Linguistic cohesion in middle-school texts:
A comparison of logical connectives usage in science and social studies textbooks

Diego Román
Southern Methodist University

Allison Briceño
San José State University

Hannah Rohde
University of Edinburgh

Stephanie Hironaka
Mills College

Linguistic cohesion in middle-school texts:
A comparison of logical connectives usage in science and social studies textbooks

Abstract

Learning from textbooks is challenging because students must understand novel concepts while also comprehending the language used to convey those concepts. In the domain of science, one posited reason for the perceived difficulty in the reading comprehension of science texts is the low frequency of logical connectives (words that signal relationships between sentences and ideas). To test this claim and discuss its potential effects on the reading comprehension of texts used at the middle school level, this study measured whether the usage of logical connectives (e.g., therefore, so) differed between science and social studies textbooks. Our findings from a large corpus of 12 science and 12 social studies textbooks showed that science texts contained a higher rate of logical connectives than social studies texts. This main effect of subject area also interacted with grade level: The rate of logical connectives usage increased over grade levels in science but not in social studies. Our results showed further differences in the types of logical connectives used across subject areas, with science texts favoring inferential connectives (e.g., furthermore) and social studies texts favoring contrastive connectives (e.g., however). The implications of these findings for the development of science-specific literacy practices are discussed here.

Introduction

Much of the knowledge transfer that occurs in academic contexts relies on students' ability to understand textbooks, so literacy serves as both a goal and a tool for the acquisition of science knowledge. However, traditional strategy instruction is insufficient for teaching students to comprehend science texts (Fisher, Grant & Frey, 2009), and since reading comprehension is typically the domain of language arts classes, science teachers are typically not prepared to support students' reading of texts (McTigue & Slough, 2010).

The challenge in extracting information from a textbook is rooted in the difficulty of simultaneously understanding novel concepts and understanding the language used to convey the concepts (Sanders, Land, & Mulder, 2007). The language of science texts is sometimes neglected by literacy researchers, who instead may focus entirely on vocabulary or prior knowledge (e.g., Fisher, Grant & Frey, 2009). In fact, successful interpretation of science texts requires the inference of different linguistic dependencies—those that hold within sentences and also across sentence—which are often signaled by logical connectives (LCs), or signal words. Since comprehension entails making inferences, LCs can increase comprehension (Best et al., 2005), particularly those that imply causation (Smoklin et al., 2009), such as *because, as a result*, and *in consequence* (Meyer, 2003).

However, since readability scores are often calculated partly based on sentence length, "extra" words such as LCs might be removed by publishers to reduce the calculated grade level of a text. Ironically, words that help students to comprehend science texts are removed to meet the requirements of computer-generated readability scores (McTigue & Slough, 2010).

Logical Connectives in Science Texts

Consider the following excerpt from a science textbook:

1 Asteroids are left over from the formation of the solar system about 4.6 billion years ago. It is thought that they crashed into the inner planets during the early period of our solar system. Asteroids lack enough gravity to have an atmosphere. Consequently, their surfaces have many craters from impacts with other objects (*Focus on Physical Science*, Glencoe, 2007, p. 489).

In order for a reader to assign meaning to the individual sentences, there are basic structural and semantic relationships that must be inferred (e.g., "it is thought that they crashed during the early part of our solar system" identifies when the crashes occurred not when the thoughts occurred). For readers to assign meaning to the text as a whole, they must establish that the sentences relate to each other in meaningful ways. In the context of textbooks, the ability to infer these intersentential relationships allows a student to participate in text-based learning. Yet, what is particularly challenging about understanding a text is that these intersentential relationships can be signaled in a variety of ways and quite often are not signaled at all (Taboada & Mann, 2006).

Interpreting (1) as a coherent passage requires that the last sentence be understood as a description of the result of asteroids' limited gravity. This relationship is signaled overtly in (1) with the logical connective *consequently*, though it could also have been left implicit. The meaning of this passage –that limited gravity results in a lack of atmosphere that renders asteroids vulnerable to frequent impacts that in turn results in numerous craters on the surface of asteroids– depends crucially on the meaning associated with the logical connective *consequently*, as evidenced by the alternative interpretation that arises when the connective is changed: "Asteroids lack enough gravity... *because* their surfaces have many craters." In that case, it would be the cratered surface that is identified as the cause of the limited gravity. A logical connective not only signals how two clauses are intended to relate, but it can also generate inferences of its own. In the case of (1), the consequence described in the last sentence can only be understood if an additional piece of information is added—that the lack of an atmosphere makes asteroids vulnerable to repeated impact from other objects. This information is not stated directly in the text but must be inferred in order for the passage to make sense. An example like this highlights the chain of pragmatic reasoning that readers need in order to process coherent discourse.

A large body of literature in linguistics has addressed the factors that contribute to the meaning, production, and processing of coherent discourse (e.g., Asher & Lascarides, 2003; Taboada & Mann, 2006). In that literature, logical connectives are frequently cast as surface realizations of some of the deeper factors that contribute to the establishment of a coherent discourse. The two relations discussed in (1) represent just two of a larger inventory of possible discourse relations that help structure a coherent text. There remains, however, a serious lack of quantitative description of the range and frequency of linguistic structures that appear in science courses at different grade levels.

In this paper we argue that analyzing logical connectives usage offers an important step in understanding the variation and challenges inherent in the language that students encounter in textbooks. To this end, we focused on the use of overt LCs in middle school textbooks. Examining the role of LCs in science texts problematizes the question of whether science teachers are content teachers or teachers of content, language and literacy. Building on the growing interest in developing subject-specific literacy strategies, this study compared discourse-connective usage across two subject areas –science and social studies– asking how often sentences contained logical connectives and what variation existed among these logical connectives. Although science textbooks were the focus of our analyses, we decided to use social studies textbooks, a non-STEM area as a comparison group. The main aims of our study were to investigate claims about the uniqueness of the language of science, to provide a snapshot of language registers (i.e., clusters of linguistic features) used in two subject areas, and to discuss how having accurate descriptions of these linguistic features can impact the reading comprehension of science texts.

