Language Muse: Automated Linguistic Activity Generation for English Language Learners

Nitin Madnani **Jill Burstein** **John Sabatini** **Kietha Biggers** **Slava Andreyev**
{nmadnani,jburstein,jsabatini,kbiggers,sandreyev}@ets.org

Educational Testing Service
Princeton, NJ 08541

Abstract

Current education standards in the U.S. require school students to read and understand complex texts from different subject areas (e.g., social studies). However, such texts usually contain figurative language, complex phrases and sentences, as well as unfamiliar discourse relations. This may present an obstacle to students whose native language is not English — a growing sub-population in the US. [1] One way to help such students is to create classroom activities centered around linguistic elements found in subject area texts (DelliCarpini, 2008). We present a web-based tool that uses NLP algorithms to automatically generate customizable linguistic activities that are grounded in language learning research.

1 Introduction

Recent educational standards adopted by several states in the U.S. (CCSSO, 2010) explicitly emphasize the need for students to read progressively more complex texts in different subject areas, to prepare for college and careers. However, to accomplish this, learners need to have a grasp of linguistic features related to vocabulary, word senses, figurative language, English conventions, and discourse structures.

English language learners (ELLs) generally struggle to acquire English language skills: reading, writing, speaking, and listening. These learners could be disadvantaged further if there were simply an increase in the complexity of texts without concurrent scaffolding to help them with the demands likely to enter the curriculum as a result of the new Standards (Coleman and Goldenberg, 2012). This suggests the need for subject area teachers to incorporate a more linguistically-based approach to support content comprehension (Christie, 1989; Christie, 1999). Yet, teachers often lack the training necessary to identify English language features that may challenge diverse groups of ELLs (Slavin and Cheung, 2004; Walqui and Heritage, 2012).

In this paper, we present *Language Muse*, an open-access, web-based tool that can address these needs.[2] Specifically, *Language Muse* can help subject area teachers support ELLs by automatically generating customizable activities derived from actual texts used in their classrooms. The activities are generated using several existing NLP algorithms and are designed to help ELLs with multiple aspects of language learning needed to support content comprehension: vocabulary, grammatical structures, and discourse & text organization.

Although *Language Muse* is related to existing work in the NLP literature on automatic question generation (Mitkov and Ha, 2003; Brown et al., 2005; Heilman and Smith, 2010), it can generate multiple activities for teachers' own texts, cover a significantly larger set of language constructs, and offer teachers much more customizability.

In subsequent sections, we first provide a description of the *Language Muse* NLP. Next, we describe how teachers interact with the backend and create activities. Finally, we present the results of a survey conducted with actual ELL teachers, and conclude with future work.

2 NLP Backend

Language Muse relies on a backend that uses NLP techniques and resources to identify a variety of linguistic features contained in an input text. The features being identified can be categorized as: (a) lexical entities (single word and multi-word expressions), (b) syntactic structure, and (c) rhetorical

[1] See http://www.ncela.us/files/uploads/9/growingLEP_0809.pdf

[2] http://languagemuse.10clouds.com.

Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics—System Demonstrations, pages 79–84,
Berlin, Germany, August 7-12, 2016.

