Language Technology and Its Role in Language Teaching and Learning

Takako Aikawa
Massachusetts Institute of Technology

Abstract

This paper discusses the use of language technology for language teaching and learning. I specifically discuss how two language technologies (i.e., NLP (natural language processing) and MT (machine translation)) can be related to language teaching and learning. First, I provide a brief introduction of NLP and explain how NLP can help our daily interactions with computer. Then I discuss some details about MT while addressing challenges related to MT. Also, I present features from Google Translate, which I think would be useful for language learning. At the end of the paper, I discuss issues about MOOCs (Massive Open Online Courses). The discussion related to MOOCs is anecdotal but I provide some useful references that can help us understand the current status of MOOCs. I hope that this paper is informative and can help us envision the direction of language teaching in the future.

Keywords: Natural Language Processing (NLP), Machine Translation (MT), Massive Open Online Courses (MOOCs), Flipping

1. Introduction

It has become obvious that technology can play an important role for language teaching and learning. Many teachers have already incorporated various types of technology (e.g., Skype, iBook, Facebook, Wiki, Forums, Blogs, etc.) into their language curricula. The use of such technology in our classroom is becoming more and more

popular, and many studies have been done that discuss how such technology can be used for language teaching.[1]

However, not many studies have been done to investigate the potential use of natural language processing (NLP) technology and machine translation (MT) technology. Both NLP and MT are to deal with human languages, and they tackle various types of linguistics or translation issues of human languages. It would be natural for language teachers to entertain potential use of these technologies for language teaching. One of the goals of this paper is to introduce some of these language technologies and advocate them for language teaching. Another goal is to discuss issues on Massive Online Open Courses (MOOCs). I provide information on the current status of MOOCs while presenting my anecdotal opinion about MOOCs.

The outline of the paper is as follows: Section 2 provides a brief overview of NLP technology. I explain how NLP technology is being used in our daily life while pointing out the importance of the role of NLP in facilitating interactions between humans and computer. Section 3 focuses on MT. First I provide a brief introduction to MT technology. Then, I present features of MT technology that can be useful for language learners. In Section 4, I switch gears a little bit and discuss some issues related to MOOCs. I present an overview of the current status of MOOCs and my anecdotal opinions about MOOCs. Section 5 provides concluding remarks. I contend that in adopting technology for language teaching, we should not underestimate the importance of re-examination of our existing curricula and that of cross-disciplinary collaboration. I hope that my discussion in this paper is informative and helpful for us to envision and shape the direction of language teaching in the future.

2. Natural Language Processing

NLP is the technology that deals with natural human languages. Different types of NLP technologies exist. For instance, NLP can be used for parsing a sentence, so that we can get the basic syntactic structure of a given sentence automatically. Further, some parsers can generate the so-called "logical form (LF)" (i.e., basic semantic structure) of a sentence. For instance, Figure 1 and Figure 2 are the snapshot of the syntactic tree and that of the LF using the Japanese parser developed at Microsoft Research, respectively. [2]

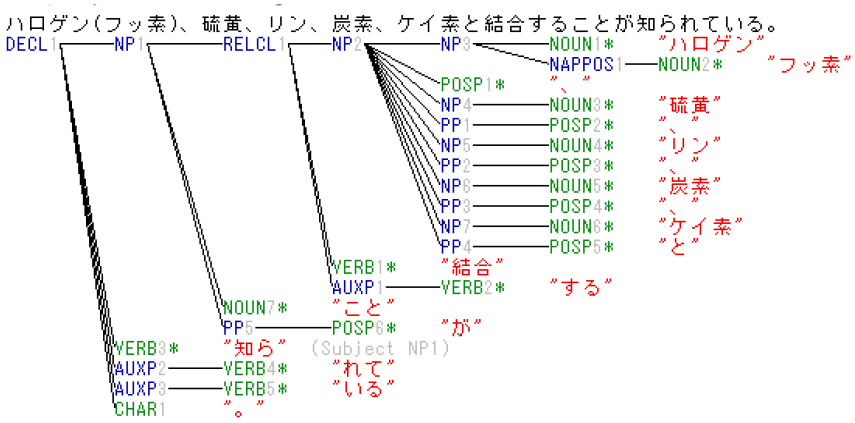

Figure 1: Parser tree of the sentence, "ハロゲン(フッ素)、硫黄、リン、炭素、ケイ素と結合することが知られている。"

