of using spacebar presses to advance the text becomes less reflective of reading times in multiple readings. The common strategy
is refined in multiple readings becoming the basis for increases in shared order and stronger attraction to shared patterns in paired undergraduates’ performances.

7.2. Phrase units

We have speculated that phrase units are not optimal units for text presentation in the self-paced reading task. The outcome of the CRQ analyses bears this out because no reliable effects of reading fluency or multiple readings were observed. If meaningful patterns emerge across phrase unit reading times they do not distinguish paired graduate students’ performance from paired undergraduates and they are not affected by single versus multiple readings of the text (i.e. for %REC all $F(1, 20) < 1.10$, for %DET all $F(1, 20) < 1.84$, and for MAXL all $F(1, 20) < 2.58$, see Fig. 10).

7.3. Sentence units

We have suggested from previous results that sentences are optimal units for self-paced reading and that graduate students are clearly more fluent readers than undergraduates. In the sentence unit condition, the quantities of shared patterns of performance between graduate students remain constant across multiple readings. The entrainment to text that they share does not change with multiple readings ($t(10) = 1.14, p = 0.28$ for %REC, $t(10) = 0.06, p = 0.96$ for %DET, $t(10) = 1.27, p = 0.23$ for MAXL). Constant results across single versus multiple readings also agree with the results from individual RQA analyses of sentence unit reading times, which also found little change of graduate student performances in multiple readings. This makes sense if graduate students are at the performance ceiling of reading fluency, which we can see most clearly once the presentation units are right
for revealing fluency in self-paced reading. Sentence units are natural units of story structure and they are rate limiting compared to spacebar pressing.

The present sample of undergraduate readers are less fluent readers than the sample of PhD students in English literature. By less fluent, we mean that the undergraduates have a greater capacity for change in reading performance. In the CRQ analyses they show a shared capacity for change, and multiple readings reveal this extra capacity for shared changes in paired undergraduate performances. The advantage they gain in multiple readings makes sense if we imagine that the paired undergraduates better entrain to the common text, which produces a greater proportion of shared structure and stronger attraction to the shared structure in sentence reading times [i.e. an increase in shared %DET in multiple readings, $t(10) = 4.85, p < 0.001$, see Fig. 11(b), and an increase in shared MAXL in multiple readings, $t(10) = 1.48, p = 0.17$, see Fig. 11(c)].

8. Discussion

A logic of concatenated effects has guided reading research for over 100 years. This guiding logic is one of divide and conquer, seeking to isolate effects that, in turn, identify the basic mental building blocks essential to reading fluency. The present alternative strategy is to unravel and describe complex patterns of interdependence in data, to quantify changes in the interactions between readers and texts. In the present study, we find evidence of complex higher-order relationships in the dynamics of story reading, discovered in the patterns of reading times. Based on these results, we put forward two speculative hypotheses:

First, the three text units used in the self-paced reading task — words, phrases, or sentences — impose different constraints on reading performance and yield patterns of reading times that refer jointly to task demands and readers. Although all three involve reading, they reflect reading differently in the reading time data they produce — so differently that they cannot be equated along a common quantitative dimension. Task demands in these three conditions are sufficiently different that they essentially create three different tasks.

Second, PhD candidates in English Literature are more fluent readers than freshman undergraduate students; the graduate students displayed qualitatively different reading performance across the three text-unit presentation conditions, and between single and multiple readings of the story. We summarize the basis of these conclusions next.

8.1. Task differences

Reading performances of the three types of text units (words, phrases, and sentences) reveal qualitative differences in long-range correlational structure (i.e. fractal dimension), as well as dissociative changes in participants' performance over reading units and multiple readings. Yet the ordinary goal of task contrasts is to find their common core. Presumably the common core would refer to reading itself, distinct from the idiosyncrasies of reading tasks. Alternatively, why not take the complicated patterns that performance reveals closer to face value? In doing so, we may suggest that reading itself changes among the three tasks. No reading process is shared identically across the three tasks, at least none is revealed in the data.

Conditions that advance the text in word units differ greatly from both phrase unit and sentence unit conditions. The fractal pattern across spacebar press times is much stronger for word unit presentations. As we already suggested, the prominent fractal pattern is consistent with the idea that spacebar pressing takes longer, on average, than word reading times. The average spacebar press time of roughly 250 msec to word units is not much longer than average simple reaction times, for instance, when participants are dedicated exclusively to producing a rapid response. On this basis, we interpreted the spacebar press to be rate limiting; the spacebar press primarily shapes performance across word reading trials, not the other way around. The likely effect on spacebar pressing performance by text and word properties is to occasionally perturb the pattern of spacebar pressing such that it departs toward slower reading times. This might happen to rare words or in difficult passages of text, but it cannot have happened too often given the average pressing time.

