

The Typology of Electronic Texts for a Corpus to be Representative

Svetlana Savchuk¹

Abstract

The paper regards the electronic communication from the point of view of text typology. This problem has rather practical, than theoretical significance, taking into consideration the problems of the text selection for a representative corpus and of their annotation. But to solve this practical problem we need to solve a set of theoretical ones. The main theoretical obstacle we ought to surmount is the internet genres specification.

As far as this question is concerned, we may tell the difference between two groups of texts. On the one hand, in Internet we deal with the genres identical to the traditional, so to speak “paper” genres, which function in the other field of communication: scientific, official, advertisement, religious, journalistic, fiction (web-texts from e-libraries, scientific and educational portals; news sites, etc.). These texts and these genres don’t differ from the traditional texts and genres, or, to be precise, the only difference between them concerns the method of their broadcast (electronic or non-electronic).

On the other hand, in Internet we have a lot of genres (usually they are used in everyday online-communication), in which the features of non-formal spontaneous interaction and of public communication combine (these are chats, BBS, forums, conferences, e-mail letters, blogs, etc.). The complex interaction of different factors has an influence on the linguistic characteristics of these texts; first of all, we mean interesting combination of the written and the spoken features in these these texts.

The paper describes typological features of the specific genres of Russian sector of the Internet (RUNET), which are to be included in the Russian National Corpus and to form the sub-corpus of electronic communication.

¹ *e-mail*: savsvetlana@mail.ru