(a) Lexical Entities	
Cognates	Identified using a manually-created dictionary that was verified by a native Spanish speaker (most U.S. ELLs speak Spanish as their native language).
Academic Words & Definitions	Words that describe complex and abstract concepts, and are used across disciplines, e.g., *analyze*, *benefit*. Identified using a manually-created list and definitions extracted using the Wordnik API (`http://developer.wordnik.com`).
Frequent Concepts	Words that appear repeatedly across the text. Identified using a heuristic that measures repetitions across paragraphs.
Multiword Expressions	Idioms, Phrasal Verbs, etc. Identified using a rank-ratio based collocation detection algorithm trained on the Google Web1T n-gram corpus (Futagi et al., 2008)
Contractions	Identified using regular expressions defined on constituency parses.
Complex Words	Morphologically complex or irregular verbs. Identified using a rule-based morphological analyzer (Leacock and Chodorow, 2003).
Morphological Variants	Generated using an algorithm that first over-generates variants using rules and then filters using co-occurrence statistics computed over Gigaword.
Synonyms	Generated using a thresholded combination of WordNet (Fellbaum, 1998), a distributional thesaurus (Lin, 1998), and SMT-based paraphrases (Bannard and Callison-Burch, 2005).
Antonyms	Generated using WordNet.
Homonyms	Generated using a manually-created list (Burstein et al., 2004).
(b) Syntactic Structure **Note**: All modules below use regular expressions on constituency parses.	
Relative Clauses	Sentences containing an explicit relative clause.
1+ Clauses	Sentences containing 1 independent clause and $>= 1$ one dependent clause. Note that this can also include sentences with relative clauses.
Complex NPs	Noun phrases with a hyphenated adjective or a prepositional phrase modifier.
Complex Verbs	Verb phrases with $>= 2$ verb forms, e.g., *will have gone*, *plans to leave*
(c) Rhetorical and Discourse Relations **Note**: All modules below use an adapted rule-based discourse analyzer (Burstein et al., 1998).	
Cause-Effect	Terms indicating a cause-effect relation between text segments, e.g., "The discovery of fossils of tropical plants in Antarctica *led to the hypothesis* that ..."
Compare-Contrast	Terms indicating a comparison or contrast between text segments, e.g., "He was a wise and patient leader; *however*, his son ..."
Evidence & Details	Terms indicating specific evidence or details between text segments, e.g., "Recent theories, *such as the influence of plate tectonics on the movement of continents*, have ..."

Table 1: The inventory of features provided by the backend NLP engine in Language Muse.

and discourse relations. Before we describe each category in detail, it is important to note that since the primary use case for *Language Muse* is to help teachers plan appropriate classroom activities for ELLs, it is important for the automatically generated activities to be as accurate as possible. Therefore, for many of the features, we rely on manually crafted resources, either directly or indirectly as a filter for a noisier statistical approach.

Table 1 shows the linguistic features that our system can identify in the three aforementioned categories and provides a brief description of the backend module is used to generate it.

3 Activity Generation

In this section, we describe how users interact with *Language Muse*, i.e., how they can automatically generate linguistic activities for any text and customize them to their own liking. *Language Muse* is completely free to use for all teachers. Teachers request an account using the form on the web site and receive their login information via email.

Once a teacher logs into *Language Muse*, she can get started either by choosing one of the 33 texts that we provide across three different content areas (English Language Arts, Science, and Social Studies), or by uploading her own classroom text (in plain text/.doc/.docx formats). The system currently limits the texts to 5000 words. All texts uploaded by a teacher are saved into her personal library for later re-use.

The text is then sent to the NLP backend for processing, which returns a JSON object containing all identified (or generated) linguistic features. At that point, the teacher can generate any of the available linguistic activities, each of which is based on one of the linguistic features. All activities were designed based on input from ELL content-area teachers. There are a total of 24 activities, grouped according to whether an activity is word-based, sentence-based, or paragraph-based. This form of hierarchical grouping is based on ELL literature which suggests that each level in the hierarchy presents distinct challenges and opportunities for language learning. Table 2 shows a few of the available activities and provides a brief description.

3.1 Recommended Activities

Based on the number of feature instances detected by the backend, *Language Muse* may recommend certain activities over others to the teacher. For example, if there were more words with synonyms but only a few cause-effect terms, it might recommend the *Synonyms in Paragraphs* activity but not the *Cause/Effect Relationships* activity. Some activities may also be unavailable since no instances of the corresponding linguistic feature could be detected by the backend. *Language Muse* makes a visual distinction between recommended activities, possible activities, and unavailable activities as shown in Figure 1. Clicking on an activity shows its description and a sample question.

