

According to the title, the thesis is composed of three parts. At the beginning the author introduces the concept of a parallel corpus defining it as a set of texts in different (at least two) languages, composed of original-translation couples. A terminology is provided to name different sets of texts in different languages. To have a general overview of this specific field, readers can be acquainted with the present existing parallel corpus. A definition of the project for creating a bidirectional French-Czech Czech-French parallel corpus is given in order to use it in linguistic research, notably research upon inflected expressions. French and Czech texts composing the parallel corpus Fratchque come from literature; the author also gives the explanation regarding the reason why other kinds of texts have not been taken into account. This corpus, conceived for PC-based researches, exists only in digital format. Having in mind the purpose of representing modern language, only texts after the year 1945 have been selected. New couples of French-Czech Czech-French texts could be easily added thanks to the files structure stored on a hard disk and managed by ParaConc. The corpus is not marked up explicitly by XML tags because the tagging is not necessary for the proper functioning of ParaConc – this step, considered as important, is postponed.

The selection principles are presented as a list in relevant order. The impossibility to create a representative corpus (according to sophisticated sociolinguistic criteria) by only one person is compensated by the intent to offer a greater variety of possible choices. This should minimize the risk that a particular style of an author or a translator could influence different aspects of the language represented in the corpus. ...