

Tato práce se zabývá několika kombinovanými metodami morfologického zpracování tureckých jazyků, zejména turečtiny. Součástí našich snah bylo i obstarání větších zdrojů jazykových dat, než jaké jsou v současnosti k dispozici, a jejich zpřístupnění veřejnosti. Počítačové zpracování turečtiny zahrnuje specifickou sadu problémů spojených zejména s vysoce produktivní, aglutinační morfologií. Rozsah veřejně dostupných dat je s ohledem na čistě statistické metody nedostatečný a pro účely strojového učení jsou tato data příliš řídká. Z tohoto důvodu vyhodnocujeme veřejně dostupný morfologický analyzátor TRmorph, založený na konečných převodnicích, tedy na pravidlech. Snažíme se rozšířit záběr a slovník tohoto analyzátoru; kombinujeme statistické metody s heuristikami pro rozpoznávání pojmenovaných entit (a konstrukci zeměpisných slovníků), zjednoznačnění morfologické analýzy a zpracování víceslovných výrazů. Výsledky dosavadních experimentů s heuristickými přístupy ukazují slibné rozšíření pokrytí textu TRmorphem. Statistické metody používáme jako záložní řešení pro jemnější úlohy, které nelze snadno zachytit heuristickými pravidly. Tímto způsobem náš hybridní systém rozšiřuje uplatnění morfologického analyzátoru, jenž je sám postaven čistě na pravidlech.