3.2 Same Feature, Multiple Activities

Some activities are based on the same underlying linguistic feature but use it differently, depending on their level. For example, a word-level activity asks students to match words in one list to words in another list based on how similar they are in meaning. That activity uses automatically generated synonyms for the words in the text and then automatically populates the two lists – one with the original words and the other with the generated synonyms. There is a similar paragraph-level cloze activity that shows students a paragraph from the text and asks them to replace pre-identified words with their synonyms such that the meaning is unchanged. This activity uses the same underlying feature — automatically generated synonyms — but presents it differently. This exposes ELLs to a different part of the language construct.

3.3 Automatically-generated Answers

The questions for all activities are automatically generated based on linguistic features in the text. However, for 15 of those activities, *Language Muse* also automatically populates an answer key for the teacher. For example, for the word-based synonym activity described in §3.2, we know which pairs of words in the two lists match each other since the synonyms were automatically generated. Automatically-generated answers reduce the time that a teacher needs to edit an activity for her classroom. See Table 2 for additional examples.

3.4 Customizability

It is impossible for *Language Muse* to always provide exactly what every teacher is looking for. Therefore, almost all aspects of the activities it generates can be customized to suit a teacher's needs. Among other things, the teacher can choose to:

- edit the instructions shown to the students,
- hide any or all automatically chosen words/sentences/paragraphs,

Sentence activities
Multiple Clause Sentences. Shows multi-clause sentences and asks students to break them up into two or more shorter sentences. Although the sentences for the activity are identified automatically in the text, the answers are not generated automatically. Example: *Organelles are structures visible within a cell that have their own structure.* ⇒ (1) *Organelles are structures visible within a cell.* (2) *Organelles have their own structure.*
Cause/Effect Relationships. Shows sentences containing causal relationships and asks students to identify the cause, the effect, and the connector word that denotes the causal relationship. The sentences with causal relationships in the text are identified automatically but only the connector word part of the answer is automatically generated. Example: Off the coast of Canada, commercial cod fishing had to stop because the population of cod collapsed. ⇒ *The population of cod collapsed off the coast of Canada* (cause), *Commercial cod fishing had to stop* (effect), *because* (connector).
Homonyms in Sentences. Shows sentences with blanks and asks students to fill in the right word from a list that contains homonyms as distractors. Examples can be seen in Figure 2.
Paragraph activities
Variant Word Forms in Paragraphs. Uses inflectional and derivational word variants generated by the backend. Shows students a paragraph of text with blanks and asks them to fill in the right morphological variant. Answers and distractors are automatically generated. Example: *Scientists suspect that there are more than 10 million _____ (different/difference) types of life forms on Earth.*
Phrasal Verbs. Asks students to pick the correct preposition to complete the phrasal verbs found in the paragraph. Answers and distractors are automatically generated. Example: *People usually think _____ (at/on/of) the heart, lungs and brain as vital organs.*

Table 2: A subset of the linguistic activities available to teachers in *Language Muse*.

