Chapter XXII
Concordancing 2.0: On Custom-Made Corpora in the Classroom

Jarosław Krajka
Warsaw School of Social Psychology, Poland

ABSTRACT

This chapter contrasts the use of corpora and concordancing in the Web 1.0 era with the opportunities presented to the language teachers by the Web 2.0 stand-alone concordancing software, which make it much easier to access, compile, and consult the corpora that are more relevant for particular classroom contexts. It is argued here that once trained in the basic corpus consultation procedures with demo interfaces, teachers can exercise their autonomy by using texts available locally and globally to compile custom-made collections. In the chapter the two basic approaches to custom-made concordancing, namely the Web as Corpus and the compilation of ad-hoc collections will be described, together with a summary of sample tools. It is hoped that given careful selection of relevant sources, the learning process will become significantly enhanced thanks to more authentic and relevant language data, promoting teacher autonomy and discovery-based procedures.

INTRODUCTION

Corpora and concordancing have been widely used in ELT by materials writers and lexicographers, also to assist teachers in making informed choices about curriculum development, vocabulary selection and lexical testing. However, these tools were not of wide use by teachers in the pre-Web 2.0 stage, mainly due to lack of proper information and training, largely limited access, as well as lack of full relevance for some contexts (e.g., LSP).

With a computer being a standard tool of a contemporary language teacher, together with greatly facilitated Internet access and much higher bandwidth, the reflection on the incorporation of concordancing procedures by language teachers in materials development or vocabulary selection seems to be of prime importance. The possible impact of hands-on concordancing on teacher autonomy (and, in effect, learner autonomy), resulting in much greater awareness of the teaching
content and judicious use of coursebook materials, is another factor that calls for the wider implementation of corpus linguistics procedures, on various planes and in different respects, in actual teacher training. Concordancing 2.0 takes as its primary characteristics teacher independence in selecting materials for a corpus, more effective information retrieval, setting criteria for corpus compilation or choosing browsing tools for text analysis.

The aim of the present chapter is to expose the shift from the use of ready-made online corpora, often limited in size, scope and functionality due to their demonstrative nature, which is tentatively termed “Concordancing 1.0,” to the unrestricted selection and compilation of corpora, as well as analysis with the use of widely accessible text analysis tools (“Concordancing 2.0”). The chapter will address the issues involved in implementing tailor-made corpora in language learning, as regards the process of corpus compilation, the browsing procedures and the corpus-based teaching activities. After establishing such introductory notions as definitions, arguments for and against in-class concordancing and sample corpus-based language learning activities, the present chapter will analyse two main approaches to preparing custom-made corpora – the Web as Corpus and using document files. A description of concordancing tools occupies the most prominent place in the chapter, as well as a discussion of the process of corpus compilation. The chapter also aims to investigate the in-class implementation procedure, proposing a multi-staged training process reflecting teachers’ and learners’ growing independence in concordancing.

BACKGROUND: OPPORTUNITIES AND DRAWBACKS OF IN-CLASS CORPUS CONSULTATION PROCEDURES

There are numerous studies reporting the investigation of the effectiveness of corpus-based procedures in foreign language instruction. These range from the use of small corpora tailored to students’ needs (Aston, 1997) to promoting large corpus concordancing (Bernardini, 2000; de Schryver, 2002); improving writing performance at lower (Yoon & Hirvela, 2004; Gaskell & Cobb, 2004) and advanced levels (Chambers & O’Sullivan, 2004); grammar presentation (Hadley, 2002) and rule inferencing (St. John, 2001). An extensive body of research can be, quite naturally, found in the area of vocabulary acquisition (Cobb, 1997; Cobb, 1998) and teaching foreign language reading, not only assisted by concordancers themselves, but performed in the wider context of a resource-assisted environment, encompassing for instance concordance, dictionary, cloze-builder, hypertext, and a database with the interactive self-quizzing feature (Cobb et al., 2001; Horst et al., 2005). Some studies reported on the relation between the effectiveness of corpus-consultation procedures and strategy training (Kennedy & Miceli, 2001; St. John, 2001; Chambers, 2005), indicating the need to reflect on the conscious and gradual introduction of the tool in the classroom. The perspective that is most relevant for the purposes of the present chapter is represented by the increase of writing proficiency due to learner corpus self-compilation (Lee & Swales, 2006).

In introduction, some space should be devoted to the definition of a corpus. Crystal (1991) defines it as a collection of linguistic data, either written texts or a transcription of recorded speech, which can be used as a starting-point of linguistic description or as a means of verifying hypotheses about a language. In a similar vein, Sinclair (1991) adds that corpora are made of naturally occurring language, while Krishnamurthy (2001) points out the genuine communicative situations that are recorded without any editing to create corpus contexts. McEnery and Wilson (1996) enumerate four crucial characteristics of a corpus, namely sampling and representativeness, finite and fixed size, machine-readability and standard reference. Four other criteria indispensable for a body of