

Tato studie opakuje výzkum Linlin Sun & Davida C. Saavedry zaměřený na určování gramatického statusu jednotek v čínštině pomocí kvantitativních metod. V návaznosti na jejich práci používáme binární logistický regresní model pro výpočet skóre gramatického statusu vybraných lexikálních jednotek. Jako zdroj dat nám slouží Lancasterský korpus čínštiny (LCMC), který obsahuje současné standardní čínské texty.

V práci podrobně popisujeme metriky a modelovací přístup použitý k odvození skóre gramatického stavu a porovnáваме naše výsledky s výsledky autorů původní studie.

V naší analýze dále posuzujeme vhodnost vybraných metrik, hodnotíme přesnost predikce binárního logistického regresního modelu a analyzujeme, kde se po přiřazení gramatického statusu jednotky nachází na lexikálně - gramatické škále, a zda tomu odpovídá jejich „tradiční“ zařazení do slovních kategorií.

Klíčová slova: gramatický status, korpus, čínština, kvantitativní lingvistika, replikační studie