Moreover, undergraduate readers fall into an increasingly rigid mode of responding, as they entrain more stably to spacebar pressing task demands. Perhaps they see the task as emphasizing a rigid pace to reveal word units, dealing with the consequences as they come, as concerns reading the story. This may explain why undergraduates share more stable dynamic structure across
spacebar press times in multiple readings, where the text is now familiar. Graduate students do not perform an identical task, however; they do not discover so rigid a pace. In contrast, graduate students adapt more flexibly to the word unit conditions and their performance exhibits a less stable, more flexible mode of spacebar pressing, also close in speed to simple reaction times. Altogether then self-paced reading of word units exhibits unique dynamics compared to the other reading conditions and between graduate and undergraduate readers.

Conditions that advance the text in phrase units differ from sentence unit conditions, as well. The prominent fractal pattern of word unit conditions is absent in almost all phrase unit conditions, and only differs from surrogate data in multiple readings by graduate students. Local details of phrase units themselves dominate reading times, such as the number of words that compose each phrase. For instance, although lacking the higher-order structure that nonlinear analyses reveal, the linear correlations between phrase lengths and reading times were statistically reliable (average $r^2 = 0.44$ for undergraduates first reading and $r^2 = 0.37$ in multiple readings; average $r^2 = 0.43$ for graduate students first reading and $r^2 = 0.50$ in multiple readings).

Perhaps the close reading of phrases is necessary because phrases are more ambiguous than sentences — i.e. reading the phrase “Drakh knew it meant time away from his family,” provides a basis to foster expectations about what will follow, but is by no means uniquely predictive of content or length of the phrase that will follow. Also the practical meaning of the phrase depends on what follows thereafter (e.g. “time away from his family could be something joyful, sorrowful, or simply unimportant”). Garden-path sentences are a prominent example of such “feedback-loops” in text comprehension, where ambiguity in one part of a sentence is resolved through a word or phrase that comes late in the sentence (e.g. [Frazier & Rayner, 1982]). Sentence surprise endings might perturb on-going comprehension of the story, when presented in phrase units, requiring a closer reading in self-paced phrase units. For instance, although lacking the higher-order structure that nonlinear analyses reveal, the linear correlations between phrase lengths and reading times were statistically reliable (average $r^2 = 0.44$ for undergraduates first reading and $r^2 = 0.37$ in multiple readings; average $r^2 = 0.43$ for graduate students first reading and $r^2 = 0.50$ in multiple readings).

Perhaps the close reading of phrases is necessary because phrases are more ambiguous than sentences — i.e. reading the phrase “Drakh knew it meant time away from his family,” provides a basis to foster expectations about what will follow, but is by no means uniquely predictive of content or length of the phrase that will follow. Also the practical meaning of the phrase depends on what follows thereafter (e.g. “time away from his family could be something joyful, sorrowful, or simply unimportant”). Garden-path sentences are a prominent example of such “feedback-loops” in text comprehension, where ambiguity in one part of a sentence is resolved through a word or phrase that comes late in the sentence (e.g. [Frazier & Rayner, 1982]). Sentence surprise endings might perturb on-going comprehension of the story, when presented in phrase units, requiring a closer reading in self-paced phrase units.

Dependencies between sentences exist as well, of course, which is the basis for higher order story structure. Also, presently, the fractal pattern of performance is more prominent in sentence unit than phrase unit conditions, so sentences are not given the close reading, to the same degree as phrase units. In the next section, we discuss results from sentence unit conditions that distinguish graduate student and undergraduate readers in fluency — at least that is our hypothesis. Please note, however, we do not claim that sentence unit conditions replicate the standard reading conditions of continuous text, or reading in any other context other than self-paced reading of sentence units. Maybe sentence units will prove better to distinguish adult fluency, but it could also turn out that word units better distinguish fluency in beginning readers.

Our discussion of how the three task conditions impose different constraints on participants, hopefully raises some concern about any claim that a laboratory task is transparent to a cognitive process. We conclude, from these and other results, that reading performance is always constituted in the light of contextual constraints and task demands, and is never transparent to a reading process. Contextual constraints are not merely perturbations of performance that can be minimized, context plays a constitutive role in reading and language comprehension and in the performance of all tasks that include printed text [Van Orden et al., 1999].

