Syntactic Complexity of Reading Content Directly Impacts Complexity of Mature Students’ Writing

Yellowlees Douglas¹ & Samantha Miller²

1 Center for Management Communication, Hough Graduate School of Business, Warrington College of Business Administration, University of Florida, Gainesville, USA
2 Heavener School of Business, Warrington College of Business Administration, University of Florida, Gainesville, USA

Correspondence: Yellowlees Douglas, Associate Professor, Center for Management Communication, Hough Graduate School of Business, Warrington College of Business Administration, University of Florida, Gainesville, FL 32611-7150, USA.

Received: April 2, 2016               Accepted: April 26, 2016               Online Published: May 4, 2016
doi:10.5430/ijba.v7n3p71               URL: http://dx.doi.org/10.5430/ijba.v7n3p71

Abstract
Increasingly, schools and colleges of business focus on the quality of their students’ writing, reflecting complaints from business and industry about the quality of writing of entry-level employees. These concerns about student writing have led to some changes in the curricula and admittance of students into graduate programs, including analytical writing essays on the Graduate Management Admission Test (GMAT). However, research also suggests that reading content and frequency may exert more significant impacts on students’ writing than writing instruction and frequency. This study surveyed a cohort of MBA students on their regular reading content and sampled their writing. We then used algorithm-based software to assess the syntactic complexity of both reading content and writing samples. Our findings reveal strong correlations between students’ most common reading content and their writing on widely-used measures of writing sophistication: mean sentence length and mean clause length. Several mechanisms may account for the dramatic influence exerted by reading content on mature students’ writing—including synchrony, priming, and implicit learning. But, irrespective of these mechanisms, undergraduate and graduate programs in business should emphasize ongoing reading of syntactically complex content both during and after students’ schooling to address the sophistication of their writing.

Keywords: writing, reading, syntactic complexity, MBA education

1. Introduction

1.1 The Impacts of Reading on Writing Skills

For decades, researchers have understood that reading and writing skills more or less track closely, particularly during students’ elementary and secondary education. However, most studies have focused on the acquisition of skills necessary for students to grapple with grade-appropriate reading assignments and to produce similarly grade- or age-appropriate writing (Nelson, Perfetti, Liben, & Liben, 2012). Developmental delays impact reading and writing skills more or less equally (Schoneil, 1942). In addition, researchers have tracked growing complexity in reading with the same markers of complexity in writing skill (Johnson, 1980; Loban, 1963). Even in studying the acquisition of second and third languages, reserachers have established that fluency and syntactic complexity in writing in students’ native languages closely influence the same factors in their non-native languages (Lu & Ai, 2015).

Using a similar focus, psycholinguists have also discovered that our speech reflects the language we encounter most frequently, with even fleeting grammatical constructions in spoken dialogues eliciting similar grammatical constructions from others (Niederhoffer & Pennebaker, 2002). This influence may stem from synchrony present in interpersonal communication that may dictate even the respiratory rate of conversation partners (Pennebaker, 2011). However, this influence may also reflect the effects of priming, even when either the questions or target text use seldom-encountered grammatical constructions, like double objects (J. K. Bock, 1986; K. Bock, Dell, Chang, & Onishi, 2007; K. Bock & Griffin, 2000). While the laboratory settings for most studies may limit the applicability of
their findings to our understanding of long-term effects, at least one study has documented the effects of synchrony over long periods through correspondence between writers (Ireland & Pennebaker, 2010). Nevertheless, even these findings reflect increases and sharp declines in synchrony based on the writers’ emotional attitudes toward one another, rather than the influence of even the more seasoned or renowned writer on the junior (Ireland & Pennebaker, 2010).

Still, after nearly three-quarters of a century of research on the interaction between reading content and writers’ output, we still lack a clear sense of the extent of this influence. Some studies demonstrate that reading more clearly influences the quality of students’ writing than the inverse (Shanahan & Lomax, 1986). More tellingly, the amount of time away from school that students spent reading clearly influenced their facility at writing (Monk, 1958). While the literature is mostly scarce and somewhat mixed, students’ writing improved more markedly when their classroom instruction included more reading (Siedow, 1972) than when students focused more on producing regular writing (Stotsky, 1983).