Logical Connectives and Textual Cohesion

Halliday and Hasan (1976) state that cohesion is the "semantic relation between an element in the text and some other element that is crucial to the interpretation of it" (p. 8). In other words, textual cohesion is achieved when the interpretation of one element is properly connected to its linked counterpart. Logical connectives are one of the factors that contribute to cohesion in texts.

Logical connectives are commonly defined as words that can guide the reader in the interpretation of text by signaling semantic relations between discourse segments (Fraser, 2006). Although researchers have used different terms to describe these expressions—logical connectives (Gardner, 1975; Osborne, 2002), discourse markers (Fraser, 2006), discourse connectives (Blakemore, 2002), and conjunctive elements (Halliday & Hasan, 1976)—we adopted the term 'logical connectives' in this paper because it has been previously used in descriptions of science texts used in schools (see Osborne, 2002).

Fraser (2006) points out that every logical connective signals one of four types of relationships: elaborative, contrastive, inferential, or temporal –categories exemplified by the expressions *in addition*, *but*, *so*, and *then*, respectively. We used Fraser's inventory of possible relations and their associated logical connectives because it provided us with a concise and explicit set of categories and connectives.

Previous research has highlighted the importance of understanding how logical connectives create cohesive texts. This type of research has shown that the use of cohesive devices correlates positively with more coherent and understandable text (Degand & Sanders, 2002) as well as improved comprehension (Ozuru, Dempsey & McNamara, 2009). Most cohesive texts seem to benefit struggling readers and English learners (Sanders, Land, & Mulder, 2007) as well as highly-skilled readers (Ozuru, Dempsey & McNamara, 2009).

In the register of science, Halliday and Martin (1993) have posited, however, that science texts have a lower frequency of logical connectives, a feature they named semantic discontinuity. Halliday and Martin described cases of semantic discontinuity as those in which writers “make semantic leaps, across which the reader is expected to follow them [the writers] in order to reach a required conclusion” (1993, p. 82). The following example from Halliday and Martin shows in brackets the logical connectives that the reader would have to insert to comprehend the meaning of the text:

2 The factories have become cleaner, [so] the countryside has become cleaner, and [so] there are getting to be more of the light colored pepper moths (p. 83).

Yet, Halliday and Martin’s work regarding the prevalence of semantic discontinuity in the register of science seems to be qualitative in nature and not based on textbooks used at the K-12 levels. Therefore it is difficult to know how common semantic discontinuity is in science textbooks and how science texts compare to other subjects areas.

The Register of Science

The new science standards in the United States call attention to the language demands of science texts (Hakuta, Santos & Fang, 2013). Although these language demands offer opportunities for science teachers to support science learning and language development for all students (Bunch, 2013), the linguistic features of academic language are rarely discussed explicitly in schools (Fang & Schleppegrell, 2010), and the complexity of science textbooks is not well understood (Frantz, Starr & Bailey, 2015). To that end, Frantz, Starr and Bailey (2015) call for a text’s syntactic difficulty to be explicitly included in measures of text complexity, as grammar contributes to text meaning and impacts comprehension.

Learning in an academic context requires that students read and write using specialized academic styles (Snow, 2008). Various authors have argued that one feature that distinguishes academic styles of language is specialized vocabulary (Dockrell, Braisby & Best, 2007; Snow, 2008). Some of this vocabulary complexity can be attributed to the ambiguity between genre-specific uses of terms and their everyday use (e.g., everyday notions of ‘work’ versus ‘work’ used in physics). Others have proposed that the difficulty of the language employed in science texts lies not only in technical vocabulary, but also in the way the sentences themselves are constructed (Bailey et al., 2007; Frantz, Starr & Bailey, 2015; Halliday & Martin, 1993). According to this view, academic language has a set of specialized language registers, or clusters of linguistic features, and each subject has its own academic register.

Examination of the academic registers in the content areas has primarily been conducted within the framework of Halliday’s Systemic Functional Linguistics (Bunch, 2013; Turkan et al., 2014; Román & Busch, 2015). Halliday and Martin’s (1993) book, *Writing Science*, posits that the science register is characterized by an array of science-specific features: such as nominalizations that introduce abstraction (e.g., *pollution, overconsumption*) and the use of passives can contribute to ambiguity by removing explicit agents from the text. Lastly, as we mentioned earlier, the register of science has been associated with a lack of transitional or cohesive elements, i.e., logical connectives (McNamara, 2001; Osborne, 2002).

It is this last point, regarding the frequency of logical connectives that this study aimed to measure in the context of a large textbook corpus. By using a large corpus, the goal was to undertake a systematic analysis of the particular words that are used as logical connectives in order to compare the language of science with the language used in a non-science subject area. The corpus approach taken here is in keeping with studies by Biber and colleagues (2004) and Conrad (1996) regarding differences across registers and subject areas.

The Importance of Language in Comprehension of Science Texts

In the United States, recent policy initiatives such as the Next Generation Science Standards (NGSS Lead States, 2013) and the Common Core State Standards for English Language Arts and Literacy in Science (CCSS Initiative, 2010) have reignited the interest of educators and researchers in developing science-specific literacy strategies. These policy initiatives have acknowledged the recommendation of various researchers that addressing the challenges of the register of science should play a prominent role in science instruction (Norris & Phillips, 2003; Shanahan & Shanahan, 2008; Slough et al., 2010; Wellington & Osborne, 2001).

Yet, some science educators still consider that addressing literacy development in the teaching science is a "radical proposal" (Alberts, 2010, p. 405) due to the widespread [mis]conceptualization of science as primarily a hands-on subject (Wellington & Osborne, 2001). In fact, few science teachers comprehend "the vital role literacy plays in enhancing rather than replacing science learning" (Pearson, Moje & Greenleaf, 2010, p. 462), and even literacy-friendly science teachers rely on literacy strategies that do not necessarily promote the learning of science concepts (Bunch, 2013; Moje, Stockdill, Kim & Kim, 2010). This focus on general cognitive strategies has raised the question of what emphasis should be placed on the linguistic characteristics of science texts (Moje & Speyer, 2008) and as to whether the difficulties of comprehending subject-specific texts are merely conceptual or whether the complexity of the writing can be attributed to linguistic properties (Halliday & Martin, 1993). Furthermore, most studies of textbooks have focused on cognitive strategies (e.g. predicting, summarizing, inferring) rather than the texts themselves (Fang & Schleppegrell, 2010).