8.2. Reading fluency

Graduate students produced a pattern of scaling exponents consistent with greater fluency in the sentence unit condition; although, admittedly, our interpretation relies on what has been learned about other cognitive and motor behaviors, and in particular fluent human gait [Van Orden et al., 2011]. Presented with novel cognitive and motor tasks, participants’ scaling exponents generally approach $\alpha = 1$ after much practice. But in adults’ preferred gaits, the most fluent behavior studied to date, scaling exponents fall within a narrowly constrained range of scaling exponents about $\alpha \approx 0.60$. Departures from adults’ preferred gaits, however, whether walking or running, or speeding up or slowing down, yield scaling exponents that reapproach $\alpha = 1$, a finding that is reliably present across a variety of gait measurements [Hausdorff et al., 1996; Jordan et al., 2007a, 2007b]. Seemingly, fluent gaits find default scaling exponents like those of slightly whitened, or partly random, fractal patterns. Consequently, flexibility is enhanced at that same time that demands for voluntary control in preferred gaits is reduced [Kloos & Van Orden, 2010].
Fluent reading may also find default scaling exponents that are slightly whitened. Yet preferred gait is usually measured on a treadmill, which reduces uncontrolled sources of perturbation, to give a clearer default picture of fluent gait. Here, however, reading time measurements were taken as the story accumulated, in units that changed in length randomly, in a pattern of change statistically equivalent to white noise. Thus the superficial default for entrained reading times is a random signal; which reading times, of both graduate and undergraduate readers, track reliably (e.g., for sentence lengths and reading times, \( r^2 = 0.64 \) and \( r^2 = 0.58 \) for graduate student’s first and multiple readings, respectively, and for undergraduates \( r^2 = 0.58 \) and \( r^2 = 0.54 \)). Nevertheless, graduate students’ and undergraduates’ scaling exponents to sentence-units were reliably different from \( \alpha = 0 \) (see Fig. 5); so sentence reading times do not simply reflect the default surface properties of text.

When reading sentence units, graduate student and undergraduate readers differ after the first reading, as scaling exponents of graduate students move closer to \( \alpha = 0 \) in multiple readings. In contrast, when word or phrase units are presented in multiple readings, scaling exponents of graduate students change in the same direction as those of undergraduates, toward \( \alpha = 1 \). The fact that graduate students’ change toward \( \alpha = 0 \), when reading sentence units, does not stem from a closer reading of sentences with multiple readings; the random pattern of sentence lengths is more strongly correlated with reading times in first readings for instance, as noted. This change may reflect a greater capacity to flexibly entrain to the story structure, yet another source of perturbation to reading times, despite the odd presentation mode of one sentence unit at a time. This conclusion draws heavily on the analogy with fluent gait, of course; fluent reading expresses flexibility in a more whitened scaling exponent.

Flexibility implies more behavioral options to choose from to accomplish the same task, or to adjust reading behavior online, and will thus result in more varied and variable outcomes. The subtle flexibility of expertise has not been captured in traditional studies [Dreyfus, 1992]. In linear analyses, for example, the data features of flexible outcomes will not be seen as positive facets of performance. They will appear superficially to be no change (flexible adjustments to sustain performance), instability (multiple means or ways to produce successful outcomes), or possibly nonstationary data outcomes (multiple choices among strategies). The data outcomes of the graduate student readers, when reading sentence units, illustrate this kind of pattern throughout.

For instance, the primary characteristics of the more fluent graduate students include stable reading times of sentence units, scaling exponents that become more idiosyncratic in multiple readings, and RQA and CRQ variables in sentence reading that show little or no change from single to multiple readings. By comparison, when undergraduates read sentence units they produce the contrasting positive or null changes in the same variables, including faster reading times, more determinism, and equivalent or stronger attraction to the surface structure in multiple readings. We interpret these differences between undergraduates’ and graduate students’ performances to indicate that the undergraduate participants have a greater capacity for improvement toward reading fluency.

Additional evidence is present in contrasts among sentence unit versus word and phrase unit conditions in graduate student performance, where graduate students appear more like undergraduates in word and phrase unit conditions, although not identical. Reading phrase units, for example, graduate students and undergraduates are qualitatively similar, changing in the same direction in multiple readings, making positive or null gains in speed and stability of reading times. Also, in word unit conditions, neither graduate students nor undergraduates change appreciably in reading times (spacebar presses are rate limiting) and both produce scaling exponents that change toward \( \alpha = 1 \) in multiple readings, consistent with gains in skill in their spacebar-tapping text-advance strategy. Yet the RQA and CRQ results suggest that undergraduates simply become more stable and more strongly attracted to a strategic rhythm of tapping, to advance the text, whereas graduate students appear more flexible in using this strategy with multiple readings, and more idiosyncratic in their patterns of responding.