Yet several key questions remain unanswered: (1) Can syntactic sophistication of our reading influence our writing? (2) Do these effects persist over long periods of time? (3) Are adults as susceptible to these effects as children perhaps mimicking the language they struggle to master? This study hypothesized that clear correlations exist between the regular sources read by adults and adults’ writing, evident in two markers of syntactic complexity: mean length of sentences and mean length of clauses.

1.2 Feedback from Business and Industry: Improve Writing

Changes in workplace technologies have placed an even heavier emphasis on reading and writing skills than they had in the twentieth century workplace. Employees now send and receive more messages than ever before, while applications like email have eliminated editors and support staff who would formerly have edited writing for managers. As a result, corporations increasingly complain about the skills of established and entry-level employees alike. For example, the National Association of Colleges and Employers Job Outlook reported that organizations sought employees who wrote well but were increasingly dissatisfied with employees’ writing skills. In fact, when organization ranked the skills they sought and the skills their employees possessed, writing skills had the largest gap between employer expectations and employee delivery (Employers, 2011). That same year, a study of accounting firms found that employers expected but were only “marginally satisfied with” eight different kinds of writing skills required of their employees, from writing clearly and precisely to the effective use of email (Middleton, 2011).

In response, schools and colleges of business have increased their focus on writing, requiring more writing of undergraduate and graduate students alike—without noticeably impacting the quality of graduates’ writing. At the same time, the Graduate Management Admission Test (GMAT), the most widely used criteria for admission to US graduate programs in business, has placed greater emphasis on student writing via an analytical writing assessment, scored by both human readers and algorithm-driven software (Council, 2016). Nevertheless, poor writing skills remain the top pet peeve among corporate recruiters of graduates of MBA programs (Korkki, 2007). But the problem may not lie in the frequency of writing assignments in courses or in the lack of investment by adjunct faculty in addressing student writing issues or even in the lack of workplace focus in writing instruction in courses (Washington, 2014). Instead, the problem may lie in two entirely different areas: the role of domain-based knowledge in determining the sophistication of graduates’ writing (McCutchen, 1986) and in the content of their writing before, during, and after their education (Stotsky, 1983).

1.3 The Relationship between Reading and Writing

Surprisingly, graduate programs in business have been slow to recognize the interaction between the syntactic sophistication of reading content and students’ writing. While most programs expect students to remain au fait with developments in business and industry by reading the likes of The Economist and The Wall Street Journal, few courses require students to read business-related periodicals. Moreover, even when faculty commonly reference academic journals in their daily instruction, faculty generally assign course textbooks, many of which are poorly written and receive minimal editing due to publishers’ cutting costs by requiring writers to submit camera-ready texts for rapid publication. However, a wealth of studies document the impact of reading content on writing, reflected in the measurable syntactic complexity of subjects’ writing. Some researchers have dubbed this near-mimicry “structural priming,” where passively constructed questions elicit passively constructed answers, and rare inversions of word orders in questions trigger the same inversions in responses (Martin J Pickering & Ferreira, 2008). These priming effects also influence a reader’s description of a picture, with subjects’ descriptions of the pictures mimicking the syntactic features of a text they had read earlier, down to using the same unusual constructions like double objects, featured in the priming text (J. K. Bock, 1986). Moreover, priming effects can prove startlingly
durable. Researchers have found significant syntactic repetitions between priming effects, even when subjects provided verbal responses to written texts or written responses to verbal texts (K. Bock & Griffin, 2000; Martin J Pickering, Branigan, & McLean, 2002). In addition, priming effects exert such strong memorability for syntactic structures that subjects’ writing samples reflect structural priming across as many as ten intervening prompts in laboratory settings (K. Bock et al., 2007). Unfortunately, to date, the strongest data comes directly from laboratory settings and not from studies of long-term reading and samples of writing taken from outside the confines of a one- or two-hour laboratory study. Does structural and syntactic priming or even the synchrony observed by psycholinguists in interpersonal communication (Ireland & Pennebaker, 2010; Niederhoffer & Pennebaker, 2002; Pennebaker, 2011) apply to real-world reading content and writing produced in organizational settings?

