

УДК 81'27

## О СПОСОБАХ СОЦИОЛИНГВИСТИЧЕСКОЙ БАЛАНСИРОВКИ УСТНОГО КОРПУСА (на примере «Одного речевого дня») <sup>1</sup>

**Екатерина Михайловна Баева**

к. филол. н., доцент кафедры английского языка

Санкт-Петербургский государственный университет

199034, Санкт-Петербург, Университетская наб., 11. kaverita@yandex.ru

Статья посвящена описанию различных способов социолингвистической балансировки, которые используются при составлении устных корпусов того или иного языка. Дается обзор социолингвистических принципов, лежащих в основе формирования некоторых западных и российских корпусов, а также ставится проблема социолингвистической несбалансированности устного корпусного материала, с которой часто сталкиваются лингвисты, изучающие языковое существование современного города. Основное внимание в статье сосредоточено на создающемся в Санкт-Петербурге звуковом корпусе «Один речевой день». Для достижения максимальной сбалансированности и представительности материала при формировании данного корпуса авторами была разработана собственная методика. На основе данных Федеральной службы городской статистики и общих тенденций лингвистического описания современного города был составлена анкета для информантов, которая в определенной степени стала ориентиром для подбора кандидатов, осуществляемого исследователями в частном порядке. В данной анкете отражены такие социальные факторы, как пол, возраст, место рождения и наиболее длительного проживания информантов, их прошлые и нынешние профессии, характер и уровень образования, родной язык, другие языки, которыми владеют информанты, а также место рождения и социальное происхождение их родителей. Значимость каждого фактора объясняется и обсуждается в статье отдельно. Описываются сбалансированные группы и подгруппы, на которые потенциально делится материал данного устного корпуса.

**Ключевые слова:** социолингвистика; социолингвистическая балансировка; устный корпус; лингвистическое описание; методика корпусного исследования; звуковой корпус «Один речевой день».

### ON MEANS OF SOCIOLINGUISTIC BALANCING OF A SPOKEN CORPUS (drawn on “One Speaker’s Day”)

**Ekaterina M. Baeva**

Associate Professor in the Department of English Language

Saint Petersburg State University

The article considers different means of sociolinguistic balancing involved in forming spoken corpora. It gives a review of sociolinguistic principles which underlie some of the foreign and Russian corpora (for instance, the British National Corpus and the Russian National Corpus) and states the problem of unbalanced material, which linguists often encounter. The main object of the study is the sound corpus “One Speaker’s Day”, being recorded in Saint Petersburg, and the methods of its balancing and forming used by its developers. It is the second part of the corpus, the first being an annotated collection of texts which were analyzed by the means of binary oppositions. It has not always proved suitable for the corpus. In order to obtain sociolinguistically balanced material for research, the authors designed a special questionnaire based on the data of Federal Agency of Urban Statistics and general trends of linguistic description of a city. This questionnaire was used to select the candidates to participate in recording for the project. It includes such important sociolinguistic characteristics as gender, age, place of birth and places of living of the informants, their former

and current occupations, their profession and education, their native language and other languages they speak, as well as their parents' place of birth and social origins. The importance of each factor is described and justified in this study. Using certain factors, for example gender and native language, it is possible to divide the informants of a corpus research in two well-balanced groups, but others, such as age and occupation, demand a more precise classification. Therefore, the research aims to categorize the material of "One Speaker's Day" by possible balanced groups and subgroups and states the necessity for a detailed analysis. It suggests that certain problems of balancing a corpus mentioned in the review can be avoided using this method.

**Key words:** sociolinguistics; sociolinguistic balancing; spoken corpus; linguistic description; sound-ing corpus; One Speaker's Day; method of corpus studies.