The role of a corpus of spoken language in the research on anglicisms

Marcin Zabawa (University of Silesia, Katowice)

The aim of the present paper will be to discuss the role of a corpus in the study of anglicisms in Polish. It would seem that the lexical influence of English upon written Polish, particularly the language of the press, has been researched relatively thoroughly; the influence on spoken informal Polish, however, has attracted much less attention of linguists, as gathering a sufficient amount of spontaneous spoken language is inherently a very difficult and time-consuming process. This tendency still continues, as much attention is devoted to new vocabulary items appearing in written Polish, especially the one used in the mass media, whereas the area of spoken language is generally neglected.

It seems that the most objective method of researching anglicisms in spoken informal language is the use of the corpus. The author will argue that the best solution is to gather the corpus on one's own, rather than use an already existing corpus. Unfortunately, the entire process is difficult, to say the least. However, the author is not going to discuss general issues related to the process of collecting a spoken corpus, as they have already been discussed in the literature on sociolinguistics. Rather, the speaker will concentrate on the problems connected with collecting data for the examination of anglicisms, such as the choice of informants and the choice of material. Sadly, it seems that the use of the corpus for the research on anglicisms has quite a few weaknesses (e.g. a scarcity of borrowed words) and the speaker will not hesitate to discuss them. The author will also compare the use of the corpus in the study of anglicisms with other methods of gathering the data, such as guided conversations, questionnaires, etc. It is hoped that the speaker will manage to argue that the use of the specially designed corpus, despite all the difficulties involved, is by far the best method for researching anglicisms in spoken language.

The second part of the paper will be devoted to the study of English lexical and semantic loans (carried out by the present author) based on the corpus of spontaneous spoken language. The corpus in question comprises 20 conversations (ca 61,000 words altogether). Special attention will be devoted to statistical analysis which will not be available without a corpus, such as the frequency of borrowings in spoken language, TTR, the classification of borrowed words (according to the part of speech, the degree of assimilation, the semantic field, etc.) and the relation between the usage of a given loan and other criteria (such as the topic of the conversation, the sex and age of informants, etc.).