1.4 Hypothesis

This study aimed to examine the correlation in syntactic complexity between the graduate business students’ writing and their most frequently read sources. We hypothesized that MBA student writing would reflect the syntactic complexity of the sources they read on a regular basis and also sought to explore whether, in adults, the average amount of non-school reading time would also correlate with the complexity of their writing.

2. Methods

This study relied on two different instruments: an online survey of reading habits and sources and algorithm-based software that assesses syntactic complexity of sample texts from both student writing and reading sources. The survey, conducted independent of any course or classroom setting, solicited a writing sample as well as participants’ self-reported reading materials, drawn from sources students in an MBA cohort regularly read. The survey, sent out via an emailed link, asked students to indicate which sources they regularly read from among business-related periodicals, newspapers, and online news outlets. In addition, the survey also asked participants to identify the type of sources they most regularly read: online-only content, newspapers, genre fiction, news-related weekly periodicals, general interest non-fiction, literary or critically acclaimed fiction, and academic journals (see Table 1). Participants also indicated how frequently they had read fiction in the past year and the average amount of hours they spent reading weekly. Since all students who received the survey had already completed a mandatory writing course for all MBA students, GEB 5212 Professional Writing in Business, the survey requested that participants merely copy and paste the second paragraph from a cover letter for a job application, an assignment students completed in the course and which many students use to secure their first post-MBA jobs.

Table 1. Survey options for reading materials

<table>
<thead>
<tr>
<th>Business-focused Magazines</th>
<th>General Interest Newspapers</th>
<th>Online Sources</th>
<th>Most Frequently Read Content</th>
</tr>
</thead>
<tbody>
<tr>
<td>The Economist</td>
<td>The Wall Street Journal</td>
<td>BuzzFeed</td>
<td>Web-based content not published in print</td>
</tr>
<tr>
<td>Forbes</td>
<td>New York Times</td>
<td>Tumblr</td>
<td>Daily newspapers</td>
</tr>
<tr>
<td>Business Week</td>
<td>USA Today</td>
<td>Reddit</td>
<td>Weekly magazines</td>
</tr>
<tr>
<td>Money</td>
<td>Gainesville Sun</td>
<td>Huffington Post</td>
<td>Academic journals</td>
</tr>
<tr>
<td>Fast Company</td>
<td>(Don’t read any newspaper regularly)</td>
<td>Medium</td>
<td>General interest non-fiction</td>
</tr>
<tr>
<td></td>
<td>Ozy</td>
<td></td>
<td>Fantasy, Science-fiction or Mystery fiction</td>
</tr>
<tr>
<td></td>
<td>(Don’t read any of these sources regularly)</td>
<td>Literary Fiction</td>
<td>(Don’t read any of these sources regularly)</td>
</tr>
</tbody>
</table>
2.1 Software for Lexical and Syntactic Analysis

Several platforms measure syntactic complexity, including Coh-Metrix (Scott A Crossley & McNamara, 2011; Graesser, McNamara, Louwerse, & Cai, 2004). Coh-Metrix measures five factors that the literature on reading has established as central to a text’s readability (Just & Carpenter, 1987; McNamara, Crossley, & McCarthy, 2010; Zwaan & Radvansky, 1998). These five factors include three directly relevant to the analysis of business writing: word concreteness and both referential and inferential cohesion (Graesser et al., 2014; Graesser et al., 2004). While Coh-Metrix is well-validated, this tool for assessing syntactic sophistication has several potential confounders for this particular study. First, Coh-Metrix has greater accuracy in assessing narrative over the expository or argumentative texts central to writing in business (Scott A Crossley & McNamara, 2011; Hiebert & Mesmer, 2013). Second, Coh-Metrix also focuses on word concreteness, presenting a potential confounder to our study’s analysis of our participants’ reading sources, as our university’s MBA curriculum requires students to complete a writing course, GEB 5212, that relies heavily on psycholinguistic and neuroscience research to teach writing. Early in the course, students learn to prefer concrete nouns and action verbs, as well as active construction, over abstract nouns, non-action verbs, and passive construction. These textual features, our participants had already learned, bolster the concreteness and memorability of texts. As a result, Coh-Metrix could potentially produce higher scores from our participants’ writing samples over their reading samples, as participants received direct instruction, explicit feedback, and grades based on their use of concrete details in their writing. Third, Coh-Metrix scores writing based on the repetition of words between sentences—known as referential continuity—a feature that increases reading speeds and eases readers’ comprehension of texts (Garnham, Oakhill, & Johnson-Laird, 1982). Similar to concreteness, referential continuity represents an additional confounder, as our students also receive explicit instruction on linking sentences via three different devices—common grammatical subject, sequencing, and transitions—that promote referential continuity in their writing (Douglas, 2015). And, finally, Coh-Metrix also scores writing samples for inferential cohesion (Kintsch, 1988; Kintsch & Van Dijk, 1978; Zwaan, 1994), an area in which our participants also received direct instruction through principles for increasing textual coherence and lowering demands on readers’ inferential processing (Douglas, 2015). As a result, we excluded Coh-Metrix from our methods for analyzing both reading and writing samples to avoid skewing our outcomes.