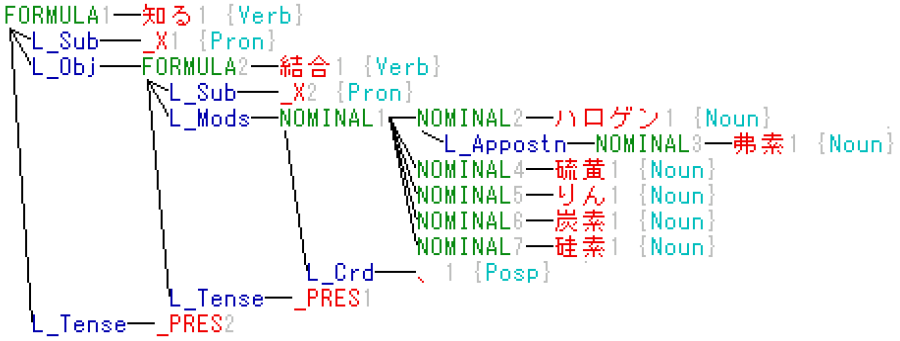

Figure 2: Logical Form based on the parse tree in Figure 1.

NLP plays a critical role for facilitating interactions between humans and computer, and it is everywhere when looking at our daily life. Take for instance search engines (e.g., Google, (https://www.google.com/), Bing (http://www.bing.com/), etc.). When we

send a query to a search engine (whatever it may be), the engine returns most relevant pages from millions of web sites based on the given query. How would that be possible? Further, the search engine can rank these pages (called "PageRank") based on various types of criteria. How can such ranking be done appropriately in such a short period of time?[3]

Another example of NLP technology that can be found in our daily life is a spell checker or a grammar checker. Speech is also a part of NLP. For instance, speech recognition or text-to-speech technologies are being used extensively in our mobile applications or navigation systems.

Sentiment analysis can be found everywhere as well. Sentiment analysis is to understand the "emotional attitude" of users (e.g., positive, negative, etc.), and it uses the so-called "machine-learning" techniques.[4] I will not get into any technical details on sentiment analysis here, but you can imagine the following scenario, for instance. Suppose that you are an owner of a big car company and you want to analyze which car(s) have positive or negative feedback based on users' blogs or feedback on the Web. Theoretically, it is possible for you to read each of the users' comments and classify their comments accordingly, so that you can identify which car(s) have positive or negative feedback. But this is not realistic to do. Sentiment analysis technology can do such a task instantly for you, and many companies are using this technology to analyze their customers' feedback.[5]

The applications mentioned above are just a small set of the examples where NLP technology is involved. NLP is everywhere in our daily life and it plays a critical role in facilitating interactions and communications between humans and computer.

3. Machine Translation (MT)

In this section, I first provide an overview of an MT system in general without getting into any technical details. Then, I present some linguistic features provided by Google Translate, which might be useful for language learners.

3.1. A Brief Overview of Machine Translation

Language translation is not a trivial task for a computer to tackle, and MT itself utilizes various types of NLP technologies. MT raises many intriguing linguistic or language acquisition-related questions like the following:

- "How can computer learn mappings between two human languages?"
- "How can computer understand differences in word order between languages like English and those like Japanese?"[6]
- "Does computer understand the meaning of a sentence in one language and translate it into another language?"
- "Is the way machine learn a language same as the way humans learn a new language?"

There are various reasons why MT is so challenging for us to tackle. First, human languages are so different from each other. Machine does not know anything about syntax or semantics of human languages. Yet, it has to learn all sorts of lexical mappings between two languages in order to translate from one language into another. At the same time, machine has to learn differences in word order, inflection, number/gender agreements, etc. between two languages for successful language translations. This is not a trivial task at all, even for humans.