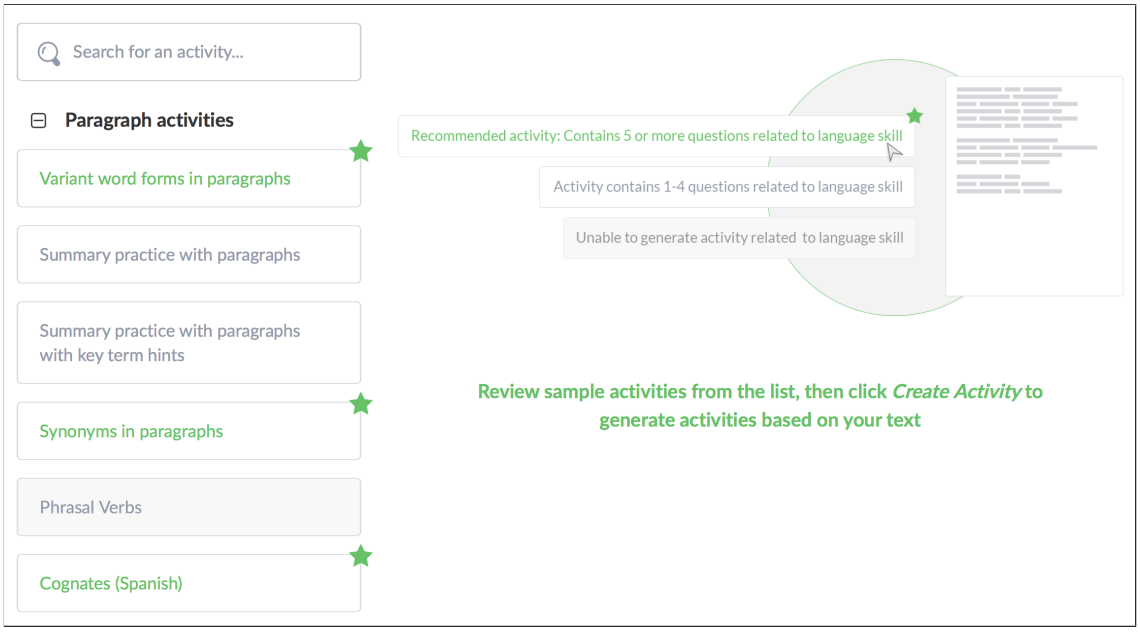

Figure 1: Activities where 5 or more questions can be generated are recommended by *Language Muse* and marked with a star. Activities with fewer than 5 questions are not marked but can still be chosen by the teacher. Activities with no available questions are greyed out and cannot be chosen.

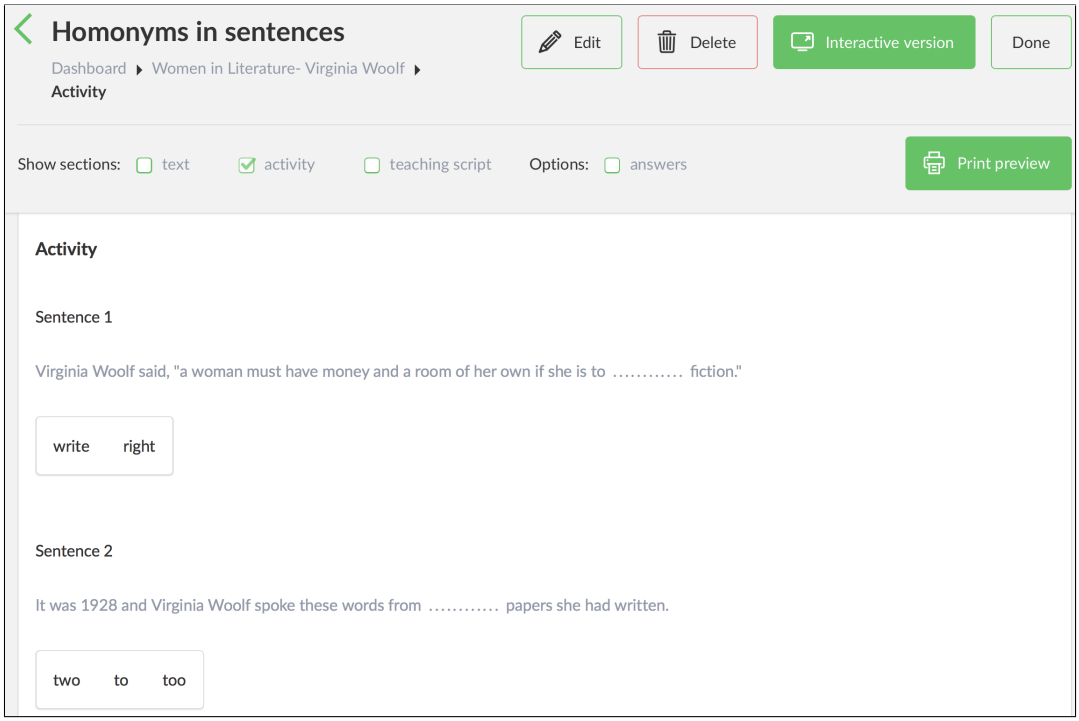

Figure 2: The Homonyms in Sentences activity generated from a Language Arts text on Virginia Woolf. The questions, answers, and distractors are all automatically generated.