We elected, instead, to use an algorithm-based tool for analysis that focuses exclusively on syntactic complexity to avoid potential confounders that could skew the analysis of our participants’ texts over their reading sources. In the past five years in graduate business programs, non-native speakers of English students increasingly outnumbered native speakers (Kozel, 2010). As English as a Second Language (ESL) speakers have begun to dominate graduate education, tools to assess their mastery of their second language (L2) have increased in accuracy, focusing exclusively on structural features of language and enabling researchers to assess whether fluency in native languages translates to fluency in L2 (Lu, 2010, 2011). This exclusive focus on syntactic features, present in Ai’s L2 Syntactic Complexity Analyzer, lends itself to quantitative assessment and also avoids confounders inherent in Coh-Metrix. As a result, we relied on Ai’s L2 Syntactic Complexity Analyzer to evaluate participants’ writing samples and samples from selected periodicals and from books that fell within the genre categories our students read most frequently (see section 2.3, below for examples of texts). This web-based analyzer counts the occurrences of fourteen key indicators of syntactic complexity (Lu & Ai, 2015), as indicated in Table 2. These key indicators represent widely agreed-upon measures of sentences’ structural complexity (Goldman & Lee, 2014; Gordon, Hendrick, & Johnson, 2004; Hiebert, 2011). In addition, Ai’s L2 analyzer also incorporates one measure, mean length of sentences, that figures centrally in the most widely-used platform for analyzing and labeling the readability of texts, the Lexile® framework (Lennon & Burdick, 2004). Currently, the Lexile® framework evaluates both grade-appropriateness of texts in elementary and second education throughout the US, in addition to providing reading-level scores for over 100,000,000 books, articles, and websites world-wide (MetaMetrics, 2015).
2.2 Participants
A total sample size of 65 MBA students received an email containing a link to our survey, which elicited a 74% response rate (n=45). Participants in the MBA cohort included non-native speakers of English, as well as a range of 23-42 years of age.

2.3 Selection of Reading Samples
Reading materials, even within a single publication, can vary significantly in syntactic complexity, even within syndicated newspapers, which, in the US, aim to make all articles comprehensible to adults with a fifth-grade education. Nevertheless, even in breaking news coverage, articles containing international news are more challenging to readers than articles covering domestic news (Raze, 1969). Moreover, other types of articles—lifestyle features and science-focused coverage—differ significantly in readability from news articles (Abdelmutti & Hoffman-Goetz, 2009). Thus, the potential for even intra-publication variance is significant.

To avoid variances within publications stemming from types of text samples, we selected a single, unfolding news story that received coverage across our survey’s target publications, the unfolding contest for the Republican presidential nomination as the race began its preliminary stages in late 2015. We thus minimized intra- and inter-publication variance by selecting reading samples featuring common themes, contestant names, and content across newspapers, magazines, and online journalism. In addition, we selected general non-fiction, genre fiction and literary fiction samples based on a business-relevant focus (e.g., The Big Short) or featured in other course syllabi (e.g., Jane Austen, Game Theorist; Thinking, Fast and Slow), best-seller status (e.g., Unbroken), recent critical acclaim in the press (e.g., City on Fire), or works that had become sources of films (e.g., World War Z) or television series (e.g., Songs of Fire and Ice). For academic journals, we selected business-relevant publications, including American Economic Review, Psychological Science, Journal of Sociology, and Administrative Sciences Quarterly.