Undergraduate reading performance increases in stability with repeated readings, showing a latent capacity to become more fluent, despite good comprehension. And, as noted, comprehension scores do not reliably distinguish adults’ fluency [Hembrouck & Tindal, 2006]. Thus gains toward expert fluency are gains toward flexible, effortless, immediate accommodation of text structure. Fluent control of behavior by experts is not only flexibly
skilled, but also partly off-loaded to the available constraints of the tools and tasks in question, which for gait would include the steady pace of the treadmill, and which for reading would include the story structure. This emerging picture of fluent behavior is reminiscent of Martin Heidegger’s phenomenology of “ready-to-hand”, to describe tools and tasks that are transparent to one’s intentions of use — the available constraints to the mind and body are transparent at their interface to the required constraints of the world, like hand in glove, so to speak (Dotov et al., 2010).

Thus, reading the story text in sentence-units seemed to situate graduate student readers best to demonstrate characteristics of reading fluency, as “ready-to-hand”, especially in contrasts with word and phrase units. For instance, graduate students read phrase units more quickly in multiple readings, but not so sentence units in multiple readings. Given sentence units, graduate readers are sufficiently fluent in the first reading to perform at ceiling, “ready-to-hand”. Nonetheless, a more abstract text might have made them look like they gain in sentence reading fluency in multiple readings, perhaps, as the phrase-unit conditions do, and this fluid “now you see it, now you do not” aspect of fluency appears to be inevitable. Fluency must be measured and measurements conflate the capacities of readers with the demands of reading tasks (cf. [Holden et al., 2010; Van Orden et al., 2010]).

8.3. What next?
In designing this exploratory study, we naturally made choices about where to start that limited the scope of the present research. One choice was to control for text by using a single story in all conditions and examine readers at different skill levels. An obvious extension of the present study would be to examine the same readers in all conditions, including texts that differ in difficulty. Variations in difficulty of text might be particularly suited to illuminate the relationship between the dynamics of the reading process, effort (or effortlessness) in reading, and actual text comprehension. We have not yet conducted such a study.

Another important extension would be to examine reading of children in the process of acquiring literacy (through cross-sectional and longitudinal investigation) or children who suffer from developmental dyslexia. In that regard, Maarten Wijnants and Anna Bosman, colleagues at Radboud University in Nijmegen, have observed changes in scaling exponents toward $\alpha = 1$ that distinguish dyslexic children from nondyslexic children of the same age in a word naming task. The scaling exponents of dyslexic children were not reliably different from the $\alpha = 0$ of white noise, the scaling exponents of nondyslexic children were reliably different in the direction of $\alpha = 1$, but not so far in that direction as undergraduate readers performing the same task.

We should also cross-validate our results using different behavioral measures of reading. A reading study that measures eye-movements could produce data amenable to nonlinear analyses. While eye-tracking studies of reading have been conducted for many years, the aims of previous studies and their procedures and methods of linear analysis differ decidedly from our approach. Measurement of reading via eye-movements could also extend to studies that do not present discrete text units. We could evaluate self-paced reading using literally self-paced eye movements. These are all goals that we are currently pursuing.

Finally, research in psycholinguistics has identified many linguistic variables and word properties that influence reading performance. These linguistic variables are interesting in the present context as well, since they are cornerstones of conventional ideas about reading. Thus, how they may modulate the dynamics of self-paced text reading will also stipulate further theory development, and we are currently expanding our analyses to include prominent psycholinguistic variables used previously in reading research. Investigating the impact of these variables, as well as their relation to dynamic measures of the reading process will be important to link our findings to the large body of the existing literature.

Of course the focus of this exploratory work has been reading fluency, specifically, not just reading performance. Reading fluency develops over the course of each individual’s lifetime, from elementary school, to high school, and as our data suggest — through university and beyond. Minimally, further details of the dynamics of reading behavior while reading continuous text will further specify the subtleties of what fluent reading entails. With that goal in mind, we have introduced methods of complexity science amenable to study the dynamics of reading behavior. The present modest success, distinguishing reading behavior of elite PhD candidates in English Literature from competent undergraduate readers, may...
promise refined metrics of reading fluency across the lifespan.

Acknowledgments

We thank Anna Hausmann for helpful comments on manuscript versions of this draft and we thank Eleanor Logan for her contribution to data collection. Preparation of this article was supported by NSF Grants to Guy Van Orden (BCS #0642716, BCS #0843133, DHB #0728743).

References


Notes in Mathematics


