

V této disertační práci zkoumáme strojový překlad mezi češtinou a ruštinou z hlediska lingvisty. Pracujeme s několika pravidlovými a statistickými překladovými systémy a pomocí změn v jejich nastavení se snažíme dosáhnout co nejlepších výsledků překladu. Jedna z otázek, které řešíme v naší práci, je nakolik příbuznost obou jazyků pomáhá strojovému překladu.

Hlavním cílem práce je lingvistický rozbor chyb ve výstupu čtyř systémů strojového překladu, dvou experimentálních - TectoMT, Moses, a dvou komerčních - PC Translator a Google Translate. Analyzujeme každý typ chyb a řešíme, zda daná chyba souvisí s rozdílem mezi češtinou a ruštinou nebo zda je zapříčiněná architekturou jednotlivých systémů. Pro některé chyby nabízíme cesty, jak je opravit.

Ve zvláštní kapitole se zaměřujeme na chyby v povrchové valenci sloves. Zkoumáme rozdíly v české a ruské povrchové valenci, popisujeme extrakci slovníku povrchových forem a tento slovník integrujeme do systému TectoMT. Dále nabízíme souhrn lingvistických pozorování o povaze rozdílů v české a ruské valenci.