Thus, researchers have advocated for studies that take into account ways in which language is used to construe knowledge in different disciplines (Sanders, Land, & Mulder, 2007). In regards to the language challenges, Snow (2008) suggests that there are two main reasons that science texts are challenging to students: Their sophisticated vocabulary and the linguistic structures of these texts. Osborne (2002) echoes this view that properties of the language contribute to obstacles students face in learning science. This is compounded by the fact that almost all of what we call "scientific knowledge" is based on language (Hines, Wible & McCartney, 2010) and that, ultimately, *doing* science depends on being able to *communicate* science (Lemke, 1990).

The use of large corpora in this study provides a means of quantifying the presence, absence, and variety of linguistic features in textbooks. For example, Freebody and Muspratt (2007) used a large corpus of high school science textbooks in Australia to characterize differences across sub-domains within science on the basis of three variables: vocabulary diversity, nature and use of high-frequency function words, and nature of different syntactic constructions. These authors found that different science disciplines showed different linguistic patterns (e.g. high vocabulary diversity in biology vs. low vocabulary diversity in physics). With the exception of studies like these, the use of corpora remains limited (Chiappetta & Filman, 2007). To address this gap, this paper reports on a corpus study of 24 middle-school textbooks in two subject areas (science and social studies) from five publishers across three grade levels.

Research questions

Given the observations discussed above, this study specifically addressed the following questions:

RQ1. Using logical connectives as a proxy for semantic discontinuity, is there evidence of semantic discontinuity in science textbooks when compared with social studies textbooks?

RQ 2. What variation exists in the categories of overt logical connectives observed in science and social studies textbooks?

Methods

In order to conduct a large-scale comparison of the use of logical connectives across science and non-science textbooks, a corpus of 12 science and 12 social studies was compiled using textbooks adopted in California (Appendix A). As textbooks have different complementary sections (e.g., review and assessment), only the main texts were chosen for analysis. These textbooks were published by 5 different publishers, with 3 grade levels per publisher (6th, 7th, and 8th grade). Thus, we believe that this corpus is representative of texts that students would find in science classrooms at the middle school level.

Corpus

The corpus in its entirety consisted of 1,667,906 words across 125,831 sentences (55,005 sentences for science, 70,826 sentences for social studies). For this study, we treated sentence boundaries as an imperfect proxy for proposition boundaries.

Annotation and evaluation

For the purposes of this investigation, we adopted Fraser's (2006) categorization of relations and their associated connectives. Fraser's approach aims to semantically classify logical connectives (LC) according to their basic meaning and function (i.e. contrastive, elaborative, inferential, and temporal connectives) in discourse. The logical connectives that Fraser proposed for each category appear in Appendix B.

The unit of analysis was the sentence: In other words, does a particular sentence in a textbook contain a logical connective or not? Sentences were automatically extracted from the text using sentence-final punctuation (period, question mark, exclamation point) as separators. The sentences were automatically coded for the presence of a logical connective and the category of that logical connective.

In order to test what factors influence whether or not a logical connective was present and what type of connective was used, we conducted two separate logistic regression analyses. The first modeled the binary outcome of LC presence/absence in each sentence, testing how that outcome was influenced by two fixed factors and their interaction: subject area (Science/Social Studies) and grade level (6th/7th/8th). The second modeled the proportions of different connective types, testing how that proportion was influenced by three fixed factors and their interactions: connective category (contrastive/elaborative/inferential/temporal), subject (Science/Social Studies), and grade (6th/7th/8th).

Both models treated textbook section (as opposed to unit or chapter) as the unit of repeated measure, so that each sentence was associated with a unique ID representing the information of publisher/subject/unit/chapter/section. The first model predicted the raw binary outcome of LC presence/absence outcome in a mixed effects logistic regression with the fixed factors listed above as well as random effects for publisher and unique ID nested within publisher. For the second model, the raw data was collapsed by unique ID to form proportions (i.e., what proportion of sentences in a particular textbook section used a contrastive connective vs. an elaborative connective vs. an inferential connective vs. a temporal connective vs. no connective). The proportions were analyzed in a mixed effect linear regression and random by-publisher slopes for grade and LC category were included where possible.

Prior to analysis, binary and numeric fixed factors (subject area and grade level) were centered to reduce collinearity and to enhance the interpretability of estimates of coefficients. The 4-level categorical fixed factor for connective type was treatment coded with Temporal connectives as the reference group. All models were fit using the lmer function in the lmer4 package in R, using maximum-likelihood (ML) estimation. For each factor and interaction in the logistic regression, we report the coefficient estimate and p-value. For the linear regression, we calculated the coefficient estimate, the standard error, and the t-value. For the binary and fixed factors, we obtained the p-values using a model comparison approach, based on a likelihood-ratio X^2 (df= 1) test of the change in the goodness of fit between the full model and a comparison model in which only the relevant fixed effect or interaction was removed. We only report the t-value for the main effects and interactions of the categorical fixed factor for connective type, since model comparison is not possible. We assumed that t-values greater than 2 represent reliable effects.

Results

For the primary question of whether subject area influences logical connective presence/absence (i.e. semantic discontinuity), the logistic regression did indeed show a main effect of subject area; however, it was not in the direction predicted: Science texts showed a higher rate of LC usage (16.2% of sentences contained a LC) than Social Studies (SS) texts (14.1%; Coeff= -0.18; p<0.001). The effect of grade level was marginal with LC usage increasing slightly across 6th grade (14.4%), 7th grade (14.6%), and 8th grade (15.8%; Coeff=0.06; p=0.10). There was also a subject area×grade level interaction whereby science showed a greater increase across grade levels than SS (Coeff= -0.18; p<0.001). The pattern of results is shown in Table 1, broken down by subject area and grade level.

