Listening. Learning. Leading.®

Research Report
ETS RR-12-21

The *Language Muse*SM System: Linguistically Focused Instructional Authoring

Jill Burstein

Jane Shore

John Sabatini

Brad Moulder

Steven Holtzman

Ted Pedersen

October 2012

ETS Research Report Series

The *Language Muse*^{SM} System: Linguistically Focused Instructional Authoring

Jill Burstein, Jane Shore, John Sabatini, Brad Moulder, and Steven Holtzman
ETS, Princeton, New Jersey

Ted Pedersen
University of Minnesota, Duluth

October 2012

Associate Editor: Joel Tetreault

Reviewers: Yoko Futagi and Tenaha O'Reilly

Abstract

In the United States, English learners (EL) often do not have the academic language proficiency, literacy skills, cultural background, and content knowledge necessary to succeed in kindergarten through 12th grade classrooms. This leads to large achievement gaps. Also, classroom texts are often riddled with linguistically unfamiliar elements, including: unfamiliar vocabulary, idioms, complex phrases or sentences, morphologically complex words, and unfamiliar discourse relations. Lack of familiarity with linguistic elements may result in gaps in a learner's comprehension of key content. It is *not* feasible for teachers to develop additional curriculum for the needs of all ELs in a classroom (who often come from culturally and linguistically diverse backgrounds.) However, it *is* feasible for teachers to develop instructional scaffolding (support) that helps ELs and can be used with all students. To develop effective scaffolding, teachers need to be able to reliably identify linguistic features in texts that could interfere with content comprehension. *Language Muse*^{SM} is a web-based application designed to support teachers in the identification of linguistic features in texts and in the development of linguistically focused instructional scaffolding. With regard to system itself, we will discuss (a) the system's motivation, (b) the system's linguistic feedback and instructional authoring components, which are driven by natural language processing, and (c) the system's infrastructure for capturing teachers' system use. In addition, we will also discuss preliminary pilot study findings with three teacher professional development programs. These findings suggest that exposure to Language Muse's linguistic feedback can support teachers in the development of lesson plan scaffolds designed to address language learning needs.

Key words: English language learning, natural language processing, educational technology, teacher education

Acknowledgments

This research presented in this paper is based upon work supported by the Institute of Education Science, U.S. Department of Education (IES), under Award No. R305A100105. Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the IES.

We have many people to acknowledge who contributed to this project. We are extremely grateful to our partners at Stanford University, George Washington University, and Georgia State University without whom we could not have conducted this research. At Stanford University, we would like to thank the following individuals: Dr. Kenji Hakuta, Martha Castellon, Diego Roman, Melissa Bennett, Shelly Frei, Christina McColley, and Kimberly Govea; at George Washington University, Dr. Amy Mazur, Dr. Patty Rice-Doran (now affiliated with Towson University, Towson, MD), Meghan Mulhern, and Charmaine Llagas; and at Georgia State University, Dr. Laura May.

We would like to thank Dr. Jeffrey Zwiers for writing instructional materials for the Language Muse intervention, and for support in scoring rubric development and scoring teacher data. At ETS, we would also like to thank Jennifer Lentini for support in scoring rubric development and scoring teacher data, and Elia Mavronikolas for scoring teacher data. At ETS, we are very grateful to the following people for critical technical support in the development of the Language Muse system: Klara Brodsky, Ramin Hemat, Chi Lu, and Srinivasa (Pavan) Pillarisetti. We would also like to acknowledge Jim Dempsey, Donna Herb, Steve Hendershott, and Mike Wagner for invaluable support in the development and management of the system architecture. We are very grateful to Jennifer Minsky for invaluable data analysis support.

Table of Contents

List of Tables

Page

List of Figures

Page

Background

The focus on all learners to read progressively more complex texts in the content areas, especially as students approach their college years, has been more recently emphasized by the Common Core State Standards initiative (Common Core). This state-led initiative is coordinated by the National Governors Association Center for Best Practices (NGA Center) and the Council of Chief State School Officers (CCSSO); the initiative has now been adopted by over 45 states for use in kindergarten through 12th grade (K-12) classrooms. That said, this initiative is likely to have a strong influence with regard to teaching standards in K-12 education. The Common Core standards describe what K-12 students should be learning with regard to reading, writing, speaking, listening, language, and media and technology. Specifically, these standards propose that *all* learners should be reading progressively more complex texts in preparation for college and that they should be continually developing their vocabulary, and understanding of word and phrase nuances (senses), and language conventions. The Common Core recently released a publishers' criteria document designed for publishers and curriculum developers that describes the type of complex elements that learners should be able to handle as they progress to the higher grades (Coleman & Pimental, 2011a, 2011b). These criteria explicitly specify that learners need to have a grasp of a number of linguistic features related to vocabulary, grammar standards and conventions, and argument structure in texts in the content areas.

An emphasis on text-based learning in curriculum standards as proposed by the Common Core is clearly becoming influential in the development of curriculum standards in the United States. At the same time, English learners (EL) in the United States often do not have the academic language proficiency, literacy skills, and cultural background and content knowledge necessary to succeed in K-12 classrooms (Center for Public Education, 2007). This creates large achievement gaps, especially for learners beyond elementary school, when the emphasis switches from *learning to read* to *reading to learn* (Francis, Rivera, Lesaux, Kieffer, & Rivera, 2006). When the goal of instruction is to teach subject-area content, the lack of familiarity with linguistic structures in a text should not interfere with content comprehension. However, this can happen, especially for ELs, when classroom texts are riddled with linguistic challenges, including: words above a learner's reading level, idioms, complex phrases or sentences, and unfamiliar or complex discourse relations that may result in gaps in explanation about key content. Further, the number of teachers trained in effective instructional strategies to meet the range of needs of ELs has not necessarily increased consistently with the rate of the EL population (Calderón & Minaya-Rowe, 2007; Gándara, Maxwell-Jolly, & Driscoll, 2005; Green, Foote, Walker, & Shuman, 2010).

One might question why use of the standard, readily available readability measures (Chall & Dale, 1995; Flesch, 1948; Stenner, 1996) does not provide a definitive solution with regard to text selection to accommodate students with specific language proficiency needs (e.g., reading below grade level). There are a number of reasons why it is not desirable for teachers to rely solely on these kinds of measures. First, there has been no empirical evidence that would suggest that these measures would facilitate the selection of appropriate alternative lower-level texts for a culturally and linguistically diverse group, such as ELs. Second, a single classroom can potentially have ELs from several cultural and linguistic backgrounds with varying levels of English proficiency. It may not even be feasible from a teacher's perspective to find a sufficient number of alternative texts that suit the needs of all individual learners. Third, all learners are expected to learn the content of the texts specified in the curriculum. Even if teachers are able to find alternative texts, there is also no guarantee that the text will have the complete set of content as specified by the curriculum standard. In addition, and consistent with the Common Core philosophy, it is pedagogically desirable to use instructional scaffolding methods with the goal of helping students become more proficient readers. Assigning texts at a lower grade level will not guarantee this outcome. What is really needed is to offer instructional scaffolding to learners that will teach them about the different kinds of linguistic features in a text. This approach will help them to understand how to read beyond reading one particular text and, over time, how to read texts with similar, and eventually greater linguistic variability. So, for instance, if a teacher instructs a learner about how to understand the structure of complex sentences in one text, the teacher can handle those types of complex sentences in the next text he or she comes across. That said, it is both feasible and pedagogically acceptable to develop instructional scaffolding that supports ELs, but can be used with all students. Finally, how meaningful are readability measures? The readability measure does not offer explicit feedback about linguistic elements in text and what is rendering a text more or less difficult. These measures also have no features that can tap into students' background knowledge, which may also be a driver of text difficulty. So, while alternative texts may exist about Christopher Columbus at a lower grade level, there is no guarantee that the alternative text that would not still mask key content for the learner, especially for an EL, because there is no guarantee that the text would be free of linguistically complex elements and conceptual obstacles to an individual student's background knowledge.

In the end, in the spirit of the teach a man to fish analogy, it makes sense to offer instruction about linguistic structures to all teachers, so that they are keenly aware of linguistic elements that may be unfamiliar and may interfere with learners' content comprehension. Equipped with this knowledge, teachers can develop appropriate instructional scaffolding, as needed. This perspective is also supported by teacher professional development programs offering certification to teach ELs. Three such programs at Stanford University, George Washington University, and Georgia State University will be discussed later in this paper. All three programs offer instruction to support teachers' understanding of linguistic features that may interfere with EL content comprehension. Preliminary findings from our pilot studies with *Language Muse*^{SM} suggest that integration of the system into these programs can support teachers in the development of lesson-plan scaffolding designed to address language learning needs.

Instructional Scaffolding Methods for English Learners

Reading is the medium through which students acquire much of their knowledge and understanding of the different subject areas, and out-of-class reading frequently forms the basis for class discussions or homework. Therefore, especially with ELs, teachers find text scaffolding methods to be critical. The idea is to provide students with a framework to access content, but not to remove the language learning potential in a text. Direct scaffolding (modification of a text with additional support) of academic content in a text can aid in the development of instruction that supports the needs of specific learners. A number of scaffolding methods are described below.

A number of research studies have suggested that elaboration of text, for example, inserting simple definitions for key concepts and important elements, can aid in vocabulary development (Hancin-Bhatt & Nagy, 1994; James & Klein, 1994), and such elaboration, or another scaffolding technique, linguistic simplification, can facilitate students' comprehension of content (Bean, 1982; Carlo et al., 2004; Fitzgerald, 1995; Francis, August, Goldenberg, & Shanahan, 2004; Ihnot, 1997; Jiménez, Garcia, & Pearson, 1996; Perez, 1981; Yano, Long, & Ross, 1994). Linguistic simplification requires teachers to go into a text and revise elements, including aspects of language like complex syntax, vocabulary, or even logic or presentation (teachers may need to improve the writing of a text), in order to make a text more coherent for learners. Elaboration might also refer to *native language support,* which, when used appropriately, can aid students in learning from text-based content (Francis et al., 2004). Texts might be translated, or *cognates*, words in two languages that are derived from the same root, inserted as support for ELs. These techniques have been found effective in both expanding English vocabulary development and aiding in comprehension of complex texts (August, 2003; Nagy, Garcia, Durgunoglu, & Hancin-Bhatt, 1993). In practice, most text modifications involve a combination of simplification and elaboration, as well as a mixture of techniques that modify language and concepts aligned with curriculum needs and the individual needs of learners.

In addition to the modification of a text directly, instructional scaffolding might be introduced at a higher level—the curricular level. At this level, teachers modify instruction with additional instructional strategies. These might be thought of as curricular modifications (Koga & Hall, 2004). Modifications might be chosen based on formative assessment administered in a classroom. These kinds of assessments determine students' need for additional preparation, perhaps directed at specific language, cultural, or historical background knowledge. This kind of modification prepares learners for a new task or text (Sparks, 2000; Switick, 1997). For example, one such enhancement that has been found to lead to improve vocabulary development involves classroom activities that focus on morphologically complex words. ELs might be presented with classroom activities in which they will work directly with prefixes, stems, and suffixes. As they learn about morphological structure, this can contribute to their understanding of future unknown words (Kieffer & Lesaux, 2007). Further, providing ELs with information about academic and content-specific vocabulary and designing instruction related to this information, can help develop learners' knowledge about the multiple ways that words might be used across content areas. For instance, teaching learners about polysemous words might help them to understand that the word *plant* in a science text about *photosynthesis* will have a different meaning that than the use of the word *plant* in a social studies text, where this may be in reference to a *factory*.

