

Tato bakalářská práce se zaměřuje na automatické skloňování českých podstatných jmen, zejména slov, která nejsou zahrnuta ve slovníku (tzv. out-of-vocabulary, OOV) - skloňování předem neviděných slov. Automaticky jsme extrahovali rozsáhlý dataset vhodný pro trénování a vyhodnocení za OOV podmínek. Dále jsme manuálně vytvořili dataset vyskloňovaných reálných OOV slov - neologismů. Vyvinuli jsme tři různé systémy: retrográdní model založený na algoritmu k-nejbližších sousedů (kNN) a dva modely sequence-to-sequence (seq2seq) založené na LSTM a Transformeru. V porovnání se stávajícím skloňovacím systémem sklonuj.cz a standardními baseline systémy ze SIGMORPHON shared tasks jsme za OOV podmínek s naším seq2seq modelem dosáhli nejlepších výsledků. Dále jsme dosáhli state-of-the-art výsledků pro 6 ze 16 vývojových (development) jazyků zahrnutých v datech SIGMORPHON 2022 shared task, při vyhodnocení za OOV podmínek (feature overlap condition) a při trénování na větším datasetu (large data condition). Na datasetu reálných OOV slov byly všechny neuronové modely překonány retrográdním modelem, který dosahuje výsledků srovnatelných s neneuronovým SIGMORPHON baseline systémem. Představujeme a zveřejňujeme skloňovací systém založený na našem seq2seq modelu jako jednoduchou pythoní knihovnu. Tento systém by mohl sloužit jako doplněk k nejlepšímu slovníkovému systému MorphoDiTa jako záloha pro OOV slova, zejména až bude rozšířen na další slovní druhy.