Data-Driven Learning of Speech Acts Based on Corpora of DVD Subtitles

S. Kathleen Kitao[1] and Kenji Kitao[2]

Abstract. Data-driven learning (DDL) is an inductive approach to language learning in which students study examples of authentic language and use them to find patterns of language use. This inductive approach to learning has the advantages of being learner-centered, encouraging hypothesis testing and learner autonomy, and helping develop learning skills. The approach has grown out of corpus linguistics, and it has been used to help students learn grammar and vocabulary usage. It is also possible to use it for teaching speech acts. In this paper, we contrast inductive and deductive learning and discuss an example of an exercise using data-driven learning to learn about the expressions and strategies used in apologies.

Keywords: speech act, apology, data-driven learning, inductive approach.

1.   Introduction

An important aspect of language education is the development of communicative competence, including use of speech acts. In this paper, we will discuss inductive and deductive approaches to learning and look at data-driven learning, one inductive approach, and its application to teaching apologies.

1.1.   Inductive and deductive approaches

One way to categorize approaches to language teaching is to divide them between deductive and inductive approaches. A teacher using a deductive approach explains grammar rules, meanings of words, etc., and then the students

---

1. Doshisha Women's College, Kyoto, Japan; kkitao217@yahoo.com

2. Doshisha University, Kyotanabe, Japan

do activities to apply what they have learned. This is a teacher-centered approach and a traditional way of teaching.

In contrast, in an inductive approach, the teacher gives students examples of the grammar point or uses of the vocabulary words and the students make observations and make generalizations from those observations. This involves hypothesis formation and testing of the hypotheses, and it is a student-centered approach.

Inductive approaches have both advantages and disadvantages. On the one hand, they allow more student participation and involvement. Students understand better and remember more, and using inductive approaches helps students develop autonomy. On the other hand, they do take longer, especially if the concepts are complex (Tian, 2005).

1.2. Data-driven learning

Data-driven learning is an inductive approach that makes use of corpora to find examples of lexical usage or grammatical points so that students can make generalizations about them. It is defined by Johns and King (1991) as "the use in the classroom of computer generated concordances to get students to explore the regularities of patterning in the target language, and the development of activities and exercises based on concordance output" (p. iii). Students use examples from a corpus to find patterns and develop rules.

While DDL was originally developed as a way to study grammar and vocabulary, it has also been used to teach speech acts and sociopragmatic guidelines.

2.   An example of data-driven learning

In this section, we will explain an example of the use of DDL to teach apologies.

2.1. Apologies

Because they are searchable in a corpus, apologies are an appropriate speech act to use for data-driven learning. Most apologies can be found using lemmatised searches of five performative words: sorry, pardon, excuse, forgive, and apologize. In one study, using a corpus of downloaded subtitles from DVDs, lemmatised searches for the five keywords found 98% of the apologies in the corpus (Kitao, 2012).

2.2. Developing materials

A corpus of DVD subtitles can by compiled by downloading the subtitles from DVDs using a program called SubRip (http://www.videohelp.com/tools/Subrip). This program creates .srt files, which, if the srt extension is changed to txt, can be searched with concordancing software. The files include minutes and seconds indicating how far from the beginning each subtitle occurred, which can be useful in finding scenes.

SubRip files are often posted online. A search for the title of a movie or television program and "srt" will find links to these files. Also, a large collection of .srt files can be found at http://www.tvsubtitles.net/. You can compile the .srt files for individual series' episodes or movies into a text file. In addition, scripts can be found at such websites as Drew's Scripts-O-Rama Index at http://www.script-o-rama.com/table.shtml.

The choice of material is important. Huang (2004) found that movies of the genre romance/comedy and dramas portrayed ups and downs in relationships that required apologies that reflect real life. Television comedies and dramas would have similar characteristics. In contrast, Huang (2004) found that science fiction and action movies concentrated more on action and had few apologies. Rose (2001) suggested that material used to teach speech acts be less than 15 years old and depict contemporary characters in real-life situations.

Depending on the type of class and goals, the teacher can have students compile their own corpora and search for themselves, search for themselves in corpora provided by the teacher, or provide examples from a corpus with explanations of the background, names of speakers, etc. If the focus of the class is corpus linguistics, one of the former two methods would be preferable; if the focus of the class is on speech acts or linguistic pragmatics, the latter is preferable. If students search for themselves, they will need to look at the scenes on DVDs in order to understand what the context of the apology was, for instance, which character spoke which line. A concordancer can be used to find the apologies (and certain other speech acts). A concordancer, developed by Laurence Anthony can be downloaded at http://www.antlab.sci.waseda.ac.jp/software.html.

