How to get closer to native Czech? Using corpora in language learning

Have you ever thought in a foreign language? Have you ever wished to reach a native-like level? The fundamental approach to acquire this faculty is exposure to the language we would like to learn. But what if a native environment is beyond our reach? How could we compensate for this unfavourable situation? Fortunately, many modern languages have to their disposal language corpora. A language corpus is an enormous set of electronic data consisting of authentic texts in various genres (fiction, no-fiction, press, private conversations, broadcasts, etc.). It makes possible for its users to look into a living language and search for not only single words but also phrases, part of speeches, letter combinations, grammatical cases, etc. Furthermore, the corpus provides many examples of language usage. It may help learners of Czech to be up to date with new trends. Based on corpus data, it is possible to learn actual Czech and be in contact with developing language. Throughout many corpus tools (e.g. SyD, Treq, Word in Context) and corpus-based exercises, learners have the opportunity to be exposed to native Czech and improve their skills. Moreover, this approach forces learners to develop quite overlooked skills of thinking. They have to first analyse a single pattern to classify it and, in the end, generalize linguistic rules. Thinking is the key element of language awareness that helps to understand language structure and makes a non-native speaker more independent. Using a language corpus reinforces inductive learning (and learner's thinking) and organizes knowledge about the acquired language.

fot. David Semerád, http://www.dsfotoportrety.cz/

Adrian Jan Zasina – holds an MA in Slavic Studies and a PhD in Mathematical Linguistics (Corpus linguistics branch). He is currently an assistant professor at the Institute of the Czech National Corpus, Charles University, Prague. His main research interests include data-driven learning, corpus linguistics, parallel corpora, the grammar of Czech and Polish, language variability, and corpus-assisted discourse studies.