Table 1

Rates of Logical Connective Usage per Grade Level and Subject Area in Relation to the Total Number of Sentences (0=absent; 1=DM present)

Grade	Science	Social Studies
6th grade	0.149	0.140
7th grade	0.147	0.145
8th grade	0.190	0.138
Average	0.162*	0.141

*Significant at p<0.001

We also calculated the proportion of sentences in each section (with its unique ID) that contained a contrastive, elaborative, inferential, and temporal connective. Figures 1 and 2 show the distribution of connectives across LC categories and grade levels for science and SS, respectively. Note that, in each grade level for each subject area, the 4 colored bars (i.e., the 4 LC categories) sum together to give the overall rate of LC usage that is depicted in the corresponding cell of Table 1.

Figure 1

Types of logical connectives in science textbooks

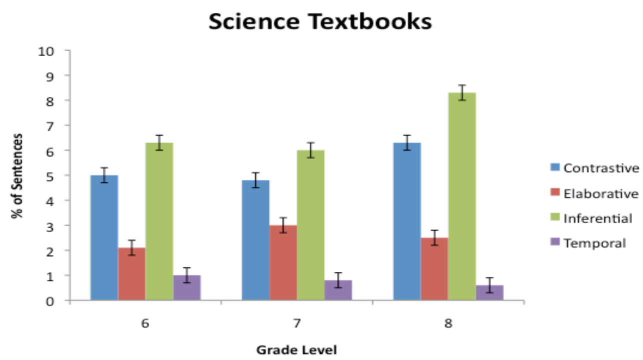
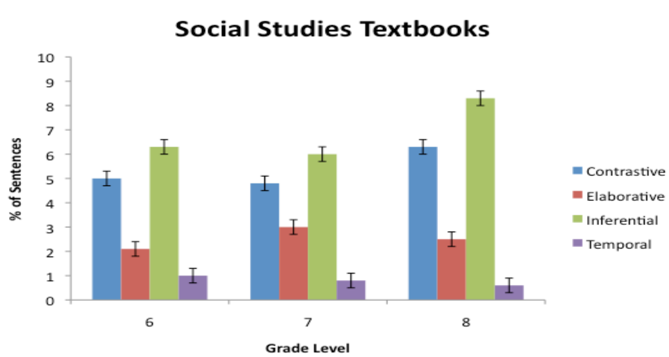


Figure 2

Types of logical connectives in social studies textbooks



For the second question of what factors influence the variation of LCs that are present, the linear regression again showed a main effect of subject area, whereby science yields higher proportions of LCs than social studies (Coeff= -0.12, SE=0.01, t-val=-11.82, $p<0.001$). The increase in the proportion of LCs across grade levels reached significance in this model (Coeff=0.03, SE=0.01, t-val=4.39, $p<0.005$). As Figures 1 and 2 show, not all LC categories were used with equal frequency. Setting temporal as the base group, the model showed that, compared to the grand mean, there were more contrastive connectives (Coeff=0.88, SE=0.02, t-val=36.18), more inferential connectives (Coeff=10.40; SE=0.03, t-val=13.05), and fewer elaborative connectives (Coeff= -0.40; SE=0.03, t-val=-11.72).

As Figures 1 and 2 also show, the most striking difference between the subject areas was the different rates of usage of the 4 LC categories. This is apparent in the subject area × connective category interactions: Again comparing to the grand mean, contrastive connectives were more frequent in social studies than science (Coeff= 0.42; SE=0.01, t-val=31.38), but elaborative and inferential connectives were more frequent in science than social studies (subject area×elaborative interaction: Coeff= -0.36; SE=0.02, t-val=-20.35; subject area × Inferential interaction: Coeff= -0.51, SE=0.02, t-val= -30.14). There were also differences by grade level. For one, overall proportions increased across grade levels for science but not social studies (subject area×grade level interaction; Coeff= -0.13, SE=0.01, t-val= -14.08, $p<0.05$). Additionally, the proportion of contrastive connectives increased over grade levels (Coeff=0.08; SE=0.01, t-val=15.21), whereas the proportion of elaborative connectives decreased over grade levels (Coeff= -0.04; SE=0.01, t-val= -6.05). The proportion of inferential connectives did not vary reliably by grade level when the data was collapsed across subject area (Coeff=0.01; SE=0.01, t-val=1.07), but the subject area×grade level×connective category interactions help clarify this.

The contrastive-connective increase over grade levels was stronger in social studies than science (subject area×grade level×Contrastive interaction: Coeff=0.11; SE=0.01, t-val=9.25), and the elaborative-connective decrease over grade levels was likewise stronger in social studies than science (subject area×grade level×elaborative interaction: Coeff= -0.15, SE=0.02, t-val=-9.99). Inferential connectives increased over grade levels and this increase was limited to science (subject area×grade level×inferential interaction: Coeff= -0.11; SE=0.01, t-val= -8.45).

Discussion

Contrary to the prediction that science textbooks would be characterized by greater semantic discontinuity (lower rates of logical connectives), the results reported here show that logical connectives are used more frequently in science than social studies texts. The variation across subject areas stems instead from the varying distributions of connective types: Science was found to favor elaborative (e.g., for example) and Inferential (e.g., therefore) connectives, whereas social studies favored contrastive connectives (Table 2). This may be a result of the differing content. Science texts describe scientific processes, which will require inferential connectives (e.g., *as a result* of mitosis…). Science texts must also provide many examples of the concepts they explore, resulting in more elaborative connectives (e.g., different biological systems, *such as* the digestive and nervous systems…). In contrast, social studies textbooks compare people, religions and events across time, resulting in more contrastive connectives.