When one of the authors developed a data-driven learning activity to teach apologies in a linguistic pragmatics course, she used a corpus of DVD subtitles that had been developed from the first three seasons (2009, 2010 and 2011) of

the US situation comedy *Modern Family* (Levitan & Lloyd, 2009). She had compiled a list of apologies from the corpus for a study of apologies (Kitao & Kitao, 2013) and she chose twenty examples of apologies. In choosing the apologies, she considered the variety of expressions and strategies and also the ease of explaining the context briefly in a way that students could easily understand. She also included examples that could be used to make specific points, such as one that used "I'm sorry" to express sympathy rather than an apology, one that used the apology in an ironic way, and one that did not have an Illocutionary Force Indicating Device (IFID) such as "I'm sorry" (Searle & Vanderveken, 1985, p. 2).

Students were given information about the characters in the series and the following instructions, along with the list of twenty apology interactions:

> The following are examples of apologies from the US family situation comedy *Modern Family*. Read the conversations and answer the following questions.
>
> 1. What is the expression used for the actual apology? (For example, "I apologize") Make a list of the common/useful expressions.
>
> 2. What other strategies are used in each conversation, that is, what else does the speaker try to do? (For example, if the speaker says, "I didn't mean to do it," they are saying that they hadn't intended to commit the offense.) Make a list of the strategies.

2.3. Classroom procedures

Students work together in groups to identify common expressions and strategies that were used in the examples and then the teacher goes through the list with the students, identifying which expressions and strategies are used in each example. Obviously, students describe the strategies in their own terms rather than using technical terms. Depending on the purpose of the class, the teacher can follow up by introducing students to a typology of apology strategies, with a lecture on theoretical aspects of apologies, with a discussion of the problems related to compiling a corpus, etc.

2.4. Examples

Below are examples of the apologies and an explanation of the strategies involved.

1. Whitney, a woman 11-year-old Manny met online, has come to meet him, believing he is an adult. She is talking to Manny's mother Gloria after learning he is only 11.

> Whitney: This is so humiliating. I am sorry.
>
> Gloria: It's okay.
>
> Whitney: He just seemed so mature online. How could I be so stupid?
>
> Manny: You're not stupid. "Stupid" is not following your heart and taking a chance on love.

Whitney expresses the emotion she is feeling ("This is so humiliating"), uses an IFID ("I am sorry"), explains why she made a mistake ("He seems so mature online"), and criticizes herself for making the mistake ("How could I be so stupid?").

2. Phil talks with an acquaintance about a problem his wife Claire has.

> Phil: But if she lets me help her, I can make her problem go away.
>
> Woman: Oh! That is such a male thing to say.
>
> Phil: Well, forgive me for being a man.

In this example, Phil is using an IFID ("Forgive me for…") and explains what he is apologizing for ("…for being a man"). In this case, students should recognize that the apology is being used ironically, since Phil has no control over the fact that he is a man.

3.   Discussion and conclusion

When using this activity in a small linguistic pragmatics class, one author found that students could identify most of the strategies in the apologies. In an informal survey of the students, they indicated that they liked the approach. One student wrote,

> "I think this is an effective way. By doing the exercise before listening to explaining, we can consider deeply, and it is important, I think. If we knew points before exercising, we only think the way that the teacher told [sic]".

A limit on the use of DDL for speech acts is that many speech acts are not easily searchable. However, it might be possible to use speech acts such as asking permission and expressing gratitude. More research is necessary on the use of DDL to teach speech acts as well as an exploration of which speech acts would work well with this approach.

References

Huang, H. (2004). Apologies in film: Implications for language teaching. *Paper presented at the 13th International Symposium on English Teaching, November 12-14, Chien Tan Overseas Youth Activity Center, Taipei*. Retrieved from http://personnel.sju.edu.tw/改善師資研究成果/92年度/黃馨週--編號89.doc

Johns, T., & King, P. (Eds.). (1991). *Classroom concordancing. English Language Research Journal, 4*. University of Birmingham: Centre for English Language Studies.

Kitao, S. K. (2012). Using a spoken corpus compiled from subtitles to study apologies. *Asphodel, 47,* 50-77.

Kitao, S. K., & Kitao, K. (2013). Apologies, apology strategies, and apology forms for non-apologies in a spoken corpus. *Journal of Culture and Information Science, 8*(2), 1-13.

Levitan, S., & Lloyd, C. (2009). *Modern Family* [Television Series]. Hollywood: American Broadcasting Company.

Rose, K. R. (2001). Compliments and compliment responses in film: Implications for pragmatics research and language teaching. *International Review of Applied Linguistics, 39*(4), 309-326. doi: 10.1515/iral.2001.007

Searle, J. R., & Vanderveken, D. (1985). *Foundations of illocutionary logic*. Cambridge: Cambridge University Press.

Tian, S. (2005). Data-driven learning: Do learning tasks and proficiency make a difference? *Proceedings of the 9th Conference of the Pan-Pacific Association of Applied Linguistics*. Retrieved from http://www.paaljapan.org/resources/proceedings/PAAL9/pdf/TianShiaup.pdf

20 Years of EUROCALL: Learning from the Past, Looking to the Future.
2013 EUROCALL Conference, Évora, Portugal, Proceedings
Edited by Linda Bradley and Sylvie Thouësny

The moral right of the authors has been asserted