These modifications, whether simplifying or enhancing a student's reading experience, are not meant to replace basic reading strategies a teacher might incorporate to support the learning of culturally and linguistically diverse students. Teachers of ELs might also incorporate questioning techniques or have students complete activities that require direct interaction with a text. For instance, students may be asked to summarize, rewrite, create, or choose a proper a visual representation or simply to ask and answer questions about a text (Biancarosa & Snow, 2004). These techniques may address a variety of learner needs and can further contribute to improved educational outcomes for ELs.

Motivation for the Language Muse System

To understand how to effectively implement instructional scaffolding, either directly in a text or in the form of supplemental classroom and homework activities, teachers first need to be able to recognize linguistic structures. Further, teachers must also have training about which linguistic structures might be unfamiliar to learners. Even with a strong linguistic awareness, if teachers have to read through texts and manually identify all of the linguistic elements that may be unfamiliar to learners, this is likely to be an extremely time-consuming task. The motivation for the Language Muse system grew from the apparent need to provide teachers with training about linguistic features in texts that may be unfamiliar to learners and to offer support to teachers that would allow them to get linguistic feedback about texts in an efficient way. Natural language processing (NLP) methods can support both of these needs. NLP methods can be used to automatically highlight relevant linguistic features in text, providing explicit feedback that can support teachers in developing scaffolded curriculum materials (texts, activities, and assessments) to better support learners' reading needs.

The Language Muse system is a web-based application designed to support teachers in the development of linguistically focused instructional authoring of content-area curriculum (Burstein, Sabatini, & Shore, in press; Shore, Burstein, & Sabatini, 2009). The application uses a suite of NLP capabilities to offer teachers explicit feedback about linguistic structures in texts to help them to develop linguistic awareness intended to support their curriculum development needs, including the development of lesson plans, scaffolded texts, activities, and assessments. In this report, we will discuss (a) the Language Muse system's motivation with regard to curriculum development to support ELs; (b) the system's specific instructional authoring components, including tools for developing lesson plans with associated activities and assessments, and a text exploration tool that uses NLP capabilities to provide explicit feedback about linguistic structures in texts; and, (c) the system's infrastructure that captures information about how teachers use the system. System use will be discussed in the context of pilot studies in three teacher professional development settings at Stanford University, George Washington University, and Georgia State University. The article will discuss pilot outcomes suggesting that explicit linguistic feedback provided by the NLP capabilities in the Language Muse system supports teachers in becoming more aware of linguistically unfamiliar structures, and in the development of instructional scaffolding that is directly connected to these structures.

The Lesson Planning Process

Previous research suggests that if teachers are likely to adopt a new technology, it needs to support and enhance their daily routine (Burstein, 2009; Burstein, Shore, Sabatini, Lee, & Ventura, 2007). The Language Muse system is intended to fit into the traditional lesson planning process. While teachers may have different lesson plan development styles, the five parts described below characterize the critical components that typically would be in a teacher's lesson plan:

1. **Identify and describe the curriculum standards and lesson objectives**. Curriculum standards typically describe what content will be taught and what aspects of language should be addressed in the lesson. Teachers typically include state standards in their plans. Lesson objectives relate to a particular standard and describe the goal of the particular lesson. Language standards and objectives may be specifically related to teaching ELs. An example of such a language standard or objective might look something like this: Students will be able to use cause-effect transitions terms in discussions, reading, and writing.

2. **Specify formative and summative assessments.** In this aspect of the lesson plan, teachers consider what kinds of assessments they will use to evaluate learners' incremental progress (formative) and their final progress (summative).

3. **Engage student background knowledge and interest.** Teachers need to develop activities with learners to draw on that background knowledge and to get students interested in a topic. Examples might include an activity that preteaches key vocabulary from the text using visuals.

4. **Develop guided practice.** Here, teachers design activities where they show students how to do something. For example, teachers may have the class review possible cause-effect relationships in history and model the use of the cause-effect diagram.

5. **Develop independent practice.** Teachers develop classroom or homework activities that students have to complete on their own. For example, teachers may ask students to identify the cause-effect sentences in a text, along with the transition words and terms that provided clues that the sentences had a cause-effect relationship.

In the sections that follow, we discuss how the different aspects of the lesson planning process are incorporated into the Language Muse system.

The Language Muse System

As mentioned earlier in this report, it is a fairly common scenario for content-area teachers to have ELs in their classrooms. At the same time, content-area teachers are not necessarily trained to deal with these students from culturally and linguistically diverse backgrounds with potentially varying levels of English proficiency. Further, not all EL students in content-area classrooms are receiving supplemental English language instruction. Many ELs who may still be reading below grade level due to language proficiency issues are mainstreamed into regular classes. These current demographics motivated the development of Language Muse. The motivating idea was to develop a system that would offer a feedback component that highlighted potential sources of linguistic difficulty in the text. This would allow teachers to use the feedback to more easily explore the linguistic features in a text. A teacher might then use the feedback to develop scaffolding to teach students how to handle potentially difficult linguistic features that in a text that could interfere with content comprehension.

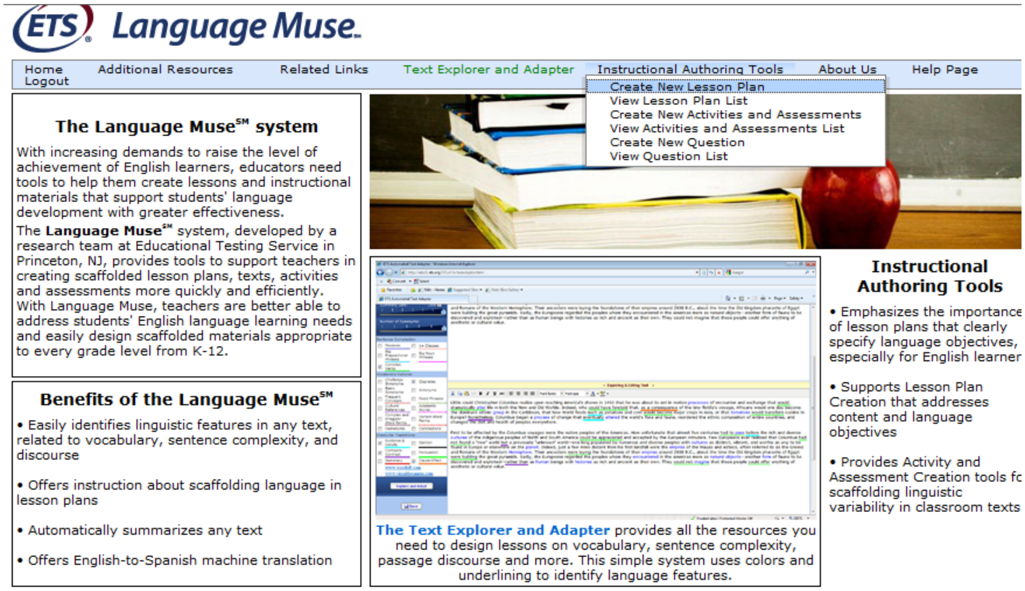
In this section we describe the Language Muse system. There are two main components: (a) the lesson planning component and (b) the Text Explorer and Adapter (TEA-Tool). The lesson planning component is described below to explain how text exploration and modification fits into the lesson planning process (see previous section).Then an in-depth description is provided of the TEA-Tool, which contain the NLP modules that provide the linguistic feedback.

Lesson Planning Components

The lesson planning component has three core modules: (a) Create a New Lesson Plan, (b) Create New Activities and Assessments, and (c) Create New Question. Figure 1 illustrates the drop-down menu of options. As Figure 1 illustrates, there are a number of instructional authoring options that support teachers in lesson plan creation, including Create New Activities and

Assessments, and Create New Question. The application also allows teachers to view lesson plans, activities, assessments, and questions that they have already created.

Teachers can begin with the Create New Lesson Plan page. On that page, there is a template that matches the five commonly used parts of a lesson plan (described in The Lesson Planning Process section). Specifically, these are: (a) standards and objectives, (b) formative and summative assessments, (c) engaging student interest/connecting to student background knowledge, (d) modeling and guided practice, and (e) independent practice. It is here where teachers can enter the critical descriptive information about a lesson plan. In addition, teachers can link a specific text to the lesson plan and invoke the activity and assessment creation capabilities. Activities and assessments created for a specific lesson plan will also be linked to that plan. Questions will be created for specific activities and assessments. These will be linked to the lesson plan through the linked activities and assessments. Teachers can access activities and assessments and the related questions through the lesson plan.[1]



The Language Muse℠ system

With increasing demands to raise the level of achievement of English learners, educators need tools to help them create lessons and instructional materials that support students' language development with greater effectiveness.

The **Language Muse**℠ system, developed by a research team at Educational Testing Service in Princeton, NJ, provides tools to support teachers in creating scaffolded lesson plans, texts, activities and assessments more quickly and efficiently. With Language Muse, teachers are better able to address students' English language learning needs and easily design scaffolded materials appropriate to every grade level from K-12.

Benefits of the Language Muse℠

- Easily identifies linguistic features in any text, related to vocabulary, sentence complexity, and discourse
- Offers instruction about scaffolding language in lesson plans
- Automatically summarizes any text
- Offers English-to-Spanish machine translation

The Text Explorer and Adapter provides all the resources you need to design lessons on vocabulary, sentence complexity, passage discourse and more. This simple system uses colors and underlining to identify language features.

Instructional Authoring Tools

- Emphasizes the importance of lesson plans that clearly specify language objectives, especially for English learner
- Supports Lesson Plan Creation that addresses content and language objectives
- Provides Activity and Assessment Creation tools fc scaffolding linguistic variability in classroom texts

***Figure 1.* Language Muse home page—lesson planning menu options.**

The Text Explorer and Adapter (TEA-Tool)

The TEA-Tool is the linguistic feedback tool in Language Muse.[2] The TEA-Tool allows teachers to explore and edit linguistic features in texts, and summarize (Marcu, 1999) and translate texts from English to Spanish.[3]

TEA-Tool use, features, and NLP methods. As discussed, the TEA-Tool uses NLP methods to provide linguistic feedback, and it supports automated text summarization and machine translation. All of the features have been informed by academic literature (discussed in the Background section) and through discussions with academics and education professionals directly involved in the design and implementation of teacher professional development programs at Stanford University, George Washington University, and Georgia State University (described in more detail later in the paper). In addition, teachers in these programs provided feedback about the application by responding to formal survey questionnaires in the context of our pilot studies which will be discussed later in this article.