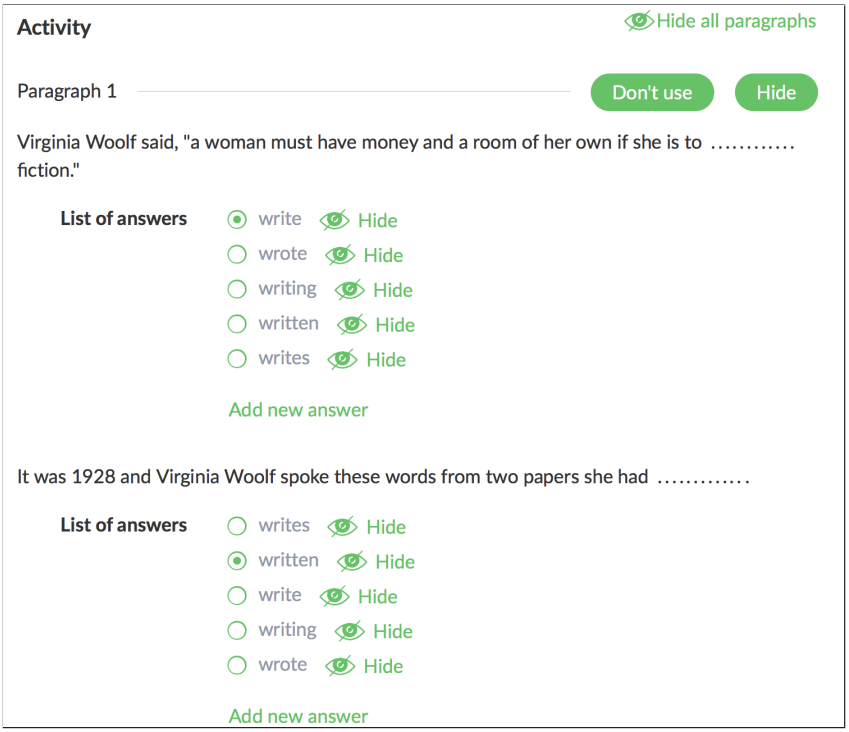

Figure 3: An example of a customizable multiple choice activity (*Variant Word Forms in Paragraphs*). Teachers can (a) change the generated correct answer, (b) hide any generated distractors, (c) add their own answers, and (d) hide a paragraph entirely or use it only as context without generating questions.

- edit the automatically generated answers or add her own, and
- edit the list of automatically generated distractors for multiple choice questions.

Figure 3 illustrates this by showing the *Variant Word Forms in Paragraphs* activity in edit mode.

4 Teacher Survey

We wanted to evaluate whether activities generated by *Language Muse* are useful to teachers. To do this, we worked with seventeen 6th-8th grade teachers who taught English language arts, science, and social studies. Four teachers had been teaching for two years or less, five for 3-9 years, and the rest for > 10 years. All but two currently had responsibility for teaching ELLs and eight had been teaching ELLs for > 5 years. We asked them to examine 9 different activities from *Language Muse* and tell us whether they would consider using them in their classrooms. Figure 4 shows the results of our survey which are encouraging for a first version.