Second, human languages are ambiguous. The same word in one language can be sometimes translated into different words in another language, depending on the given

context(s). For instance, let us take the case of translating the English word "cell" into Japanese. This word can be translated into the biological "cell" (細胞); or into the prison "cell" (牢屋); or into the "cell" phone (セル、携帯). The choice of such translations is totally dependent on the context. So, in order for machine to translate human languages properly, machine has to understand the context properly and based on that context, it has to select an appropriate translation. This again is not trivial.

Third, human languages have many idiomatic or set phrases, and machine is not good at handling such phrases. For instance, translating the simple sentence "How are you?" into Japanese (i.e., お元気ですか) is quite difficult; if machine tries to do the word-to-word translation, it would immediately fail.

These are simply the tip of the iceberg of MT-related issues, and there are many more problems to be tackled for machine to translate human languages properly. Nonetheless, existing (web-based) free MT systems such as Google Translate or Bing Translator do a reasonable translation job.[7] Moreover, these MT systems can translate into/from many languages. How can machine handle so many languages with reasonable translation quality?

To answer this question, first I would like to provide a brief history of MT. MT was initially developed based on human hand-coded rules. Such MT system is called a "rule-based MT." A rule-based MT however has several limitations in terms of scalability; that is, since it takes an enormous amount of time and human resources to build one rule-based MT system, it won't be able to scale up to many languages easily. This scalability issue forced the field of MT to switch from a rule-based approach to a statistical approach.

The statistical approach is roughly speaking to let machine learn all sorts mappings along with their probabilities via feeding lots of bilingual parallel corpus data (対訳コーパスデータ). With millions of bilingual data, machine nowadays can figure out statistically the most probable translation in the target language given an input in the source language.[8] Google provides a nice explanation about how their MT system works using a statistical approach. I recommend the readers to visit http://translate.google.com/about/ for more details about a statistical approach to MT.

3.2. Useful Features for Language Learners

This sub-section presents some linguistic features available from Google Translate (http://translate.google.com), which I think would be useful for language learning. First, their speech recognition and text-to-speech features, which are shown in Figure 3 below, can be useful for language learning; learners can use the speech recognition feature to attest whether their pronunciation is correct. If their pronunciation is bad, the system will not recognize their utterance properly. Language learners can also use the text-to-speech feature for their listening practice.

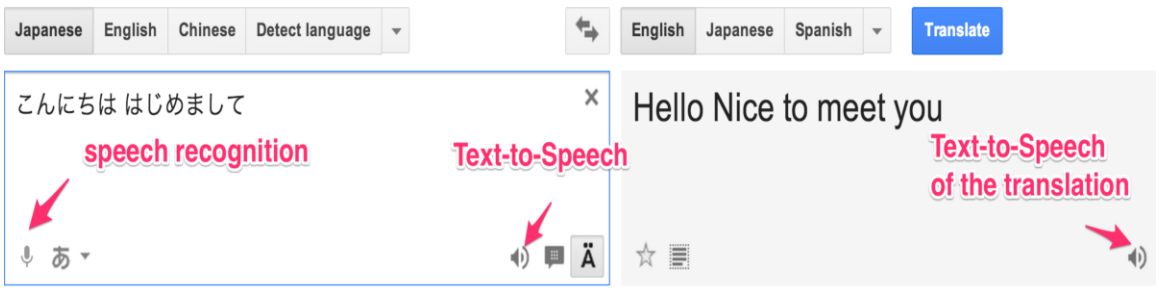

Figure 3: Speech Recognition and Text-to-Speech Features on Google Translate

Another linguistic feature that might be useful for language learning is the so-called "word-alignment" information; users can see what word(s) in the source language correspond to what word(s) in the target language as shown in Figure 4 below.

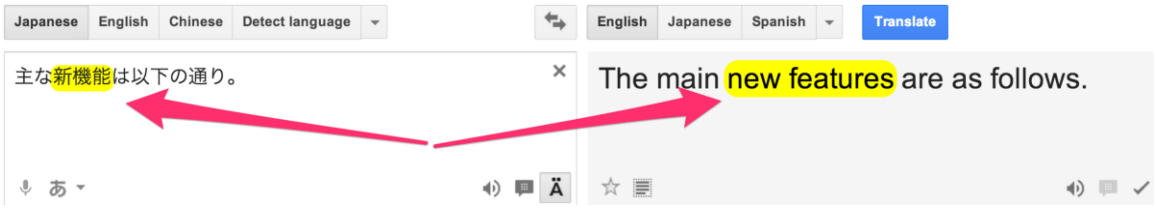

Figure 4: Word-alignment Information on Google Translate

Google Translate also provides synonyms or alternative translations as shown in Figure 5. This type of lexical information is very useful for language learners.