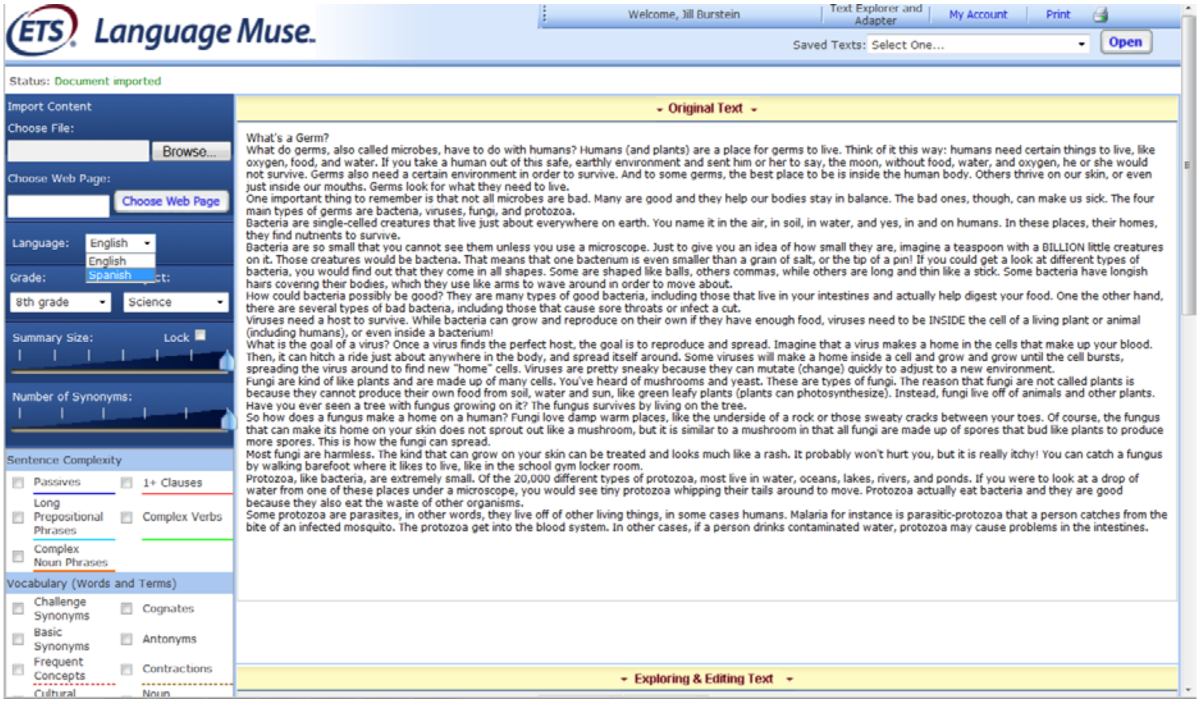
The TEA-Tool feature set has been developed over the past 5 years (Burstein, 2009, Burstein et al., 2007; Burstein et al., in press; Shore et al., 2009). The set of features reflect three linguistic categories: (a) lexical entities (single word and multi-word expressions), (b) syntactic structure, and (c) rhetorical and discourse relations. These are represented as the following categories in the TEA-Tool to be more meaningful for a teacher audience: (a) vocabulary (lexical entities), (b) sentence complexity (syntactic structure), and (c) discourse transitions (rhetorical and discourse relations). Using the category names in the interface, the finer-grained features in these categories available in the TEA-Tool are listed and described in tables in the appendix. Category names were developed for our teacher audience.

TEA-Tool use. Using the TEA-Tool is a relatively simple process. Users click the Text Explorer and Adapter link on the toolbar on the Language Muse homepage (see Figure 1). The TEA-Tool screen opens (see Figure 2). Users then have the option to upload a text (in Microsoft Word, PDF, or plain text format) or choose a web page by clicking the Browse or Choose Web Page button, respectively. As mentioned above, more technical details about system use can be found on the videos also on the Language Muse home page.

Once a text has been uploaded, the user clicks the Explore and Adapt button (see Figure 3). The text is processed, and features selected by the user are highlighted. The screenshot in Figure 3 shows the partial set of linguistic feedback features in the TEA-Tool and also gives an example of how synonyms are highlighted in blue for the user. In Figure 3, the set of synonym choices offered for *germ* are displayed. Details about the synonym capability are described in a later section. Teachers can use these choices to modify the text directly or to create activities that address synonyms to support vocabulary comprehension and vocabulary building.

Teachers can also edit the text as desired. With each iteration, clicking the Explore and Adapt button (see Figure 3) will produce a new set of feedback related to the user's feature selections.

In Figure 2, note that Summary Size slide in the left panel can be moved to the right and to the left to increase and to decrease, respectively, the summary size of the original text. The default setting is 100% of the text. Using the Language drop-down list, also in the left panel in Figure 2, users can change the default English option to Spanish to produce a Spanish translation.



Saved Texts: Select One...

Status: Document imported

Language: English

Grade: 8th grade

Subject: Science

What's a Germ?

What do germs, also called microbes, have to do with humans? Humans (and plants) are a place for germs to live. Think of it this way: humans need certain things to live, like oxygen, food, and water. If you take a human out of this safe, earthly environment and sent him or her to say, the moon, without food, water, and oxygen, he or she would not survive. Germs also need a certain environment in order to survive. And to some germs, the best place to be is inside the human body. Others thrive on our skin, or even just inside our mouths. Germs look for what they need to live.

One important thing to remember is that not all microbes are bad. Many are good and they help our bodies stay in balance. The bad ones, though, can make us sick. The four main types of germs are bacteria, viruses, fungi, and protozoa.

Bacteria are single-celled creatures that live just about everywhere on earth. You name it in the air, in soil, in water, and yes, in and on humans. In these places, their homes, they find nutrients to survive.

Bacteria are so small that you cannot see them unless you use a microscope. Just to give you an idea of how small they are, imagine a teaspoon with a BILLION little creatures on it. Those creatures would be bacteria. That means that one bacterium is even smaller than a grain of salt, or the tip of a pin! If you could get a look at different types of bacteria, you would find out that they come in all shapes. Some are shaped like balls, others commas, while others are long and thin like a stick. Some bacteria have longish hairs covering their bodies, which they use like arms to wave around in order to move about.

How could bacteria possibly be good? They are many types of good bacteria, including those that live in your intestines and actually help digest your food. One the other hand, there are several types of bad bacteria, including those that cause sore throats or infect a cut.

Viruses need a host to survive. While bacteria can grow and reproduce on their own if they have enough food, viruses need to be INSIDE the cell of a living plant or animal (including humans), or even inside a bacterium!

What is the goal of a virus? Once a virus finds the perfect host, the goal is to reproduce and spread. Imagine that a virus makes a home in the cells that make up your blood. Then, it can hitch a ride just about anywhere in the body, and spread itself around. Some viruses will make a home inside a cell and grow and grow until the cell bursts, spreading the virus around to find new "home" cells. Viruses are pretty sneaky because they can mutate (change) quickly to adjust to a new environment.

Fungi are kind of like plants and are made up of many cells. You've heard of mushrooms and yeast. These are types of fungi. The reason that fungi are not called plants is because they cannot produce their own food from soil, water and sun, like green leafy plants (plants can photosynthesize). Instead, fungi live off of animals and other plants. Have you ever seen a tree with fungus growing on it? The fungus survives by living on the tree.

So how does a fungus make a home on a human? Fungi love damp warm places, like the underside of a rock or those sweaty cracks between your toes. Of course, the fungus that can make its home on your skin does not sprout out like a mushroom, but it is similar to a mushroom in that all fungi are made up of spores that bud like plants to produce more spores. This is how the fungi can spread.

Most fungi are harmless. The kind that can grow on your skin can be treated and looks much like a rash. It probably won't hurt you, but it is really itchy! You can catch a fungus by walking barefoot where it likes to live, like in the school gym locker room.

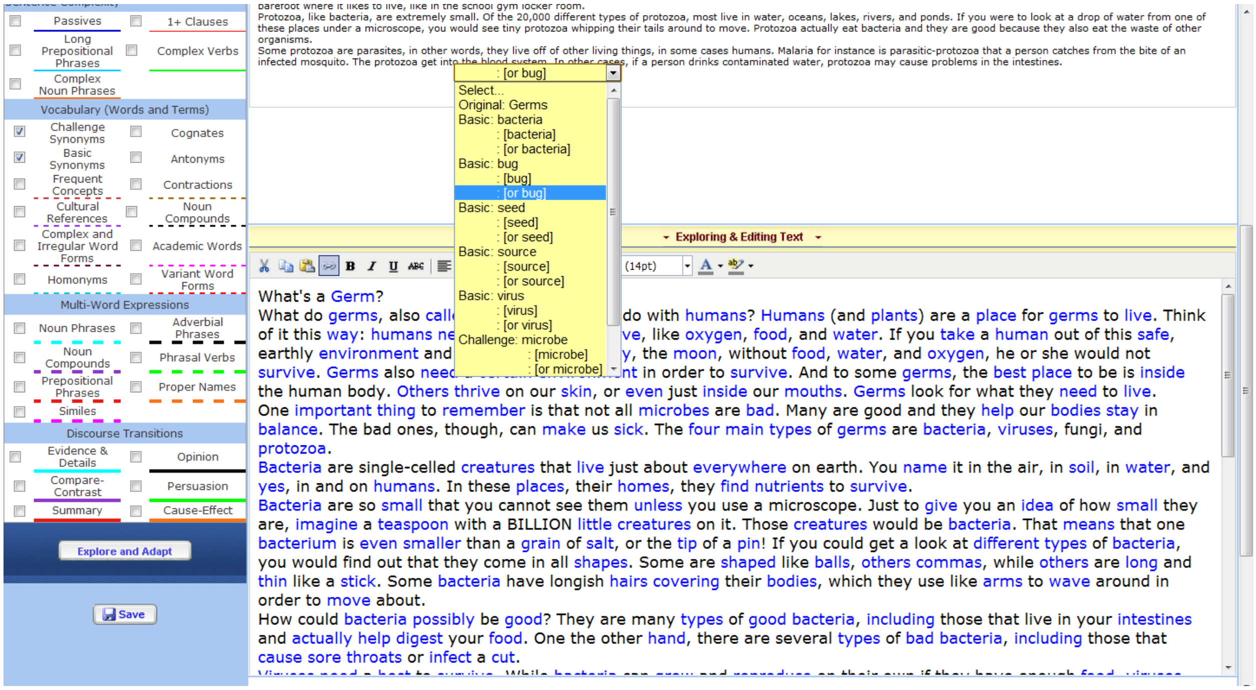
Protozoa, like bacteria, are extremely small. Of the 20,000 different types of protozoa, most live in water, oceans, lakes, rivers, and ponds. If you were to look at a drop of water from one of these places under a microscope, you would see tiny protozoa whipping their tails around to move. Protozoa actually eat bacteria and they are good because they also eat the waste of other organisms.

Some protozoa are parasites, in other words, they live off of other living things, in some cases humans. Malaria for instance is parasitic-protozoa that a person catches from the bite of an infected mosquito. The protozoa get into the blood system. In other cases, if a person drinks contaminated water, protozoa may cause problems in the intestines.

***Figure 2.* TEA-Tool options: Language and Summary Size.**

The TEA-Tool and the lesson planning process. In the context of lesson planning, teachers frequently try to incorporate supportive reading to enhance their pedagogical (teaching) goals. They may have a text identified already or may search for an appropriate reading for the lesson they have created. The Language Muse system is designed to encourage teachers to use the TEA-Tool to explore and, if necessary, modify a text chosen for a lesson. After the text is reviewed, feedback related to linguistic variability in the text can be evaluated, and a teacher can determine what, if any, language modifications, supports, activities, and assessments will be necessary in planning a lesson. Using Language Muse, the teacher can then proceed to develop the lesson plan using the lesson planning modules in the tool to develop the lesson plan description, activities, and assessments. Using the feedback in Figure 3 as an example, a teacher might want to develop a lesson activity that teaches students about synonyms related to key content in the text. Learners might then engage in some form of a class work or homework activity that requires them, for instance, to paraphrase sentences in the text by using synonyms for key content terms.

Teachers can choose from the full set of features using the checkboxes (see Figure 2 and Figure 3 and the appendix for feature set glossary) and can explore a single feature or any combination of features.