Table 2

Five Most Common Inferential Logical Connectives and Percent of All Inferential Logical Connectives by Subject Area

Science			Social Studies		
Top 5	*n*	*%*	*Top 5*	*n*	*%*
So	665	45%	Then	621	54%
Then	327	22%	So	291	25%
When you	176	12%	Because of	131	11%
As a result	125	9%	Therefore	70	6%
Therefore	113	8%	Thus	43	4%

Consider this excerpt from a science textbook, [for clarity, the LC used as examples in the discussion section have been underlined]:

3 A blood cell cannot change into a skin cell, <u>for instance</u>. <u>However</u>, humans do produce certain cells –called stem cells—that can differentiate throughout life. Stem cells exist all around the body. These cells can respond to specific needs in the body by becoming specialized. <u>For example</u>, your body needs a constant supply of new blood cells to replace older cells (*Focus on Life Science*, Prentice Hall, 2008, p. 140).

The multiple logical connectives in (3) seem to guide the reader in comprehending the network of concepts addressed. First, the elaborative LC *for instance* is used to highlight an example of types of cells that cannot further differentiate. Then, the central idea of the text (that stem cells are the only human cells that can differentiate throughout life) is introduced with the contrastive connective *however* to signal stem cells' ability. Then, another elaborative logical connective *for example* is used to indicate one of the ways in which human bodies benefit from the ability of stem cells to become specialized cells.

The higher frequency of logical connectives in science textbooks could arise for a variety of reasons. For instance, secondary science textbooks have been described as presenting high amounts of factual information, in which many scientific concepts are explained (Halliday & Martin, 1993). In these cases, students are forced to understand the first concept introduced if they are to understand conceptual sequences. In (3), for instance, students would need to understand what stems cells are, the difference between stem cells and other cells, and the function of stem cells in human bodies including an example. As these concepts are not presented in this order, but rather contrast first, then stem cell definition, and finally an example, the authors might use LCs to signal how these concepts connect.

Additionally, science textbooks contain many experimental procedures, calculations, or model descriptions. In such instances, authors might use more logical connectives to help readers. In the following text excerpt (4), the contrastive LC *however* is used to signal the difference between the axes of a position-time graph. Meanwhile, the inferential LCs *as a result* and *then* are employed to point to the result and interpretation of the calculations:

4 On a position-time graph, the slope equals the rise over the run. However, the rise is the same as the distance traveled. The run equals the time needed to travel that distance. As a result, the slope of a line on a position-time graph equals the average speed. In Figure 16, the rise is equal to 60 m and the run is equal to 3 s. Then the average speed is 20 m/s (*Focus on Physical Science*, Glencoe, 2007, p. 67).

The higher frequency of LCs in science textbooks was due to the higher rates of elaborative and inferential logical connectives. As described above, it is possible that science writers use LCs to provide readers with a coherent way of connecting a high density of concepts presented in a limited amount of text. Additionally, writers may convey that science is about explaining the natural world, and they may do this by providing specific interpretations of phenomena.

Social studies textbooks authors, on the other hand, do not seem to rely on LCs to aid in the signaling of explanations and inferences as much as science counterparts. One explanation may be that, in social studies, pronouns and other referring expressions play a more important role in creating coherent texts by creating chains of references to relevant historical figures, locations, and events, as shown in (5):

5 Across the Mississippi River lay the unexplored territory of Louisiana. This immense region stretched from Canada in the north to Texas in the south. From the Mississippi, it reached west all the way to the Rocky Mountains. *First* claimed by France, it was given to Spain after the French and Indian War. In 1800, the French ruler Napoleon Bonaparte convinced Spain to return Louisiana to France (*History Alive! The United States through Industrialism*, TCI, 2011, p. 280).

Additionally, in instances in which a historical event is described in chronological order, authors might assume that readers would use dates to differentiate between events, rather than guiding the readers via logical connectives.

The results showed that overall proportions of LCs increased in science as grade level increased, but not in social studies. As logical connectives explicitly guide the reader in understanding the text (Mayer, 2007), the increase of LC in science might be due to the more complex material discussed in later grades. Additionally, the significant grade level and logical connective interaction that was found in science textbooks could support low knowledge readers who seem to benefit from explicit texts (Kamalski, Sanders & Lentz, 2008). The same can be said about the increase of contrastive connectives identified in social studies textbooks in higher grades.

Finally, the overall greater number of logical connectives found in science textbooks may also reflect Freebody and Muspratt's (2007) findings regarding differences in the language of different scientific disciplines. They found the language of physics and chemistry seemed to convey organized taxonomies, while biology and geology were characterized by descriptive accounts of phenomena. In California, as geology is taught in 6^{th} grade as part of Earth science, biology in 7^{th} grade as part of life science, and physics and chemistry as part of physical science in 8^{th} grade, our findings could indicate that the differences in the language of these disciplines is also reflected in the textbooks we analyzed (i.e. more inferential connectives in physics and chemistry in 8^{th} grade when more theoretical constructs are discussed compared with fewer inferential connectives in geology and biology in 6^{th} grade and 7^{th} grade, when more descriptions are presented).

Implications for Classrooms, Teacher Certification and Professional Development

This type of research could also serve to refine professional development and teacher certification programs to help science educators tailor instruction to support comprehension of textbooks (Kucer, 2011). As logical connectives explicitly guide the reader in understanding the text (Sanders, Land, & Mulder, 2007), having an adequate description of the ways logical connectives are used in academic textbooks can aid in the development of content-area specific reading and writing strategies. Effective science teachers incorporate authentic literacy activities into their classroom to promote content learning (Douglas, Klentschy, Worth & Binder, 2006), so teacher preparation and professional development must aid science teachers in learning to integrate the language and literacy of science with the content.

Bunch (2013) reminds us that the Common Core State Standards require "increasingly demanding uses of language and literacy" (p. 298), and argues for teachers' acquisition of pedagogical language knowledge, or the ability and knowledge required to support students' comprehension and acquisition of the language required by the subject being taught. Similarly, Quinn, Lee & Valdés (2012) argue, " We do not suggest that science teachers should be function as language teachers, but rather as supporters of the language learning that occurs in content-rich and discourse-rich classroom environments" (p. 1). Science teachers should provide opportunities for students to use the language of science by engaging in science tasks –including designing arguments from evidence, developing explanations and designing solutions, and obtaining, evaluating and communicating information– in order to achieve "science understanding and science communication" (p. 5).

