Chapter XXIV

Corpora in the Classroom and Beyond

Rolf Kreyer
University of Bonn, Germany

ABSTRACT

Although corpus linguistic methods and research have had a considerable impact on language teaching in the last few decades, the corpus is still mostly regarded as a device in the hands of the teacher rather than the students. This is most probably due to a number of serious problems that need to be considered when allowing students unlimited access to the wealth of corpus-linguistic data. In this chapter it is argued that these problems are to be taken seriously and that a fundamental prerequisite for student use of corpora is what I call ‘corpus competence.’ Such a corpus competence will not only help students to make use of corpora in the classroom but will also prepare the grounds for use of corpora in non-institutionalized contexts and as a tool for life-long learning.

INTRODUCTION

In the early 60s, with the compilation of the BROWN corpus, modern corpus linguistics made its way into the field of linguistics and since then has been a firmly established methodology and discipline. The study of large amounts of authentic language data has led to new insights into the use of language and into our conception of the language system. These insights have also had a strong impact on language teaching. A number of articles, for instance, describe discrepancies between the language that is represented in textbooks and the language use that is attested in corpora: cases in point are the use of some and any in affirmative and negative contexts (Tesch, 1988), the use of relative pronouns in restrictive relative clauses (Kreyer, 2000), or the use of tense and aspect in conditional sentences (Römer, 2004a; Schäpers, 2005), to name but a few (for other areas of discrepancy consider Grabowski & Mindt, 1995; Römer, 2004b). Corpus-based find-
Corpora in the Classroom and Beyond

Corpora have also made their way into many textbooks and dictionaries: The German school grammar _Englische Grammatik Heute_ (Ungerer, 1999), for instance, emphasizes the fact that its descriptions of language use are (to some extent) corpus based (cf. Mukherjee, 2004) and the _Collins Cobuild English Dictionary_ aims at ‘helping learners with real English,’ that is, through authentic language data drawn from corpora. This shows that corpus linguistics is extremely influential with regard to linguistics in general and also with regard to language teaching, especially curriculum design and reference tools (see also Meunier, 2002).

The more recent past has seen attempts to make corpus-linguistic methods more directly applicable to language teaching. In this respect, two different kinds of suggestions are usually made: (1) the corpus as a tool in the hands of the teacher to prepare materials or as an aid in marking students’ writings, or (2) the corpus as a tool in the hands of the students, that is, students are encouraged to induce rules and regularities into the data that the corpus provides. Although the number of teachers that actually use corpora along the lines of (1) is still fairly small, recent studies indicate a trend towards an increasing use of corpora or corpus-based findings by teachers (see, for instance, Tsui, 2004, 2005). On the other hand, there still is “a need to convince practicing teachers to use corpora and concordances in the classroom” (Römer, 2006); teachers seem to be reluctant to give their students immediate access to corpora (Mukherjee, 2004), although there are many suggestions about how to ‘do corpus linguistics’ in the classroom (see Kreyer, 2007; Mukherjee, 2002; Rautenhaus, 1996; Tribble & Jones, 1997).

This reluctance is not unfounded. If students are confronted with raw corpus data, a number of severe problems arise, all of which have to be taken seriously. Doing corpus linguistics is a skill that has to be and, fortunately, can be trained (see Kirk, 2002; Bernardini, 2002 for some suggestions). Students need to be taught how to make use of the corpus, and they need to be made aware of its restrictions and shortcomings as well as its potential and advantages. In short, students need to be taught a set of skills that will allow them to exploit corpora on their own without the help of the teacher. Such a set of skills has been sketched out by Mukherjee (2002, pp. 179-180) under the term ‘corpus literacy’. In this chapter, I discuss each of these skills in detail, explore additional aspects and, finally, investigate the interdependencies between all of the relevant aspects. This extension of Mukherjee’s concept I will refer to as ‘corpus competence’.

Since to a large extent corpus competence means knowledge of how to deal with problematic aspects in working with corpora, I first give an overview of these problems. Then, I portray some of the benefits of language learning with the help of corpora. The remaining chapter describes aspects of corpus competence and makes suggestions as to how these can be taught at school or university. It is argued that corpus competence is indispensable if students want to work with raw corpus data and that through corpus competence the corpus can become a learning aid in non-institutionalized contexts and a tool for life-long learning.

**PROBLEMS WITH INTERPRETING RAW CORPUS DATA**

Working with a corpus is a very demanding activity for the researcher as well as the student of a foreign language. Most importantly, it is often far from trivial to interpret authentic corpus data. As Mauranen (2004) points out:

_Noticing things in corpus data is an acquired skill even for linguistically relatively sophisticated learners like L2 majors in university departments [...]. Less sophisticated students need even more tuition in making observations from corpus material. Initially, even making sense of a set of pre-edited concordance lines tends to be hard_