□ Passives □ 1+ Clauses
□ Long Prepositional Phrases □ Complex Verbs
□ Complex Noun Phrases

Vocabulary (Words and Terms)

☑ Challenge Synonyms □ Cognates
☑ Basic Synonyms □ Antonyms
□ Frequent Concepts □ Contractions
□ Cultural References □ Noun Compounds
□ Complex and Irregular Word Forms □ Academic Words
□ Homonyms □ Variant Word Forms

Multi-Word Expressions

□ Noun Phrases □ Adverbial Phrases
□ Noun Compounds □ Phrasal Verbs
□ Prepositional Phrases □ Proper Names
□ Similes

Discourse Transitions

□ Evidence & Details □ Opinion
□ Compare-Contrast □ Persuasion
□ Summary □ Cause-Effect

barefoot where it likes to live, like in the school gym locker room.
Protozoa, like bacteria, are extremely small. Of the 20,000 different types of protozoa, most live in water, oceans, lakes, rivers, and ponds. If you were to look at a drop of water from one of these places under a microscope, you would see tiny protozoa whipping their tails around to move. Protozoa actually eat bacteria and they are good because they also eat the waste of other organisms.
Some protozoa are parasites, in other words, they live off of other living things, in some cases humans. Malaria for instance is parasitic-protozoa that a person catches from the bite of an infected mosquito. The protozoa get into the blood system. In other cases, if a person drinks contaminated water, protozoa may cause problems in the intestines.

: [or bug]
Select...
Original: Germs
Basic: bacteria
: [bacteria]
: [or bacteria]
Basic: bug
: [bug]
: [or bug]
Basic: seed
: [seed]
: [or seed]
Basic: source
: [source]
: [or source]
Basic: virus
: [virus]
: [or virus]
Challenge: microbe
: [microbe]
: [or microbe]

What's a Germ?
What do germs, also call … do with humans? Humans (and plants) are a place for germs to live. Think of it this way: humans ne … ve, like oxygen, food, and water. If you take a human out of this safe, earthly environment and … y, the moon, without food, water, and oxygen, he or she would not survive. Germs also need … nt in order to survive. And to some germs, the best place to be is inside the human body. Others thrive on our skin, or even just inside our mouths. Germs look for what they need to live.
One important thing to remember is that not all microbes are bad. Many are good and they help our bodies stay in balance. The bad ones, though, can make us sick. The four main types of germs are bacteria, viruses, fungi, and protozoa.
Bacteria are single-celled creatures that live just about everywhere on earth. You name it in the air, in soil, in water, and yes, in and on humans. In these places, their homes, they find nutrients to survive.
Bacteria are so small that you cannot see them unless you use a microscope. Just to give you an idea of how small they are, imagine a teaspoon with a BILLION little creatures on it. Those creatures would be bacteria. That means that one bacterium is even smaller than a grain of salt, or the tip of a pin! If you could get a look at different types of bacteria, you would find out that they come in all shapes. Some are shaped like balls, others commas, while others are long and thin like a stick. Some bacteria have longish hairs covering their bodies, which they use like arms to wave around in order to move about.
How could bacteria possibly be good? They are many types of good bacteria, including those that live in your intestines and actually help digest your food. One the other hand, there are several types of bad bacteria, including those that cause sore throats or infect a cut.
Viruses need a host to survive. While bacteria can grow and reproduce on their own if they have enough food, viruses

***Figure 3.* TEA-Tool linguistic feedback for synonyms (*germ* ⇒ bacteria, bug, source, virus, microbe).**

TEA-Tool features and NLP. The TEA-Tool uses NLP methods and capabilities for automatic summarization, machine translation, and linguistic feedback. Teachers can use the summarization capability in the TEA-Tool to reduce the amount of text that learners are exposed to, if that method of scaffolding would be effective. Summarization can help to reduce the cognitive load, offering the learner small parts of the text at first and then increasing the amount of text, little by little. The English-to-Spanish machine translation capability can be used with students who have little or no English proficiency, which is sometimes the case. These specific

NLP capabilities are complex, and so for details the reader should refer to Marcu (1999) and SDL (n.d.). While automatic summarization and English-to-Spanish translation of texts can help teachers to develop materials for ELs, the primary focus in this section will be related to lexical, syntactic, and discourse-related feedback because the core goal here is support teachers' awareness of specific linguistic features in texts. The linguistic feedback provided by the tool includes specific information about sentence complexity, vocabulary, and rhetorical and discourse relations. The remainder of this section covers the features in Language Muse's TEA-Tool that use NLP methods to generate linguistic feedback. Evaluations of NLP methods were completed for methods not evaluated in previous research and are described in this section. In a later section, we discuss how feature use may have supported teachers in the development of lesson plan scaffolds designed to address learners' language needs.

Vocabulary. The vocabulary features in the TEA-Tool that use NLP approaches or resources are basic and challenge synonyms, complex and irregular word forms, variant word forms, and multi-word expressions.

Basic Synonyms and Challenge Synonyms: Feature description. As discussed earlier in this paper, many kinds of linguistic features in text may interfere with an EL's comprehension. Unfamiliar vocabulary is recognized as a big contributor. That said, teachers can use synonyms to support basic comprehension or vocabulary development. In the tool, the Basic Synonym and Challenge Synonym features support the comprehension and development aspects, respectively.

The TEA-Tool has a Number of Synonyms slide (see Figure 2), which allows users to adjust the number of words for which the tool will return synonyms. Outputs are based on word frequency. Frequencies are determined using a standard frequency index based on Breland, Jones, and Jenkins (1994).[4] If users want synonyms for a larger number of words across a broader frequency range that includes lower (more rare words) and higher (more common words) frequency words, then they adjust the slide further to the right. If users want to narrow down the number of words to a smaller number of lower frequency (more rare) words, then they move the slide to the left. The more the slider is moved to the left, fewer and more rare words will be addressed. For all words in the text that are within the range of word frequencies at the particular point on the slide, the TEA-Tool returns synonyms (see Figure 3). If users select Basic Synonyms, then the tool returns all words with equivalent or higher frequencies than the word in the text. In theory, these words would be more familiar and more common words that support basic comprehension if the word in the text. If users select Challenge Synonyms, then the tool returns all words with equivalent or lower frequencies than the word in the text. In this case, the teacher might want to work on vocabulary building skills. Words with lower frequencies are more likely to be unfamiliar, and so this might help the learner with new vocabulary. If the user selects both the Basic Synonyms and Challenge Synonyms features, then the tool will output the full list of basic (more familiar), and challenge (less familiar) synonyms for the word in the text. The teacher can, of course, use these synonyms to modify the text directly or to develop classroom or homework activities to support students in learning new words—whether the goal of the lesson is vocabulary comprehension or vocabulary building.

Basic Synonyms and Challenge Synonyms: NLP method. The TEA-Tool uses both a distributional thesaurus (Lin, 1998) and WordNet (Miller, 1995) to generate a comprehensive and reliable set of synonym candidates for words. In a recent annotation study, Burstein and Pedersen (2010) showed that combining a version of Lin's distributional thesaurus and WordNet yielded a higher proportion of automatically generated reliable synonym candidates. The version of the Lin thesaurus being used is a modified distributional thesaurus. Entries in this version were created for educational software and used 300 million words of text from a corpus of fiction, nonfiction, textbooks, and newswire from the *San Jose Mercury News* (Leacock & Chodorow, 2003). It is important to note that in the context of Language Muse system, teachers are looking at synonym candidates not only as substitutes for words, but also as a means of explanation. For instance, for the word *sports*, the candidates, *basketball*, *baseball*, and *football* all exemplify types of *sports*, and might offer helpful explanation to a learner who may be familiar with a particular sport. Therefore, the term *synonym* in this context is used interchangeably with the concept of *word similarity*.

Burstein and Pedersen (2010) examined the reliability of using both resources to implement a synonym identification system with greater breadth in their annotation study. Data preparation proceeded in the following way: Five texts were used to generate synonym candidates for some proportion of words in the text. The texts included two social studies texts, two science texts, and one language arts text. Texts spanned Grades 5, 7, 8, 9, and 12. The number of words per text was, by grade order, 902, 287, 374, 300, and 855. While only five files were in this task, the annotators had to evaluate thousands of synonyms. In preparation of the texts for annotation, synonyms were generated using the modified Lin thesaurus and WordNet using the following procedure: Synonyms were generated for the same set of words from each text. If these words had a standard word frequency *equal to* or *less than* the highest word frequency on the system's Number of Synonyms slide, then synonyms were selected for these words. The idea was to simulate actual system use while providing synonyms for the largest number of words in a text. This frequency also corresponded to the default standard frequency index value in the TEA-Tool (the right end of the Number of Synonyms slide).

Once words were identified as candidates for the synonyms generated, the synonym identification was performed as follows: In Lin's distributional thesaurus (1998), words that are similar to word entries are associated with a probability value that indicates the likelihood of similarity; see Figure 4, which shows an excerpt of an entry for *buy*. In Figure 4, the words each have a probability that indicates a likelihood of the word's similarity to *buy*. The higher the probability value, the more likely the word is likely to be similar to *abandon*. For example, *purchase* is more likely to be related to *buy* than *offer* or *sell*. The mean probability value across the noun, verb, and adjective thesauri (similarity matrices) is approximately 0.172. Therefore, for annotation purposes, similar words equal to or greater than 0.172 were selected as synonym candidates for the word in the text. This threshold was determined to prevent overgeneration of candidates from the Lin resource for the annotators. In the example below, *purchase*, *acquire*, and *own* would therefore be the candidates offered from the Lin-based resource.

buy	
purchase	0.368052
acquire	0.280885
own	0.193306
pay	0.152656
offer	0.147571
import	0.141104

Figure 4. **Excerpt from the Lin distributional thesaurus for the verb *buy*.**

Using WordNet, all words listed for the first three senses associated with each possible part of speech for the text word were returned as synonym candidates in addition to words from the Lin resource. Note that if a sense for a given part of speech provides only the text word itself as synonym, then it is skipped and the next sense is used. As well, WordNet entries were returned only if they were unique in terms of the synonyms returned by the Lin resource. For instance, additional synonyms from WordNet added to the Lin list for *buy* were the following: *bargain, steal, bribe, corrupt, grease one's palms.* Using the default word frequency value (described earlier), 743 words were selected from the five texts as words for which synonyms would be generated. Together, Lin and WordNet resources provided a total of 7,171 candidate synonyms for 743 words in the five texts. The number of synonyms from WordNet was 5,036 (70%), and from the Lin resource, 2,135 (30%). Annotators were given text files with the 743 words and their associated synonyms from Lin and WordNet. There was no indication in the annotator's file as to which resource the synonyms had been derived from. Annotators then had to indicate with an asterisk which synonyms were acceptable substitutions or explanations for the 743 words. The purpose of the task was to use the annotator judgments to determine if either resource was better alone or if in combination they would generate a larger number of synonym candidates. Kappa form interannotator agreement was 0.72 for judgments on synonyms from WordNet, and 0.88 on judgments for synonyms from the Lin resource. These kappa values indicate moderate to strong agreement, respectively. Table 1 indicates that Annotators 1 and 2 agree that 14% of the total of 5,036 synonyms were acceptable (YES). Table 2 shows that both annotators agreed that 28 % of the total of 2,135 Lin synonyms were acceptable (YES).

Table 1

Interannotator Agreement for WordNet Synonyms

Annotator	1—YES	1—NO	Total
2—YES	**14% (702)**	75	15% (777)
2 —NO	340	**78% (3,919)**	4,259
Total	21% (1,042)	3,994	5,036

Note. Numbers in bold represent exact interannotator agreement.