While logical connectives are intended to support comprehension by providing syntactical hints (e.g., *however* and *in contrast* change the direction of a sentence, idea or paragraph), many are also tier two vocabulary words (Beck, McKeown & Kucan, 2002). For example, words such as *consequently* and *additionally* might be new to students, just as tier 3 vocabulary words (e.g., genes, mitochondria) might be new. If signal words are unknown, it is unlikely that students would be able to use those words to support comprehension. Therefore, logical connectives might need to be taught both as new vocabulary and as signal words, helping the reader to follow the direction of the argument.

Logical connectives can also be used to help students understand how sentences and texts work. Explicit instruction, possibly through sentence combining, would help students to understand how logical connectives are used. Sentence combining has been found to be helpful for developing students' facility with more complex sentence structures (Berninger, Nagy, & Beers, 2011; Limpo & Alves, 2013), as well as for students with special needs (Saddler, 2005; Saddler, Behforooz & Asaro, 2008; Saddler & Preschern, 2007). "Sentence combining provides direct, mindful practice in manipulating and rewriting basic or kernel sentences into more syntactically mature or varied forms," explained Saddler (2005, p. 468).

For example, two sentences might be provided, such as an adaptation from (3), "A blood cell cannot change into a skin cell," and, "Humans do produce certain cells –called stem cells— that can differentiate throughout life." Students would then be asked to combine the sentences using an appropriate logical connective. A possible result might be, 'A blood cell cannot change into a skin cell, <u>however</u> humans do produce certain cells –called stem cells—that can differentiate throughout life.' Once students are familiar with different forms of logical connectives in science texts, examples from texts could be used as mentor texts to help students construct their own argumentation based on evidence (Gallagher, 2011).

Among the literacy skills students require to access science concepts, Yore and Shymansky (1991) stated to "read about science is a critical skill to have in order to develop scientific literacy" (p. 29). Since an understanding of discourse requires readers to construct a mental representation of the text, students must understand how different ideas are connected to comprehend the concepts. The findings we discussed in this paper about discourse-connective usage offer the education community crucial information by addressing the variation and challenges inherent to the language that students encounter in textbooks.

First of all, it is worth considering that overall only 16.2% in science and 14.1% in social studies in the corpus analyzed contained logical connectives. These frequencies of LCs subjects should be evaluated vis-à-vis research that has emphasized the importance of LCs in aiding comprehension (Mayer, 2007). One of the open questions for future research would consist of finding adequate descriptions of the frequencies of logical connectives in different disciplines and across grade levels. Additionally, psycholinguistic studies could be developed to quantify the adequate balance between overt and implicit discourse connectives that a text should have to be coherent for students of different reading levels. This could be particularly informative to publishers because "if research shows that text characteristics lead to processing difficulties, they should be avoided in effective text design" (Sanders, Land, & Mulder, 2007, p. 221).

Our results also showed that science textbooks at the middle school level do not present more semantic discontinuity when compared against their social studies counterparts. This finding, measured by the frequency of logical connectives, does not reflect Halliday and Martin's (1993) description of problems of scientific English. Furthermore, additional studies might compare findings against oral discourse in science classrooms to identify similarities and differences with text.

In regards to the findings about the categories of logical connectives analyzed, the higher frequency of inferential connectives and lower frequency of contrastive connectives found in science texts might indicate, for instance, that science texts present science mostly as a set of results rather than as processes, in which differing opinions have to be reconciled. Thus, science teachers may need to supplement the textbook with instruction and materials that model the crucial role that argumentation plays in scientific findings (Wellington & Osborne, 2001). Additionally, the significant grade level and logical connective interaction that was found in science textbooks, in which the number of inferential connectives increases in the higher grade levels, could help low knowledge readers who benefit from more explicit texts (Kamalski, Sanders, & Lentz, 2008).

To conclude, more studies should be conducted in ecologically valid contexts (e.g., schools), in which, researchers test students' reading comprehension using comparable texts that only vary in the number of logical connectives. We advocate for more interdisciplinary projects between educational researchers and linguists, such as the one that resulted in this paper, that could evaluate lexical, syntactic, and discourse level claims that exist around the language of schooling and could guide the development of effective literacy practices.

References

Alberts, B. (2010). Prioritizing science education. *Science, 23*(328), 405.

Asher, N., & Lascarides, A. (2003). *Logics of conversation*. Cambridge: Cambridge University Press.

Bailey, A. L., Butler, F. A., Stevens, R., & Lord, C. (2007). Further specifying the language demands of school. In A. L. Bailey (Ed.), *The language demands of school: Putting academic English to the test* (pp. 103–156). New Haven, CT: Yale University Press.

Beck, I. L., McKeown, M. G. & Kucan, L. (2002). *Bring words to life: Robust vocabulary instruction*. New York: Guilford.

Best, R.M., Rowe, M., Ozuru, Y., & McNamara, D. (2005). Deep-level comprehension of science texts: The role of the reader and the text. *Topics in Language Disorders, 25*, 65–83.

Berninger, V.W., Nagy, W., & Beers, S. (2011). Child writers' construction and reconstruction of single sentences and construction of multi-sentence texts: Contributions of syntax and transcription to translation. *Reading and Writing, 24*(2), 151–182.

Biber, D., Conrad, S., & Cortes, V. (2004). "If you look at…": Lexical bundles in university teaching and textbooks. *Applied Linguistics*, 25(3), 371-405. doi: 10.1093/applin/25.3.371

Blakemore, D. (2002). *Relevance and linguistic meaning: The semantics and pragmatics of discourse markers.* Cambridge: Cambridge University Press.

Bunch, G. (2013). Pedagogical language knowledge: Preparing mainstream teachers for English Learners in the new standards era. *Review of Research in Education, 37*(1), 298-341.

Chiappetta, E. L., & Fillman, D. A. (2007). Analysis of five high school biology textbooks used in the United States for inclusion of the nature of science. *International Journal of Science Education*, *29*(15), 1847-1867.