2.4 Selection of Average Reading Types
In identifying their most frequently-read sources, participants indicated reading web-based only sources, genre fiction, business and general non-fiction, literary fiction, and academic journals. These indications eliminated weekly business-focused periodicals and all newspapers from our correlations between reading sources and writing samples. As a result, we averaged scores on L2 measures of syntactic complexity from six representative samples for books and academic journals in each category: genre fiction, general interest non-fiction, fiction (best-sellers, classic, or critically acclaimed) and academic journals. For all web-based sources, we relied on L2 scores for the sentinel story.
we chose to avoid variance in our sampling, the unfolding race for the Republican presidential nomination in the last month of 2015.

2.5 Data Analysis

Once we obtained scores from the L2 Analyzer for fourteen measures of syntactic complexity across all reading and writing samples, we used regression models and Spearman’s nonparametric correlations to arrive at three measures which correlated significantly (p<0.05) across reading and writing samples: Mean Length Sentence (MLS), Mean Length Clause (MLC), and Complex Nominals per T-unit (CN/T). These measures correspond to features most commonly associated with syntactic sophistication (Evans, 1979; Goldman & Lee, 2014; Gordon et al., 2004). We then analyzed data using regression analysis and used both Spearman’s non-parametric and Pearson’s correlations. Specifically, we focused on correlations between

1. MLS, MLC, and CN/T in reading and writing samples;
2. MLS, MLC and CN/T in average reading type and participants’ writing samples;
3. MLS, MLC and CN/T and self-reporting reading of fiction regularly;
4. MLS, MLC and CN/T and number of hours of weekly reading.

3. Results

The age range of participants provides an excellent snapshot of the reading habits of both Gen X and Millennials but showed surprisingly few differences between the reading preferences of the two groups, particularly in their equal preferences for staples of business-focused periodicals—The Economist, Business Weekly, Forbes, The Wall Street Journal—and online-only, often amateur-provided content from web sources that included Tumblr, Reddit, BuzzFeed, and the Huffington Post. This trend counters some researchers’ arguments that Millennials’ habits clearly distinguish them from Gen Xers (Twenge, 2014). Note, however, that our study made no distinction between digital and print editions of published-in-print periodicals, as the differences between print and digital stories had no impact on syntactic complexity. Specifically, digital-only content exists for some periodicals only if users pay for premium subscriptions, and our participants most likely accessed the business-oriented dailies and weeklies via either personal subscriptions (print or digital) or through the university’s library subscriptions, virtually all read online.

3.1 Average Reported Reading Type Correlates Highly between Reading and Writing on Syntactic Complexity

Our study found significant correlations between average reading type—and our source examples of texts—and writing samples on two primary measures: Mean Length Sentence (MLS) and Mean Length Clause (MLC). Both measures showed significant correlations between reading sources and writing samples: (MLS, p=0.028; MLC, p=0.030). Of the other L2 measures, Complex Nominals per T-unit (CN/T), indicating the number of phrases or clauses begun by head nouns per clause or sentence, narrowly failed to achieve statistical significance between average reading sources and writing samples (p=0.067).

3.2 Average Reported Reading Times Statistically Insignificant

In contrast, two other measures we evaluated—reading fiction at least six times per year and the reported number of hours spent reading—failed to correlate between reading sources and writing samples on any of the L2 factors.

4. Discussion

4.1 Complexity of Commonly-read Sources Significantly Impacts Adults’ Writing

This study represents the first to correlate content read by adults in an MBA program with the quality of their writing, based on syntactic measures of writing complexity. Since psycholinguists first began measuring students’ reading and writing progress by analyzing sentence structure, researchers have relied on several reliable syntactic features to measure the evolving sophistication of students’ writing—primarily mean lengths of sentences (S.A. Crossley, Dufy, McCarthy, & McNamara, 2007; Scott A Crossley & McNamara, 2011; Cunningham & Mesmer, 2014; Schwarm & Ostendorf, 2005), as well as mean lengths of clauses (Lu, 2010; Stenner, 1996; Stotsky, 1983) and the use of complex nominals (Izquierdo & Bailey, 1998).