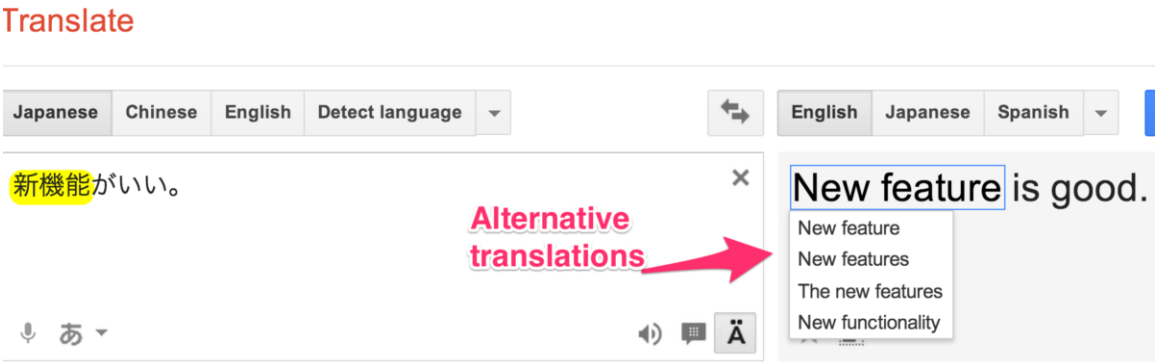

Figure 5: Display of Alternative Translations

There are many other great features available from web-based MT systems, which I think would be useful for language learning. I encourage language teachers to visit such MT sites and entertain potential usage of their linguistic features.

4. MOOCs (Massive Open Online Courses)

Here, I would like to switch gear a little bit, and would like to address topics related to MOOCs (Massive Open Online Course(s)). Many universities have already been

participating in MOOCs, and it is predicted that the number of the participants in MOOCs will continue to grow.[9]

One goal of MOOCs is to provide people more opportunities to learn openly (=freely) and provide high-quality educational contents on the web; MOOCs aim at the equity of educational opportunities.[10] This conceptual goal of MOOCs sounds good but there are many people who are skeptic about the current status of MOOCs. Many blogs and articles have been published to address problems and concerns about MOOCs (e.g., Vollmer (2012), Perna, et. al. (2013), Kim (2013), among others). The following lists some of the skepticisms on MOOCs:

- *Lack of authentic interactivity*: MOOCs won't be able to provide interactivity as we can do in classrooms.
- *Cost*: Creating instructional video(s) is quite expensive and requires a huge amount of instructor's time. How can we afford such things?
- *Assessment*: How do we know the impact of MOOCs on learning? Do we have any good way to measure it?
- *Demographics*: Are MOOCs United State-centric? Most of the current participants are from the United States. Can we scale up to other parts of the world quickly?

It seems that MOOCs have not met the original expectations yet, and many people are extremely doubtful about their slogan; that is, MOOCs can provide a solution to increase learning opportunities, raise educational standards, and decrease costs in the higher education. Scholz (2013) examines issues on MOOCs extensively, and I recommend this article to the readers who are interested in learning more about MOOCs.

Anecdotally speaking, however, I think that the direction of our future education will move toward (not away from) MOOCs regardless of these concerns and skepticisms. Our students may learn things totally differently from us. For instance, they may not

need classroom interactions to learn a new language. Or, such interactions can be done remotely in a virtual space. I contend that MOOCs (as well as technology) are expected to affect or change our language curricula and pedagogy. It is important for us to explore ideas on what to do with MOOCs earlier rather than later and get ready for MOOCs.