Table 2

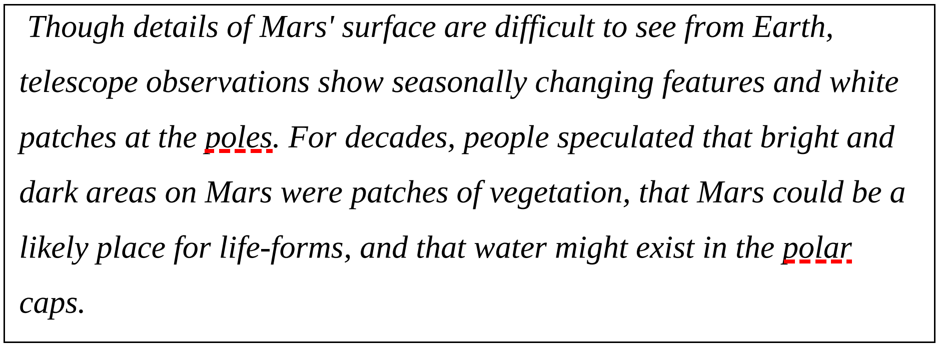
Interannotator Agreement for Lin Synonyms

Annotator	1—YES	1—NO	Total
2—YES	**28% (606)**	37	30% (643)
2—NO	67	**67% (1,425)**	1,492
Total	32% (673)	1,462	2,135

Note. Numbers in bold represent exact interannotator agreement.

The results in both tables indicate that the Lin resource and WordNet each contribute a set of *unique* and *acceptable* synonyms. Therefore, the TEA-Tool uses both resources to generate synonym candidates as shown in Figure 3. While there does appear to still be overgeneration of candidates, we believe this is an acceptable scenario in a setting where a person will examine the full set of outputs. As well, teachers can use the senses of a word that are *not* legitimate substitutes to inspire an activity that may teach learners about polysemy.

Complex and Irregular Word Forms and Variant Word Forms: Feature description. As mentioned earlier in the paper, instructional scaffolding that offers discussion and activities related to morphological structure is an effective method to build ELs' vocabulary (Keiffer & Lesaux, 2007). There are two features in the TEA-Tool that identify words with morphological complexity, specifically, words with prefixes or suffixes: Complex and Irregular Word Forms and Variant Word Forms (see Figure 3). A morphological analyzer is used to generate outputs for both features in the following way: For complex and irregular word forms, the morphological analyzer identifies and underlines words that are morphologically complex. A rollover is available for these words. Users can place their cursors over the highlighted word, and the word stem is shown (e.g., *lost* ⇒ *stem*: *lose*). For the variant word forms, the system underlines words with the same stem that have different parts of speech, such as *poles* and *polar* in Figure 5. Teachers can build instruction related to this kind of morphological variation and teach students about variation and parts of speech.



Though details of Mars' surface are difficult to see from Earth, telescope observations show seasonally changing features and white patches at the poles. For decades, people speculated that bright and dark areas on Mars were patches of vegetation, that Mars could be a likely place for life-forms, and that water might exist in the polar caps.

Figure 5. **Example of variant word forms underlined by the TEA-Tool.**

Complex and Irregular Word Forms and Variant Word Forms: NLP method. The morphological analyzer used in Language Muse was originally developed for *c-rater*™, ETS's short-answer scoring system (Leacock & Chodorow, 2003). This analyzer handles derivational and inflectional morphology. Derivational morphology includes cases where affixes can change the part of speech of a word, such as in nominalization of a verb (e.g., *buy* to *buy+er*). Inflectional morphology, on the other hand, adds grammatical markers that, for instance, change singular nouns to plural nouns (e.g., *cat* to *cat+s*) and present tense verbs to past tense verbs (e.g., *observe* to *observ+ed*). We completed an evaluation to determine the accuracy of the morphological analyzer in the Language Muse context. The evaluation was completed as follows: A set of 72 texts from fifth- though 12th-grade from social studies, science, and language arts were used. From these 72 texts, 1,000 sentences were randomly selected. The morphological analyzer was run on the 1,000 sentences, and the system identified words that were morphologically complex. Two annotators were given a file with the 1,000 sentences and the words identified as morphologically complex from each sentence. Annotators were asked to identify any words that were misidentified as morphologically complex, and to indicate words in each sentence that were morphologically complex but were missed by the morphological analyzer. We then computed agreement between each of the annotator's judgments and system judgments using precision, recall, and F-measure metrics. Definitions of *precision*, *recall,* and *F-measure* are as follows (where MC = morphologically complex):

- $Precision = \frac{|\{annotator\ MC\ words\} \cap \{system\ MC\ words\}|}{|\{\ system\ MC\ words\}|}$
- $Recall = \frac{|\{annotator\ MC\ words\} \cap \{system\ MC\ words\}|}{|\{annotator\ MC\ words\}|}$
- $F\text{-}measure = \frac{2\times\ (Precision \times Recall)}{(Precision + Recall)}$

The total number of morphologically complex words selected by annotators and the system are used to compute precision, recall, and F-measure metrics. The results appear in Table 3.

Table 3

Precision, Recall, and F-Measures for Two Annotators and the Morphological Analyzer

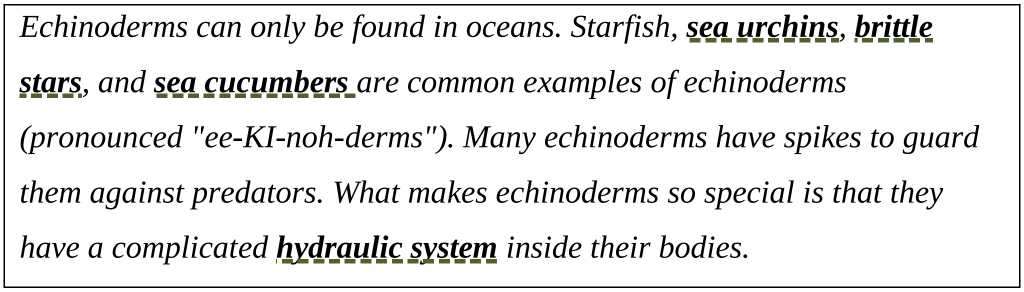
Annotator	Precision	Recall	F-measure
1	0.95 (3620/3810)	0.91 (3620/3970)	0.93
2	0.97 (3713/3810)	0.94 (3713/3933)	0.95

Results would indicate that the analyzer had a high degree of accuracy based on this annotation task.

Multi-Word Expressions: Feature description. Multiple-word expressions include structures, such as idioms (e.g., body and soul), phrasal verbs (e.g., reach into), and multiword expressions that are not necessarily idiomatic but typically appear together to express a single concept (e.g., heart disease). All of these kinds of collocations may be unfamiliar terms to ELs, and so they may interfere with comprehension of content in a text. The Multi-Word Expressions feature in the TEA-Tool is designed to identify and underline the different types of these terms. Teachers can then use this information to scaffold the text appropriately.

Multi-Word Expressions: NLP method. Two resources are used to identify collocations in texts in the context of the Multi-Word Expressions features. First, we use the WordNet 3.0 compounds list of approximately 65,000 collocational terms. Terms can be composed of two to four words (e.g., *natural language, natural language processing, natural language processing application*). We also use a collocation tool that was designed to identify collocations in test-taker essays (Futagi, Deane, Chodorow, & Tetreault, 2008). Details about how this collocation detection system works and complete evaluations can be found in Futagi et al. (2008). This tool is currently used in *e-rater*®, ETS's essay scoring system (Attali & Burstein, 2006). Futagi et al.'s collocation tool essentially identifies collocations in a text that occur in seven syntactic structures that are the most common structures for collocations in English based on *The BBI Combinatory Dictionary of English* (Benson, Benson, & Ilson, 1997). For instance, the following examples are given in Futagi et al.: Noun *of* Noun (e.g., swarm of bees), and Adjective + Noun (e.g., strong tea), and Noun + Noun (e.g., house arrest).

The collocation tool uses a reference database containing collocations that have been created from the Google N-gram Corpus,[5] which is one terabyte. However, the majority of the data turn out to be almost entirely nonword strings, which are unusable for collocation reference. Therefore, the data have been filtered to keep only the usable strings, and the final size of the corpus is about one third of the original (approximately one billion n-grams retained). The tool identifies bigram, trigrams, and 4-grams in text and computes point-wise mutual information values between these n-grams extracted from the text and collocations in a reference database. For the purpose of identifying collocations for the Multi-Word Expressions feature in Language Muse, we do the following: The list of WordNet compounds is matched against n-gram sequences in the text. Any matches are considered possible outputs for the Multi-Word Expressions feature. In addition, the collocation tool is also used to extract n-grams from the text, which are then matched against the reference database of collocations, and point-wise mutual information (PMI) values are computed between the n-gram sequences found in the text and collocations found in the reference database. Thresholds were determined using the point-wise mutual information values to prevent overgeneration of collocations that might not be useful. . For instance, some collocations in the text with low PMI values may just be noncollocational bigrams, such as *decorate walls*, whereas others with higher PMI values, such as *good tidings*, do qualify as acceptable collocations. Once the matches with the WordNet compounds have been identified and the collocations identified by the collocation tool have been found, the nonoverlapping collocations found by each resource are then used by the TEA-Tool to highlight fixed phrases in the text, as in Figure 6.



*Echinoderms can only be found in oceans. Starfish, **sea urchins**, **brittle stars**, and **sea cucumbers** are common examples of echinoderms (pronounced "ee-KI-noh-derms"). Many echinoderms have spikes to guard them against predators. What makes echinoderms so special is that they have a complicated **hydraulic system** inside their bodies.*

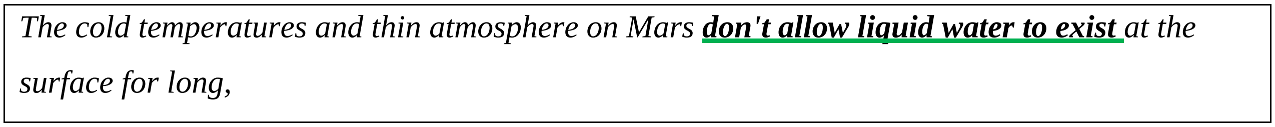
***Figure 6.* Example of nonoverlapping collocations highlighted by the TEA-Tool through the Multi-Word Expressions feature.**

Collocations, such as hydraulic system in the example above, may be unfamiliar to ELs. Teachers may want to offer additional explanation or activities concerning this term and other collocations to teach ELs about this type of structure in English.

Sentence Complexity: Feature description. Complex phrasal or sentential structures can introduce potential difficulty in a text. The following Sentence Complexity features can be selected in the TEA-Tool: Long Prepositional Phrases, which identifies sequences of two or more consecutive prepositional phrases (e.g., “He moved the dishes from the table to the sink in the kitchen”); Complex Noun Phrases, which shows noun compounds composed of two or more nouns (e.g., emergency management agency) or noun phrases with hyphenated modifiers (e.g., shark-infested waters); Passives, which indicates passive sentence constructions (e.g., The book was bought by the boy.); 1+Clauses, which points out sentences with at least one dependent clause (e.g., The newspaper noted that there have been no recent weather advisories.); and Complex Verbs, which identifies verbs with multiple verbal constituents (e.g., would have gone, will be leaving, had not eaten).

Sentence Complexity: NLP method. Rule-based NLP is used to identify all of the Sentence Complexity features in the TEA-Tool: Long Prepositional Phrases, Complex Noun Phrases, Passives, 1+Clauses, and Complex Verbs. Using a shallow parser developed for e-rater (Burstein, Kukich, Wolff, Lu, Chodorow, Braden-Harder, et al., 1998), rules were developed to identify the features described above. The parser had been evaluated for prepositional phrase and noun phrase detection in the context of c-rater (Leacock & Chodorow, 2003). The module to identify passive sentence construction had been previously developed and evaluated for use with *Criterion*®, ETS's online essay evaluation service (Burstein, Chodorow, & Leacock, 2004), and sentences structures identified by the *1+ Clauses* option had been evaluated in earlier versions of e-rater (Burstein, Kukich, Wolff, Lu, Chodorow, Braden-Harder, et al., 1998).