4.2 Why Complex Nominals Failed to Predict Complexity

Correlations between reading sources and writing samples likely failed to achieve statistical significance due to writing instruction that directed students to avoid using complex nominal that lower clarity in writing. In fact, complex nominal can prove so challenging to comprehension that even a third of native speakers commonly misidentify the head noun in nominals containing as few as three words (Limaye & Pompian, 1991; Price, 1974).
4.3 Frequency of Reading Matters Only over Extended Periods

As expected, we found no correlation between syntactic complexity of writing and the number of hours participants reported spending reading on a weekly basis. We primarily expected this outcome as a natural extension of our participants’ enrollment in an MBA program, some of whom were completing a two-year program in a single year, leaving most participants with little leisure time for reading. However, this outcome also suggests a potentially important implication. The amount of time spent reading for adult writers may prove less influential than patterns of reading over longer periods of time, as adults have left behind the developmental stages where frequency of practice in reading can exert strong influences on writing, via either the acquisition of syntactic knowledge or of domain-based knowledge. Perhaps still more important, the duration of time spent reading mattered far less than the syntactic complexity of the majority of their reading.

4.4 Structural Priming, Synchrony, and Syntactic Complexity

The syntax of reading content impacts the syntax of adults’ writing, likely due to the same mechanisms of structural priming that informs the transfer of syntax from speech to speech (Hartsuiker, Bernolet, Schoonbaert, Speybroeck, & Vanderelst, 2008) and from speech into writing (K. Bock et al., 2007). Structural priming may impact the way we produce writing, just as structural priming impacts our production of speech (Chang, Dell, Bock, & Griffin, 2000; M. J. Pickering, Branigan, Cleland, & Stewart, 2000), due to either our comprehension of language influencing our subsequent production or to structural priming as a form of implicit learning, so durable that structural priming effects lasted over a week in aphasic patients (Saffran & Martin, 1997).

4.5 Implications for Business Curricula

Students who read academic journals, “literary” fiction, or general non-fiction wrote with greater syntactic sophistication than students who read genre fiction (mysteries, fantasy, or science fiction) or exclusively web-based content aggregators like Reddit, Tumblr, and BuzzFeed. In fact, when we examined the scores on L2 measures of syntactic complexity, writing with the lowest syntactic complexity was associated with heavy or exclusive reading of web-based content from the likes of BuzzFeed, Tumblr, and Reddit. In contrast, students with the highest scores of syntactic complexity in their writing read academic journals more frequently than their peers. In addition, these findings may reflect structural priming and synchrony working together to reflect the cadence of writing. Cadence, which readers perceive even during silent reading, involves our perceptions of syntax, likely due to reading activating our brain’s centers for speech, vision, and hearing, all of which work together to enable us to speak, read, and write. In particular, our ability to read and write involves Broca’s area, which enables us to perceive rhythm and syntax; Wernicke’s area, which impacts our perception of words and meaning; and the angular gyrus, central to our perception and use of language. In addition, Broca’s and Wernicke’s areas are wired together by a band of fibres, the arcuate fasciculus, while the angular gyrus itself sits as the junction between the occipital (visual functions) and temporal (auditory functions) intersect (Douglas, 2015). This neuroanatomy may predispose even adults to mimicry and synchrony with the language they routinely encounter in their reading, directly impacting their writing. Based on our findings, graduate business curricula could benefit significantly from incorporating more content drawn from academic journals, rather than textbook paraphrases of significant studies. In addition, courses in organizational behavior, economics, marketing and finance could enhance MBAs’ writing skills by requiring students to apply domain-specific knowledge from business-related news in the likes of The Economist and Wall Street Journal in case-based assignments and tests.

References


Monk, R. H. J. (1958). *A Study to Determine the Relationship between Children's Home Environments and their School Achievemeni in Written English*. (PhD.), University of Washington, Ann Arbor, MI: University Microfilms.