Common Core Standards Initiative (2010). *Common Core State Standards for English Language Arts & Literacy in History/Social Studies, Science, and Technical Subjects*. Retrieved from http://www.corestandards.org/the-standards

Conrad, S. (1996). Investigating academic texts with corpus-based techniques: An example from biology. *Linguistics and Education, 8*, 299-326.

Degand, L., & Sanders, T. (2002). The impact of relational markers on expository text comprehension in L1 and L2. *Reading and Writing, 15*(7-8), 739-757.

Dockrell, J., Braisby, N., & Best, R. (2007). Children's acquisition of science terms: Simple exposure is insufficient. *Learning and Instruction, 17*, 577-594.

Douglas, R., Klentschy, M. P., Worth, K., & Binder, W. (Eds). (2006). *Linking science and literacy in the K-8 classroom*. Arlington, VA: National Science Teachers Association.

Fang. Z., & Schleppegrell, M. J. (2010). Disciplinary literacies across content areas: Supporting secondary reading through functional language analysis. *Journal of Adolescent & Adult Literacy, 53*(7), 587-597. doi: 10.1598/JAAL.53.7.6

Fisher, D., Grant, M., & Frey, N. (2009). Science Literacy Is > Strategies. *Clearing House, 82*(4), 183-186.

Frantz, R. S., Starr, L. E., Bailey, A. L. (2015). Syntactic complexity as an aspect of text complexity. *Educational Researcher, 44*(7), 387-393. DOI: 10.3102/0013189X15603980

Fraser, B. (2006). Towards a theory of discourse markers. Approaches to Discourse *Particles*. New York: Elsevier.

Freebody, P., & Muspratt, S. (2007). Beyond generic knowledge in pedagogy and disciplinarity. *Pedagogies: An International Journal, 2*(1), 35-48.

Gallagher, K. (2011). *Write like this: Teaching real-world writing through modeling & mentor texts*. Portland, ME: Stenhouse Publishers.

Gardner, P. L. (1975). Logical connectives in science: A preliminary report. *Research in Science Education*, *5*(1), 161-175.

Glencoe/McGraw-Hill. (2007). *Focus on Physical Science* (California ed.). Columbus, OH: Glencoe/McGraw-Hill.

Hakuta, K., Santos, M., & Fang, Z. (2013). Challenges and opportunities for language learning in the context of the CCSS and the NGSS. *Journal of Adolescent & Adult Literacy, 56*(6), 451-454.

Halliday, M., & Hasan, R. (1976). *Cohesion in English.* New York: Longman Inc.

Halliday, M., & Martin, J. (1993). *Writing science*. Pittsburg: University of Pittsburgh Press.

Hines, P. J., Wible, B., & McCartney, M. (2010). Learning to read, reading to learn. *Science, 23*(328), 447. doi: 10.1126/science.328.5977.447

Kamalski, J., Sanders, T., & Lentz, L. (2008). Coherence marking, prior knowledge, and comprehension of informative and persuasive texts. *Discourse Processes, 45*(4-5), 323-345.

Kucer, S. (2011). Going beyond the author: what retellings tell us about comprehending narrative and expository texts. *Literacy, 45*(2), 62-69.

Lemke, J. (1990). *Talking science: Language, learning, and values*. Westport, CT: Ablex.

Limpo, T., & Alves, R. A. (2013). Teaching planning or sentence-combining strategies: Effective SRSD interventions at different levels of written composition. *Contemporary Educational Psychology, 38*(4), 328–341. http://doi.org/10.1016/j.cedpsych.2013.07.004

Mayer, R. (2007*). Learning and instruction.* Upper Saddle River, NJ: Prentice Hall.

McNamara, D. S. (2001). Reading both high and low coherence texts: Effects of text sequence and prior knowledge. *Canadian Journal of Experimental Psychology, 55,* 51-62.

McTigue, E. & Slough, S. W. (2010). Student accessible science texts: Elements of design. *Reading Psychology, 31*(3), 213-227.

Meyer, B. J. F. (2003). Text coherence and readability. *Topics in Language Disorders, 23*(3), 204-224.

Moje, E., & Speyer, J. (2008). The reality of challenging texts in high school science and social studies. In K. Hinchman & H. Sheridan-Thomas (Eds.), *Best practices in adolescent literacy instruction* (pp. 185-211). New York: Guilford.

Moje, E. B., Stockdill, D., Kim, K., & Kim, H. J. (2010). 20 The Role of Text in Disciplinary Learning. *Handbook of reading research*, *4*, 453.

NGSS Lead States. (2013). Next Generation Science Standards: For states, by states. Washington DC: National Academies Press.

Norris, S. P., & Phillips, L. M. (2003). How literacy in its fundamental sense is central to scientific literacy. *Science Education, 87*(2), 224-240. doi: 10.1002/sce.10066

Osborne, J. (2002). Science without literacy: A ship without a sail? *Cambridge Journal of Education, 32*(2), 203-218.

Ozuru, Y., Dempsey, K. & McNamara, D. S. (2009). Prior knowledge, reading skill and text cohesion in the comprehension of science texts. *Learning and Instruction, 19*, 228-242.

Pearson, P. D., Moje, E., & Greenleaf, C. (2010). Literacy and science: Each in the service of the other. *Science, 23*(328), 459-463. Doi: 10.1126/science.1182595

Prentice Hall. (2008). *Focus on Life Science* (CA Ed.). Upper Saddle River, NJ: Pearson Prentice Hall.

Quinn, H., Lee, O. & Valdés, G. (January, 2012). *Language demands and opportunities in relation to Next Generation Science Standards for English language learners: What teachers need to know*. Paper presented at the Understanding Language Conference, Stanford, CA.

Román, D., & Busch, K.C. (2015). Textbooks of doubt: Using systemic functional analysis to explore the framing of climate changes in middle-school textbooks. *Environmental Education Research*, DOI: 10.1080/135004622.2015.1091878

Saddler, B. (2005). Sentence combining: A sentence-level writing intervention. *The Reading Teacher, 58*(5), 468-471).