We did complete an evaluation of complex verbs, as the set of rules was fairly complex and was designed to handle complexity, such as in Figure 7.



The cold temperatures and thin atmosphere on Mars ***don't allow liquid water to exist*** *at the surface for long,*

Figure 7. **Example of a complex verb phrase underlined by the TEA-Tool.**

To examine the accuracy of the complex verb identification module, an annotation task similar to that completed for the evaluation of the morphological analyzer was completed. Two annotators were given a set of 1,035 sentences that had been randomly selected from the set of 72 texts described earlier in the section about morphologically complex words. For each of the sentences, the complex verbs identified in the sentence were displayed. Two annotators were asked to indicate if any of the complex verbs were incorrect and also to indicate if any were missed. We then computed agreement between each of the annotator's judgments and system judgments using *precision*, *recall*, and F-measure metrics. The results appear in Table 4.The total number of complex verbs selected by annotators and the system are used to compute precision, recall, and F-measure metrics. Definitions of precision, recall, and F-measures are as follows (where CV = complex verb):

- *Precision* = $\frac{|\{annotator\ CV\ words\} \cap \{system\ CV\ words\}|}{|\{\ system\ CV\ words\}|}$

- *Recall* = $\frac{|\{annotator\ CV\ words\} \cap \{system\ CV\ words\}|}{|\{annotator\ CV\ words\}|}$

- *F-measure* = $\frac{2\times\ (Precision \times Recall)}{(Precision + Recall)}$

Table 4

Precision, Recall, and F-Measures for Two Annotators and the Complex Verb Detection Module

Annotator	Precision	Recall	F-measure
1	0.89 (183/205)	0.75 (183/264)	0.81
2	0.89 (184/205)	0.58 (182/310)	0.71

Annotator 2 did seem to find additional complex verbs that were missed by the module. However, overall results would indicate that the complex verb detection module had a reasonably high degree of precision based on this annotation task. It is desirable in the application to have a trade-off between precision and recall, where precision is higher. It is preferable for the system to generate a smaller proportion, but to generate these proportions correctly.

Discourse Transitions: Feature description. Discourse-relevant cue words and terms are highlighted when the following Discourse Transitions features are selected in the TEA-Tool: Evidence & Details, Compare-Contrast, Summary, Opinion, Persuasion, and Cause-Effect.

Discourse Transitions: NLP method. The Discourse Transition features in the TEA-Tool are outputs from a discourse analyzer from an earlier version of e-rater. Essentially, the system identifies cue words and phrases in text that are being used as specific discourse cues. For instance, the term *because* is typically associated with a cause-effect relation. However, some words need to appear in a specific syntactic construction to function as a discourse term. For instance, the word *first* functions as an adjective modifier and not a discourse term in a phrase, such as “the first piece of cake.” When *first* is sentence-initial, as in, “First, she sliced a piece of cake,” then it is used as a discourse marker to indicate a sequence of events. Only in the latter case would the system identify *first* as a discourse marker. For system details and relevant evaluations, see Burstein, Kukich, Wolff, Lu, Chodorow, Braden-Harder, et al. (1998) and Burstein, Kukich, Wolff, Lu, and Chodorow (1998).

Pilot Study

In this section, we describe the pilot research and preliminary findings that suggest that exposure to Language Muse's linguistic feedback can support teachers in the development of lesson plan scaffolds designed to address learners' language needs.

Site Descriptions

As part of a 3-year grant from the Institute of Education Sciences, U.S. Department of Education, the Language Muse system has been integrated and delivered as part of Stanford University and George Washington University's (GWU) EL in-service teacher preparation courses. A third site at Georgia State University (GSU) has also been included in the set of pilot sites.

Stanford's courses are offered entirely online to teachers as part of a professional development program that awards the California State Cross-Cultural Language and Academic Development (CLAD) certificate through itsCalifornia Teachers of English Learners (CTEL) certification process. By state law, all California teachers of ELs must obtain a CLAD/CTEL or equivalent certification. GWU offers a hybrid (online/face-to-face) course series that takes place in teachers' school settings. In-class professional development is augmented in online forums for purposes of discussion, coursework submission, and materials dissemination. Courses are offered as a part of the requirements that lead toward a certificate or master's degree in bilingual or bilingual special education. GSU's Urban Accelerated Certification and Master's Program is a 2-year alternative certification program at the master's degree level for those with an undergraduate degree or higher in an area other than education who are interested in teaching in urban elementary schools. Students in this program spend the first year as full-time students taking courses and completing field experiences in schools in the metro Atlanta area. This program qualifies candidates who have successfully met all requirements to be recommended to the state for certification in early childhood education (pre-K through 5th grade) with an English to speakers of other languages endorsement from the state. At all three sites a common goal is to prepare educators to meet the needs of culturally and linguistically diverse students.

Teacher Cohorts

During the course of the 3-year grant, Language Muse is being piloted at the three sites with new teacher cohorts as each new course sequence begins. We describe and report on the initial three cohorts for which we have complete data sets and for which analyses are underway.

These initial cohorts contain 69 teachers: 28 from Stanford, 19 from GWU, and 22 from GSU. All teachers from the GSU site were preservice teachers. This means that teachers in this program were learning how to be teachers and did not currently hold teaching positions. Some had student teaching experience. The teachers in the Stanford and GWU cohorts held teaching positions in elementary, middle, and high schools. Teachers had a range of teaching experience from less than a year of teaching experience to as much as 37 years of teaching experience. Teachers taught in a range of content areas, including social studies, science, math, language arts, music, art, computers, physical education, and health.

Language Muse Intervention

As stated earlier, the motivation for the Language Muse system was to offer instruction about linguistic structures to teachers, so that they become keenly aware of linguistic elements that may interfere with learners' content comprehension of a text. Equipped with this knowledge, teachers can develop appropriate instructional scaffolding. Consistent with this, one of the main hypotheses of this research is that as teachers become more aware of linguistic difficulty in text, they can develop teaching materials that offer instructional scaffolding that supports learners' language needs. The TEA-Tool, Language Muse's linguistic feedback component, offers feedback to support teachers in developing awareness about potentially difficult linguistic features in classroom texts.

All three sites agreed to integrate Language Muse into their coursework as an intervention to support coursework instruction and goals. The following activities were integrated into the courses at each site, and teachers completed each activity as part of the pilot intervention:

1. Background survey to collect information about teachers' professional background
2. Pre- and posttests that evaluated teachers on the following: (a) knowledge of linguistic structures (e.g., morphologically complex words, complex verb phrases) and (b) ability to identify linguistic features in a text that were likely to interfere with

content comprehension and knowledge about how to build instructional scaffolding for these features

3. One assigned reading that discussed linguistic features that were potentially difficult for ELs, how Language Muse could be used to explore these features, and how instructional scaffolding could be developed for these features
4. Language Muse self-guided instruction and practice
5. Three videos demonstrating how to use Language Muse
6. Up to two practice activities requiring students to use the system
7. Lesson plan assignment in which teachers developed a lesson plan using the tool. The lesson plan assignment required that they used the TEA-Tool to explore at least one text and that a lesson plan be designed that included instructional scaffolding for that text
8. Perception survey to collect teachers' perceptions of Language Muse

Preliminary Findings

For this paper, our goal was two-fold: (a) to show that the NLP-driven and other linguistic feedback from the TEA-tool could support teachers in their ability to identify linguistically difficult features in text and (b) to evaluate if the feedback supports teachers in the development of relevant and potentially effective instructional scaffolding that supports learners' language needs. In light of these goals, we conducted (a) an evaluation of the relationship between lesson plan scores and (b) a qualitative analysis of teachers' inclusion of TEA-Tool linguistic feedback in developing instructional scaffolding for the lesson plan assignment.

Lesson plan assignment. The lesson plan assignment instructed teachers to use Language Muse to build a lesson plan for a target student population. Teacher cohorts at the different sites were given slightly different instructions about the target learner population depending on the goals of the assignment at each site. However, all teachers had to use Language Muse when producing a lesson plan in the following ways:

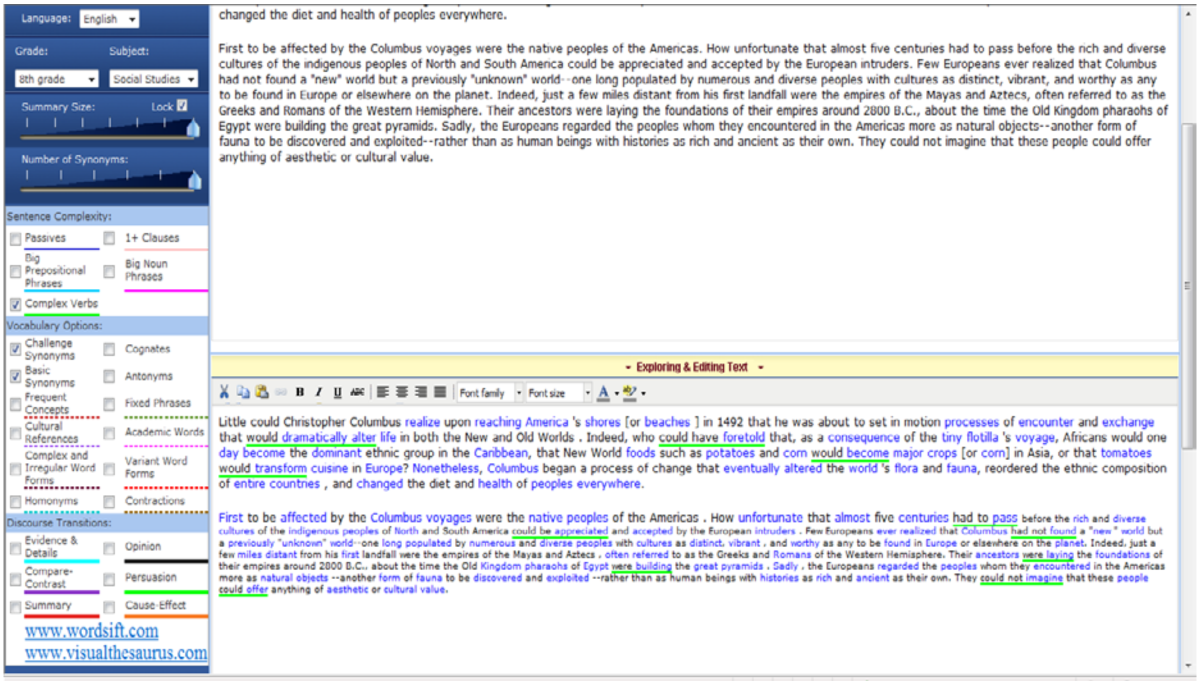
- Processing or exploring at least one text using the TEA-Tool. Teachers could select the TEA-Tool features of their choice to explore any lexical, syntactic, and discourse features in the text.

- Creating one lesson plan using Language Muse's lesson planning (instructional authoring) components; the lesson plan needed to include (a) a completed lesson plan template describing all of the elements of the plan and (b) at least two instructional scaffolds in the form of activities and assessments.