5. Concluding Remarks

It is clear that we are in the middle of transitioning from the traditional classroom-based teaching style to new digital or online-based teaching style(s). This transition forces us to re-think about what activities should be done in classroom as technology enables us/students to do lots of things outside classroom. In response to that, many educators started advocating and experimenting "flipping".[11]

Flipping involves the inversion of the activities that used to be done in the classroom and those that used to be done at home. Under the traditional curriculum, a teacher delivers instructions to students in the classroom and students do homework at home. Under the flipped classrooms, students get instructions via instructional videos at home, and in the classroom, they do homework with their teacher(s) and classmates. This inversion of the activities naturally shifted the role of teachers "from sage on the stage to guide on the side" as discussed in King (1993).

Here, I would like to point out two things in adopting technology for language teaching. First, if we are to integrate technology into our language curricula, we need to take a closer look at the existing curricula and pedagogy and bring in some changes, so that we can better utilize our classroom time.

Second, I would like to emphasize the importance of cross-disciplinary collaboration. Language teachers tend not to know details about software engineering issues. Thus, even if language teachers have ideas on what technology they want to use for teaching, they have little idea on how to get the technology or where to start. On the other hand, software engineers do not know much about language teaching or pedagogy. Educational software developed by developers or programmers alone are thus most likely to fail to adequately meet the demands of language teachers. I believe that if teachers and engineers can collaborate with each other, we can move forward much faster. We should never underestimate the power of cross-disciplinary collaboration.

References

Eaton, Sarah E. (2012) *Skype for Literacy and Language Learning: "How To" Tips and Best Practices for Teachers*. Retrieved from http://www.springinstitute.org/Files/howtouseskypeforlanguageteachingandliteracyspdf.pdf.

Harry (2013, September) *How Search Engines Rank Web Pages*. Retrieved from http://searchenginewatch.com/article/2064539/How-Search-Engines-Rank-Web-Pages

Kim, Joshua (2012, May), *Playing the Role of MOOC Skeptic: 7 Concerns*. Retrieved from http://www.insidehighered.com/blogs/technology-and-learning/playing-role-mooc-skeptic-7-concerns#ixzz2nYxMgUeW

King, Alison (1993) *From Sage on the Stage to Guide on the Side.* College Teaching Vol. 41, No. 1, pp. 30-35.

Perna, Laura and et. al. (2013, December) *The Life Cycle of a Million MOOC Users*. Retrieved from http://www.gse.upenn.edu/pdf/ahead/perna_ruby_boruch_moocs_dec2013.pdf

Scholz, Claudia W (2013) *MOOCs and the Liberal Arts College*. Retrieved from http://jolt.merlot.org/vol9no2/scholz_0613.htm

Vollmer, Timothy (2012, November) *Keeping MOOCs Open*.
Retrieved from http://creativecommons.org/weblog/entry/34852

Footnotes

[1] For instance, Skype seems to provide video clips and various articles related to the use of Skype in classroom. Please visit https://education.skype.com/. Also, visit http://www.skypelearn.com/teach-languages-using-skype.html where you can be a language teacher or a language learner using Skype. See also Eaton (2012) for more references.

[2] See http://research.microsoft.com/en-us/projects/japanesenlp/ for more details.

[3] Harry (2013, September) provides good explanations to these search engine-related questions. Also, there is a video clip on that website called "Understanding Signals" provided by Google. This video explains in simple terms how a search engine works, and it is worth watching it.

[4] For more details, see http://en.wikipedia.org/wiki/Machine_learning.

[5] For more details, see http://en.wikipedia.org/wiki/Sentiment_analysis.

[6] Languages like English are often classified as "SVO" (subject-verb-object) languages whereas those like Japanese as "SOV" (subject-object-verb) languages in linguistics

[7] Google Translate is available from http://translate.google.com/, and Bing Translator is available from http://www.bing.com/translator.

[8] Google Translate and Bing Translator are both statistical MT systems.

[9] The website (http://www.mooc-list.com/) provides the list of the universities and the educational institutions currently participating in MOOCs.

[10] For more details on the historical development of MOOCs, see the Wiki page of MOOCs (http://en.wikipedia.org/wiki/Massive_open_online_course).

[11] Terms such as "blended learning" or "hybrid learning" are also used to refer to the approach that combines the online delivery of instructions and the face-to-face instructions.