Saddler, B., Behforooz, B., & Asaro, K. (2008). The effects of sentence-combining instruction on the writing of fourth-grade students with writing difficulties. *The Journal of Special Education, 42*(2), 79–90. http://doi.org/10.1177/0022466907310371

Saddler, B. & Preschern, J. (2007). Improving sentence writing ability through sentence-combining practice. *Teaching Exceptional Children, 39*(3), 6-11.

Sanders, T., Land, G., & Mulder, G. (2007). Linguistic connectives of coherence improve text comprehension in functional contexts. *Information Design Journal, 15*(3), 219-235.

Shanahan, T., & Shanahan, C. (2008). Teaching disciplinary literacy to adolescents: Rethinking content-area literacy. Harvard Educational Review, 78(1), 40-59.

Slough, S. W., McTigue, E. M., Kim, S., & Jennings, S. K. (2010). Science textbooks' use of graphical representation: A descriptive analysis of four sixth grade science texts. *Reading Psychology*, *31*(3), 301-325.

Smolkin, L. B., McTigue, E., Donovan, C. A., & Coleman, J. (2009). Explanation in science trade books recommended for use with elementary students. *Science Education, 93*, 587–610.

Snow, C. (2008). What is the vocabulary of science? In A. Rosebery and B. Warren (Eds.), *Teaching science to English language learners* (pp. 71-83). Arlington, VA: NSTA Press.

Taboada, M., & Mann, W. C. (2006). Rhetorical Structure Theory: Looking back and moving ahead. *Discourse Studies, 8*(3), 423-459.

TCI. (2011). *History Alive! The United States Through Industrialism.* Rancho Cordova, CA: Teachers' Curriculum Institute.

Turkan, S., De Oliveira, L., Lee, O., & Phelps, G. (2014). Proposing a knowledge base for teaching academic content to English Language Learners: Disciplinary linguistic knowledge. *Teachers College Record, 116,* 1-30.

Wellington, J., & Osborne, J. (2001). *Language and literacy in science education*. Glasgow: Open University Press.

Yore, L. D., & Shymansky, J. A. (1991). Reading in science. *Journal of Science Teacher Education, 1*(2), 29-36.

Appendix A: List of Textbooks Analyzed in Corpus

CPO Science. (2007). *Focus on Earth Science*. (CA Ed.). Cambridge, MA: Cambridge Physics Outlet.

CPO Science. (2007). *Focus on Life Science*. (CA Ed.). Cambridge, MA: Cambridge Physics Outlet.

CPO Science. (2007). *Focus on Physical Science* (CA Ed.). Cambridge, MA: Cambridge Physics Outlet.

Glencoe/McGraw Hill. (2006). *Discovering Our Past: Ancient Civilizations*. (California ed.). Columbus, OH: Glencoe/McGraw-Hill.

Glencoe/McGraw Hill. (2006). *Discovering Our Past: Medieval and Early Modern Times* (California edition). Columbus, OH: Glencoe/McGraw-Hill.

Glencoe/McGraw Hill. (2006). *Discovering Our Past: The American Journey to World War I.* (California ed.). Columbus, OH: Glencoe/McGraw-Hill.

Glencoe/McGraw-Hill. (2007). *Focus on Earth Science* (California edition). Columbus, OH: Glencoe/McGraw-Hill.

Glencoe/McGraw-Hill. (2007). *Focus on Life Science* (California edition). Columbus, OH: Glencoe/McGraw-Hill.

Glencoe/McGraw-Hill. (2007). *Focus on Physical Science* (California edition). Columbus, OH: Glencoe/McGraw-Hill.

Holt, Rinehart, and Winston Inc. (2007). *Earth Science* (CA Ed.). Boston, MA: Houghton Mifflin Harcourt Publishing Company.

Holt, Rinehart, and Winston Inc. (2007). *Life Science* (CA Ed.). Boston, MA: Houghton Mifflin Harcourt Publishing Company.

Prentice Hall. (2006). *Ancient Civilizations* (CA Ed.). Upper Saddle River, NJ: Pearson Prentice Hall.

Prentice Hall. (2008). *Focus on Earth Science* (CA Ed.). Upper Saddle River, NJ: Pearson Prentice Hall.

Prentice Hall. (2008). *Focus on Life Science* (CA Ed.). Upper Saddle River, NJ: Pearson Prentice Hall.

Prentice Hall. (2008). *Focus on Physical Science* (CA Ed.). Upper Saddle River, NJ: Pearson Prentice Hall.

Prentice Hall. (2006). *Medieval and Early Modern Times* (CA Ed.). Upper Saddle River, NJ: Pearson Prentice Hall.

TCI. (2011). *History Alive! The Ancient World*. Palo Alto, CA: Teachers' Curriculum Institute.

TCI. (2011). *History Alive! The Medieval World and Beyond.* Palo Alto, CA: Teachers' Curriculum Institute.

TCI. (2011). *History Alive! The United States Through Industrialism.* Palo Alto, CA: Teachers' Curriculum Institute.

Appendix B: List of Logical connectives

Contrastive Connectives
but, alternatively, although, conversely, despite (this/that), even so, however, in spite of, in comparison, in contrast, instead, nevertheless, nonetheless, notwithstanding, on the other hand, on the contrary, rather, regardless, still, though, whereas, yet, similarly

Elaborative Connectives
above all, alternatively, analogously, besides, correspondingly, equally, for example, for instance, further, furthermore, in addition, in other words, in particular, likewise, more accurately, more importantly, more precisely, moreover, on that basis, otherwise, rather, similarly

Inferential Connectives
after all, all things considered, as a conclusion, as a consequence of, as a result, because, because of, consequently, for this reason, for that reason, hence, it follows that, in this/that/any case, on this/that condition, on these/those grounds, so, then, therefore, thus, when you, equally

Temporal Connectives
eventually, finally, immediately, afterwards, in the meantime, meanwhile, originally, subsequently, lastly