Lesson plan scoring. Of the 69 participating teachers, 52 teachers used the tool to create a lesson plan as part of their coursework. The set of 52 plans were downloaded from the tool, and printed for scoring purposes. Each lesson plan was assigned two scores by two human raters, both of whom work in education. One rater has teaching experience and the other rater works in literacy research. Raters were trained to assign two scores to each of the plans: (a) the *language skills evaluation score* was based on how well the plan addressed language and language skill objectives in the lesson in general, and (b) the *English language-specific evaluation score* was based on how well the plan addressed potential areas of linguistic or cultural complexity in the lesson that might present unique challenges to ELs. A Pearson correlation was used to compute interrater agreement. Correlations were 0.72 and 0.74 for the language skills evaluation score and the English language-specific evaluation score, respectively.

The score scale for each of the two scores was 0 through 2, where 0 indicated the lowest quality score, and 2 was the highest quality score. The two rater scores were averaged to compute the final score for each of the two scores, the language skills evaluation score, and the English language-specific evaluation score. The two final scores were used in the statistical evaluations described in a later section.

Qualitative analysis coding. In Language Muse, users specify which texts are associated with the lesson plan. These saved texts are created and saved in the TEA-Tool and are easily accessible. When a saved text is opened from the TEA-Tool, the features selected by the user to explore the text are shown (see Figure 8). Saved texts, along with user feature selections, are stored in the system's database. For the set of 52 lesson plans, one of the authors manually reviewed each of the lesson plans along with the saved text(s) that the teacher had explicitly associated with the lesson plan. The author used a coding scheme of 0, 1, or 2. These codes were independent of the lesson plan scores. These codes indicated the following: 0 indicated that the lesson plan did not include instructional scaffolding based on a TEA-Tool feature, 1 indicated that the lesson plan included one activity or assessment that was based on a single TEA-Tool feedback feature (e.g., the teacher selected the Challenge Synonyms and Basic Synonyms features and developed an activity related to synonyms), and 2 indicated that the lesson plan included two or more activities or assessments that were based on two or more TEA-Tool feedback features (e.g., the teacher selected the Complex Verbs feature and created an activity related to complex verb structure, and the teacher selected the Compare-Contrast feature and created a related activity).



Language: English

Grade: 8th grade

Subject: Social Studies

Summary Size: Lock ☑

Number of Synonyms:

Sentence Complexity:

☐ Passives
☐ 1+ Clauses
☐ Big Prepositional Phrases
☐ Big Noun Phrases
☑ Complex Verbs

Vocabulary Options:

☑ Challenge Synonyms
☐ Cognates
☑ Basic Synonyms
☐ Antonyms
☐ Frequent Concepts
☐ Fixed Phrases
☐ Cultural References
☐ Academic Words
☐ Complex and Irregular Word Forms
☐ Variant Word Forms
☐ Homonyms
☐ Contractions

Discourse Transitions:

☐ Evidence & Details
☐ Opinion
☐ Compare-Contrast
☐ Persuasion
☐ Summary
☐ Cause-Effect

www.wordsift.com
www.visualthesaurus.com

changed the diet and health of peoples everywhere.

First to be affected by the Columbus voyages were the native peoples of the Americas. How unfortunate that almost five centuries had to pass before the rich and diverse cultures of the indigenous peoples of North and South America could be appreciated and accepted by the European intruders. Few Europeans ever realized that Columbus had not found a "new" world but a previously "unknown" world--one long populated by numerous and diverse peoples with cultures as distinct, vibrant, and worthy as any to be found in Europe or elsewhere on the planet. Indeed, just a few miles distant from his first landfall were the empires of the Mayas and Aztecs, often referred to as the Greeks and Romans of the Western Hemisphere. Their ancestors were laying the foundations of their empires around 2800 B.C., about the time the Old Kingdom pharaohs of Egypt were building the great pyramids. Sadly, the Europeans regarded the peoples whom they encountered in the Americas more as natural objects--another form of fauna to be discovered and exploited--rather than as human beings with histories as rich and ancient as their own. They could not imagine that these people could offer anything of aesthetic or cultural value.

Little could Christopher Columbus realize upon reaching America 's shores [or beaches] in 1492 that he was about to set in motion processes of encounter and exchange that would dramatically alter life in both the New and Old Worlds . Indeed, who could have foretold that, as a consequence of the tiny flotilla 's voyage, Africans would one day become the dominant ethnic group in the Caribbean, that New World foods such as potatoes and corn would become major crops [or corn] in Asia, or that tomatoes would transform cuisine in Europe? Nonetheless, Columbus began a process of change that eventually altered the world 's flora and fauna, reordered the ethnic composition of entire countries , and changed the diet and health of peoples everywhere.

First to be affected by the Columbus voyages were the native peoples of the Americas . How unfortunate that almost five centuries had to pass before the rich and diverse cultures of the indigenous peoples of North and South America could be appreciated and accepted by the European intruders . Few Europeans ever realized that Columbus had not found a "new " world but a previously "unknown" world--one long populated by numerous and diverse peoples with cultures as distinct, vibrant , and worthy as any to be found in Europe or elsewhere on the planet. Indeed, just a few miles distant from his first landfall were the empires of the Mayas and Aztecs , often referred to as the Greeks and Romans of the Western Hemisphere. Their ancestors were laying the foundations of their empires around 2800 B.C., about the time the Old Kingdom pharaohs of Egypt were building the great pyramids . Sadly , the Europeans regarded the peoples whom they encountered in the Americas more as natural objects --another form of fauna to be discovered and exploited --rather than as human beings with histories as rich and ancient as their own. They could not imagine that these people could offer anything of aesthetic or cultural value.

***Figure 8.* Saved text with Complex Verbs, Challenge Synonyms, and Basic Synonyms features selected.**

Preliminary findings. A simple linear regression was run to evaluate if the lesson plan scores could be predicted based on TEA-Tool use. Specifically, we wanted to know if there was a relationship between the lesson plan score and the qualitative analysis score that told us the extent to which the teacher had used TEA-Tool features to develop instructional scaffolding in the lesson plan assignment.

The regression showed that for the language skills evaluation score there was a marginal positive relationship between the two. The correlation was 0.27 with a *p*-value of .052. This positive trend suggests that teachers who used TEA-Tool feedback to create instructional scaffolding in their lesson plans received a higher score for the language skills evaluation score.

This suggests that in its current form, the tool has promise for developing teachers' linguistic awareness, which they can then use to develop effective instructional scaffolding for difficult linguistic structures in text. While a positive correlation (0.13) was found for the English language-specific evaluation score, this correlation was not significant. What this is most likely telling us is that the current intervention may need to include more instruction related to the specific language needs of ELs. Also, with regard to the tool, existing features, such as Cultural References (e.g., name of plants, insects, animals, and foods), and Multi-Word Expressions (i.e., detection of collocations) that are more EL-specific may need to be enhanced, and new features may need to be added. More advanced features that identify figurative language could be important enhancements.

Discussion and Conclusions

As discussed earlier in this paper, students acquire much of their knowledge and understanding of the different subject areas through reading, and out-of-class reading often forms the basis for class discussions or homework. The is a growing emphasis on text-based learning in curriculum standards as proposed by the Common Core State Standards, which is becoming influential in the development of curriculum standards in the United States. When classroom texts contain linguistically unfamiliar structures, such as words above a learner's reading level, idioms, complex phrases or sentences, and unfamiliar or complex discourse relations, this may result in gaps in explanation about key content. At the same time, the number of teachers trained in effective instructional strategies to meet the range of needs of ELs has not necessarily increased consistently with the rate of the EL population.

The motivation for the development of the Language Muse system was to offer instruction for teachers that was aligned with the familiar process of lesson plan development, and, as part of this approach, to also offer automated linguistic feedback. The linguistic feedback would support the development of teachers' linguistic awareness. As teachers created lesson plans, the feedback would guide them in the identification of linguistic elements in texts that may be unfamiliar to learners and may interfere with learners' content comprehension. Using their knowledge about potentially difficult linguistic forms, teachers would be able to develop appropriate instructional scaffolding to serve learner language needs in the context of lesson planning. This perspective was supported by the partner teacher professional development programs at Stanford University, George Washington University, and Georgia State University.

As part of a pilot study funded by the IES, the Language Muse system has been integrated into three teacher professional development programs. These programs share common goals, including (a) to provide instruction to teachers about linguistic structures, and in particular, those structures that might interfere with learner comprehension of content, and (b) to provide instruction for teachers about how to design effective language scaffolding to support ELs' comprehension and language skills. In the context of the pilot integration, teachers are using Language Muse to develop lesson plans as part of the coursework. Preliminary findings are promising in this context and suggest that the more that teachers use linguistic feedback from the system, the more likely it is that they will produce a lesson plan that contains relevant language scaffolding. What we have also learned from these analyses is that it would be helpful to enhance the current set of linguistic features to produce feedback that was more fine-tuned to the specific needs of ELs, such as components that more reliably recognized figurative language. Additional instruction related to how to use the current set of features more effectively to develop scaffolding that was specific to the needs of ELs might also be effective support.

In planned future research in the context of Language Muse pilot studies, we will work with post-intervention teachers to deliver lesson plans developed with the system to ELs in the teacher classrooms. The post-intervention teachers will be a group of teachers who have finished their participation in a Language Muse pilot study at one of the three partner sites. These teachers will have agreed to use the system in their classrooms. Through this research, we will evaluate the effectiveness of instructional language scaffolding developed using linguistic feedback from the Language Muse system.

References

Attali, Y., & Burstein, J. (2006). Automated essay scoring with e-rater® v.2.0. *Journal of Technology, Learning, and Assessment, 4*(3), 1–31.

August, D. (2003). *Supporting the development of English literacy in English language learners: Key issues and promising practices* (Report No. 61). Baltimore, MD: Johns Hopkins University, Center for Research on the Education of Students Placed at Risk. Retrieved from http://www.cde.state.co.us/cdesped/download/pdf/ELL_SupportDevelopEngLangLit.pdf

Bean, T. W. (1982). Second language learners' comprehension of an ESL prose selection. *Journal of the Linguistic Association of the Southwest, 4,* 376–386.

Benson, M., Benson, E., & Ilson, R. (Eds.). (1997). *The BBI combinatory dictionary of English: A guide to word combinations* (revised). Amsterdam, the Netherlands: John Benjamins.

Biancarosa, G., & Snow, C. (2004). *Reading next: A vision for action and research in middle and high school literacy*. New York, NY: Carnegie Corporation of New York and Alliance for Excellent Education.

Breland, H., Jones, R., & Jenkins, L (1994). *The College Board vocabulary study.* (College Board Report No. 94-4; ETS Research Report No. RR-94-26). New York, NY: College Entrance Examination Board.

Burstein, J. (2009). Opportunities for natural language processing in education. In A. Gelbulkh (Ed.), *Lecture notes in computer science: Vol. 5449. Computational linguistics and intelligent text processing.* (pp. 6–27). Berlin, Germany: Springer-Verlag.

Burstein, J., Chodorow, M., & Leacock, C. (2004). Automated essay evaluation: The Criterion Online service. *AI Magazine, 25*(3), 27–36.

Burstein, J., Kukich, K., Wolff, S., Lu, C., & Chodorow, M. (1998). Enriching automated scoring using discourse marking. In M. Stede, L. Wanner, & E. Hoy (Eds.), *Proceedings of the Workshop on Discourse Relations and Discourse Marking* (pp. 15–21). New Brunswick, NJ: Association of Computational Linguistics.

Burstein, J., Kukich, K., Wolff, S., Lu, C., Chodorow, M., Braden-Harder, L., & Harris, M. D. (1998). Automated scoring using a hybrid feature identification technique. In *Proceedings of the 36th Annual Meeting of the Association of Computational Linguistics* (Vol. 1, pp. 206–210). New Brunswick, NJ: Association of Computational Linguistics.

