Implementation of Data-driven Learning (DDL) and the Effects of Machine Learning on the Classroom

Sarah Deutchman

Waseda University & Wayo Women's University

Vignette

With technology continually advancing it is important to make use of new tools in the classroom. For example, having access to corpora, large searchable collections of texts on particular subjects, is like being able to consult a native speaker at any time. Corpora can be used to search for grammatical patterns and other information about words. The process of looking for patterns is called data-driven learning (DDL). If students are taught how to do DDL they can gain more autonomy and become more aware of how language works.

However, there are some negative sides of AI and machine learning which affect us already. For example, DeepL has been rapidly improving at translating text, which can lead to the problem of students using auto translators to write their papers. This translated text can then be put through Grammarly to fix any grammatical errors. After this process is complete, the paper may seem as if it has been written by the student.

Another negative aspect of improved technology is data mining, such as Facebook quizzes that ask questions to get password information or answers to security questions. If people post these answers, they are giving hackers the ability to get into their accounts.

Objectives

The purpose of the paper is to inform fellow educators about tools that can be used in the classroom to do DDL. Additionally, this paper will explain ways to protect their data and students' data.

Practical Implications

Corpora are now easier to use as many are free and web-based. With corpora, students can see which genre a word belongs to, see how it is used in a sentence, create genre-specific vocabulary lists, and see how words connect. Three commonly used tools are Sketch Engine for Language Learning (SKELL), AntConc, and the Contemporary Corpus of American English (COCA).

SKELL is a web-based corpus that is geared toward beginners. It does not require an account and searches can be done easily. For more frequent words it gives information on collocations (i.e., words that are paired together). This is important because of language influence; students may try to pair words incorrectly based on their L1. In Japanese, you *drink medicine* but in English you *take medicine*. If students use SKELL to search they can see for themselves which words are neighbors and see how the words are used in context. With SKELL, teachers can create different activities such as fill-in-the-blank activities, multiple-choice questions, cross-out words that do not pair with each other, etc. The only downside of SKELL is that for low-frequency vocabulary words, such as uxorious, only sentences are provided.

AntConc is free software that can be used to create corpora. Student essays can be imported into the software and searched for common errors. For four years I have taught academic writing classes where students had to write essays about studying abroad. One of the reasons students gave for studying abroad was to *touch culture*. With AntConc I could search for the word *culture* and see if many students were making this mistake. After identifying a common error, I could correct it by creating activities to show students that the word *experience* should be used with *culture*. Another good point of AntConc is that Laurence Anthony, the creator of the software, has a Youtube channel that shows how to use his free software (https://www.youtube.com/user/AntlabJPN).

COCA is a free corpus with many features. For example, it can be used to compare two similar words such as *utter* and *sheer*. *Utter* is used with negative words (e.g., failure, darkness, disregard) while *sheer* is used with more descriptions of scale (e.g., number, volume, magnitude). It is also possible to link COCA with different websites such as Youglish (a searchable YouTube corpus) and Linguee (a kanji dictionary website). Two of the downsides of COCA are that users must create an account to use the

website and the sentences provided are not always complete so students should be made aware of this.

In short, teachers can create activities from more controlled to less controlled. It is possible to use corpora for all levels. If students are beginners, students can be shown corpora in their L1 and L2 and look for patterns between the two.

At the same time, plagiarism and data mining exemplify the negative aspects of advancements in technology. Unfortunately, many plagiarism detectors are not able to flag papers as being translated from other languages, or copied from other papers. This is because Turnitin is based on corpora; the software looks for similar phrases. If a paper has not been imported into the system yet, a plagiarism checker will not flag any content from that paper or essay as plagiarized. Bitwarden and Lastpass are examples of password management software. Having a password manager allows users to have more complicated passwords that will be more difficult to crack and can store complex answers to security questions. It is better to use random strings of alphanumeric characters instead of actual answers, which makes it harder for hackers to access accounts.

Reflective Conclusion

Technology is advancing quickly and already has an influence on our lives as teachers, so it is important to stay informed about innovations that have taken place. Quillbot and Wordtune are two new websites of note that have been trained on machine learning models. These sites may cause concern among those who teach academic writing because they can help to summarize, paraphrase, and automatically edit texts. Thus, educators may have to decide what constitutes plagiarism and whether these tools should be allowed in class. Additionally, I found it very encouraging that one of the people who watched my presentation could create their own activities and successfully implement them in class.

Author Bio

Sarah Miyoshi Deutchman teaches at Waseda University and Wayo Women's University as a part-time lecturer. She has taught English for over 14 years in three different countries. She has been teaching at the university level for 4 years. Her areas of research include data-driven learning, corpus linguistics, and vocabulary.

<sarah.deutchman@gmail.com>

Link to Presentation on YouTube:
https://www.youtube.com/watch?v=j26c7f3v3s4