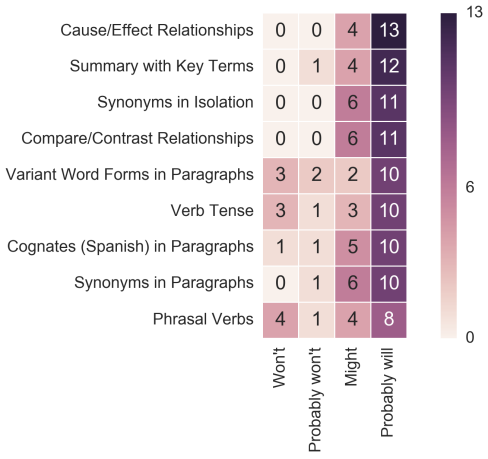

Figure 4: A heatmap showing the results of our teacher survey. Each cell shows the number of middle school teachers responding to whether they would use the corresponding activity in classrooms.

5 Conclusions & Future Work

We presented *Language Muse*, an open-access, web-based tool that can help content-area teachers support ELL students with content comprehension. We are currently working on the next version which will allow students to log into Language Muse to complete any activities assigned to them by their teachers and also receive feedback. All development on Language Muse continues to be informed by frequent and detailed interactions with teachers.

Acknowledgments

Research presented in this paper was supported by the Institute of Education Science, U.S. Department of Education, Award Number R305A140472, and by the 10Clouds front-end development team.

References

Colin Bannard and Chris Callison-Burch. 2005. Paraphrasing with Bilingual Parallel Corpora. In *Proceedings of ACL*.

Jonathan C Brown, Gwen A Frishkoff, and Maxine Eskenazi. 2005. Automatic Question Generation for Vocabulary Assessment. In *Proceedings of EMNLP*.

Jill Burstein, Karen Kukich, Susanne Wolff, Chi Lu, and Martin Chodorow. 1998. Enriching Automated Scoring using Discourse Marking. In *Proceedings of the ACL Workshop on Discourse Relations and Discourse Marking*.

Jill Burstein, Martin Chodorow, and Claudia Leacock. 2004. Automated Essay Evaluation: The Criterion Online Writing Service. *AI Magazine*, 25(3):27.

CCSSO. 2010. *Common Core State Standards for English language Arts & Literacy in History/Social Studies, Science, and Technical Subjects. Appendix A: Research supporting key elements of the Standards*. Washington, DC.

Frances Christie. 1989. *Language Education*. Oxford University Press, Oxford, UK.

Frances Christie. 1999. *Pedagogy and the Shaping of Consciousness: Linguistics and Social Processes*. Continuum, London, UK.

Rhonda Coleman and Claude Goldenberg. 2012. The Common Core Challenge for English Language Learners. *Principal Leadership*, pages 46–51.

Margo DelliCarpini. 2008. Success with ELLs. *English Journal*, 98(2):98–101.

Christiane Fellbaum. 1998. *WordNet*. Blackwell Publishing Ltd.

Yoko Futagi, Paul Deane, Martin Chodorow, and Joel Tetreault. 2008. A Computational Approach to Detecting Collocation Errors in the Writing of Non-native Speakers of English. *Computer Assisted Language Learning*, 21(4).

Michael Heilman and Noah A Smith. 2010. Good question! Statistical Ranking for Question Generation. In *Proceedings of NAACL*.

Claudia Leacock and Martin Chodorow. 2003. C-rater: Automated scoring of Short-answer Questions. *Computers and the Humanities*, 37(4).

Dekang Lin. 1998. Automatic Retrieval and Clustering of Similar Words. In *Proceedings of COLING*.

Ruslan Mitkov and Le An Ha. 2003. Computer-aided Generation of Multiple-choice Tests. In *Proceedings of the Workshop on Building Educational Applications*.

R. E. Slavin and A. Cheung. 2004. How do English language Learners Learn to Read? *Educational Leadership*, 61.

A. Walqui and M. Heritage. 2012. Instruction for Diverse Groups of ELLs. In *Understanding Language Conference*, Stanford, CA.