Burstein, J., & Pedersen, T. (2010). *Towards improving synonym options in a text modification application* (University of Minnesota Supercomputing Institute Research Report Series UMSI 2010/165). Retrieved from http://static.msi.umn.edu/rreports/2010/165.pdf

Burstein, J., Sabatini, J., & Shore, J. (in press). Developing NLP applications for educational problem spaces. In R. Mitkov (Ed.), *Oxford handbook of computational linguistics*. New York, NY: Oxford University Press.

Burstein, J., Shore, J., Sabatini, J., Lee, Y.-W., & Ventura, M. (2007). Developing a text support tool for English-language learners. In R. Luckin, K. R. Koedinger, & J. E. Greer (Eds.), *Artificial intelligence in education: Building technology rich learning contexts that work* (pp. 542-544). Amsterdam, The Netherlands: IOS Press.

Calderón, M., & Minaya-Rowe, L. (2007). *ESL—How ELLs keep pace with mainstream students.* Thousand Oaks, CA: Corwin Press.

Carlo, M. S., August, D., McLaughlin, B., Snow, C. E., Dressler, C., Lippman, D.…White, C. (2004). Closing the gap: Addressing the vocabulary needs of English language learners in bilingual and mainstream classrooms. *Reading Research Quarterly, 39*(2), 188–215.

Center for Public Education. (2007). *Research review: What research says about preparing English language learners for academic success.* Alexandria, VA: Author.

Chall, J. S., & Dale, E. (1995). *Readability revisited: The new Dale-Chall readability formula*. Cambridge, MA: Brookline Books.

Coleman, D., & Pimentel, S. (2011a). *Considerations for kindergarten through second grade curriculum materials to achieve alignment with the Common Core State Standards*. Retrieved from http://schools.nyc.gov/NR/rdonlyres/93D7B95D-A17F-4EC9-A4EE-A26CA7CCA0BC/0/PublishersCriteriaforLiteracyforK2Final.pdf

Coleman, D., & Pimentel, S. (2011b). *Publisher's criteria for the Common Core State Standards in ELA & literacy, grades 3-12*. Retrieved from http://www.isbe.net/common_core/pdf/pub_criteria_ela3-12.pdf

Coxhead, A. (2000). *The academic word list.* Retrieved from http://www.victoria.ac.nz/lals/resources/academicwordlist/

Fitzgerald, J. (1995). English-as-a-second-language learners' cognitive reading processes: A review of research in the United States. *Review of Educational Research*, *65*(2), 145–190.

Flesch, R. (1948). A new readability yardstick. *Journal of Applied Psychology, 32*, 221–233.

Francis, D., August, D., Goldenberg, C., & Shanahan, T. (2004). *Developing literacy skills in English language learners: Key issues and promising practices.* Retrieved from www.cal.org/natl-lit-panel/reports/Executive_Summary.pdf

Francis, D., Rivera, M., Lesaux, N., Keiffer, M., & Rivera, H. (2006). *Practical guidelines for the education of English language learners: Research based recommendations for instruction and academic interventions.* Portsmouth, NH: Center on Instruction. Retrieved from www.centeroninstruction.org/files/ELL1-Interventions.pdf

Futagi, Y., Deane, P., Chodorow, M., & Tetreault, J. (2008). A computational approach to detecting collocation errors in the writing of non-native speakers of English. *Computer Assisted Language Learning, 21*, 353–367.

Gándara, P., Maxwell-Jolly, J., & Driscoll, A. (2005). *Listening to teachers of English language learners: A survey of California teachers' challenges, experiences, and professional development needs.* Sacramento, CA: The Regents of the University of California. Retrieved from http://www.cftl.org/centerviews/july05.html

Green, L. C., Foote, M., Walker, C., & Shuman, C. (2010). From questions to answers: Education faculty members learn about English language learners. *College Reading Association Yearbook, 31,* 113–126.

Hancin-Bhatt, B., & Nagy, W. E. (1994). Lexical transfer and second language morphological development. *Applied Psycholinguistics, 15*(3), 289–310.

Ihnot, C. (1997). *Read naturally.* St. Paul, MN: Read Naturally.

James, C., & Klein, K. (1994). Foreign language learners' spelling and proofreading strategies. *Papers and Studies in Contrastive Linguistics, 29,* 31–46.

Jiménez, R. T., Garcia, G. E., & Pearson, D. P. (1996). The reading strategies of bilingual Latina/o who are successful English readers: Opportunities and obstacles. *Reading Research Quarterly, 31*(1), 90–112.

Kieffer, M. J., & Lesaux, N. K. (2007). Breaking down words to build meaning: Morphology, vocabulary, and reading comprehension in the urban classroom. *The Reading Teacher, 61*, 134–144.

Koga, N., & Hall, T. (2004). *Curriculum modification.* Wakefield, MA: National Center on Accessing the General Curriculum. Retrieved from http://aim.cast.org/learn/historyarchive/backgroundpapers/curriculum_modification

Leacock, C., & Chodorow, M. (2003). C-rater: Scoring of short-answer questions. *Computers and the Humanities*, *37*, 389–405.

Lin, D. (1998). Automatic retrieval and clustering of similar words. In *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics* (Vol. 2, pp. 768–774). Retrieved from http://dl.acm.org/citation.cfm?doid=980691.980696

Marcu, D. (1999). Discourse trees are good indicators of importance in text. In I. Mani & M. Maybury (Eds.), *Advances in automatic text summarization* (pp. 123–136). Cambridge, MA: MIT Press.

Miller, G. A. (1995). WordNet: A lexical database. *Communications of the ACM*, *38*(11), 39–41.

Nagy, W. E., Garcia, G. E., Durgunoglu, A. Y., & Hancin-Bhatt, B. (1993). Spanish-English bilingual students' use of cognates in English reading. *Journal of Reading Behavior*, *25*(3), 241–259.

Perez, E. (1981). Oral language competence improves reading skills of Mexican American third graders. *Reading Teacher*, *35*(1), 24–27.

SDL. (n.d.). Automated translation. Retrieved from http://www.sdl.com/en/language-technology/products/automated-translation/

Shore, J., Burstein, J., & Sabatini, J. (2009, April). *Text adaptor: Web-based technology that supports ELL reading instruction.* Paper presented at the at the annual meeting of the American Educational Research Association, San Diego, CA.

Sparks, S. (2000). Classroom and curriculum accommodations for Native American students. *Intervention in School and Clinic*, *35*(5), 259–263.

Stenner, A. J. (1996). *Measuring reading comprehension with the Lexile framework.* Durham, NC: MetaMetrics.

Switlick, D. M. (1997). Curriculum modifications and adaptations. In D. F. Bradley, M. E. King-Sears, & D. M. Switlick (Eds.), *Teaching students in inclusive settings* (pp. 225–239). Needham Heights, MA: Allyn & Bacon.

Yano, Y., Long, M., & Ross, S. (1994). The effects of simplified and elaborated texts on foreign language reading comprehension. *Language Learning*, *44*, 189–219.

Notes

[1] More details about the lesson planning component are available in a video series at the Language Muse system homepage: http://ntis31.ets.org/ETS.ATA/login.html. The username and password “ets” can be used to access the application.

[2] This tool is an enhancement of an earlier tool, Text Adaptor, which did not include the lesson planning component (Burstein et al., in press; Shore et al., 2009).

[3] The automated translation of the English-to-Spanish language pair uses a tool from SDL (n.d.).

[4] The formula to determine a word’s standard frequency index value is as follows: SFI = 10(Log10(1,000,000 * F/N) + 4), where *F* is the word frequency and *N* is the total number of words.

[5] http://www.ldc.upenn.edu/Catalog/CatalogEntry.jsp?catalogId=LDC2006T13

Appendix

List and Description of TEA-Tool Features

Table A1

Syntactic (Sentence Complexity) Features

Feature name	Description	Examples
Passives	In contrast to active sentences, in passive sentences the object (book) precedes the subject (student) in the sentence.	Active: *The student read the book aloud.* Passive: *The book was read aloud by the student.*
1+ Clauses	These are sentences that contain one independent clause and at least one dependent clause.	*The teacher read the newspaper article aloud to the class because it was relevant to the lesson.*
Long Prepositional Phrases	These are prepositional phrases that contain at least two prepositional phrases in sequence.	*The kindergarteners sat quietly on the large, round rug in the classroom.*
Complex Noun Phrases	These are noun compounds, or longer noun phrases with a hyphenated adjective modifier.	*school building, school parking lot, back-to-school night*
Complex Verbs	A complex verb is composed of at least two verbs forms.	*will have gone*, *plans to leave*

Table A2

Lexical (Vocabulary) Features

Feature name	Description	Examples
Basic Synonyms	These are more frequent synonyms, and possibly less difficult than the words in the text.	Word: *immediately* Basic synonyms: *at once, now, right away*
Challenge Synonyms	These are less frequent synonyms and possibly more difficult than the words in the text. (These may be used for vocabulary building activities.)	Word: *immediately* Challenge synonyms: *forthwith, instantly*
Antonyms	These are words that are opposites of words in the text.	*king; queen*
Cognates	These are Spanish words that sound similar to and have the same meaning as an English word.	*ceramic; cerámica*
Academic Words	Words that describe complex and abstract concepts, and are used across disciplines (Coxhead, 2000).	*analyze, approach, benefit, concept*
Frequent Concepts	These are words that appear repeatedly across a text.	*Jamestown may have been ultimately abandoned,…original Jamestown settlement became the first permanent English colony*
Multi-Word Expressions	These are multi-word expressions that have a specific meaning when they appear together. Similes are included in this category (e.g., *as happy as a clam*).	*run into, red tape*
Cultural References	These are words and phrases that may be unfamiliar to ELs due to limited exposure to U.S. culture.	*pizza, Idaho, U.S. Senate, bluebird, tulip*
Contractions	These are cases where two words have been joined for a contracted word form.	*I'll, she'd, would've*
Complex and Irregular Word Forms	These are morphologically complex or irregular verbs.	*extracurricular, writing, went*
Variant Word Forms	These are cases where word forms in a text share the same word stem, but correspond to different parts of speech.	*The teacher booked* (verb) *the bus for the field trip, and bought a few books* (plural noun) *to read on the bus.*
Homonyms	These are words that sound alike, but have different meanings.	*there, their, they're*

Table A3

Rhetorical and Discourse Relations (Discourse Transitions)

Discourse relations	Description	Example
Cause-Effect	Words or terms that indicate a cause-effect relation between text segments.	*The discovery of fossils of tropical plants in Antarctica led to the hypothesis that this frozen land previously must have been situated closer to the equator, in a more temperate climate where lush, swampy vegetation could grow.*
Compare-Contrast	Words or terms that indicate a comparison or contrast relation between text segments.	*He was a wise and patient leader; however, his son had inherited none of these traits and brought ruin down on the nation.*
Evidence & Details	Words or terms that indicate specific evidence or details between text segments.	*Recent theories, such as the influence of plate tectonics on the movement of continents, have revolutionized our understanding of the dynamic planet upon which we live.*
Opinion	Words or terms that indicate an opinion about a text segment.	*Obviously, the many glitches in this complex process should prevent us from acting rashly.*
Persuasion	Words or terms that indicate the author is trying to persuade the reader.	*Equally important, the colonists tried many industries, such as silk, wheat, glass, timber, and cotton, but none were profitable enough to sustain the colony.*
Summary	Words or terms that indicate a summary related to a text segment.	*In conclusion, family values are decaying and the government needs to take action.*