



منظور از خطایابی نحوی تشخیص و اصلاح خطاهای متن است که از اشتباه در دستور زبان ناشی می‌شوند. سامانه‌های خطایابی نحوی می‌توانند به افرادی که در حال یادگیری یک زبان جدید هستند، در یادگیری قواعد زبان کمک کنند. برای بررسی و شناسایی خطاهایی که در میان فارسی‌آموزان رایج است، برگه‌های نگارش زبان‌آموزان مؤسسه لغت‌نامه دهخدا و مرکز آموزش زبان فارسی دانشگاه بین‌المللی امام خمینی قزوین جمع‌آوری و خطاهای آن به صورت دستی برچسب زده‌شد. با استفاده از مجموعه‌ی حاصل، «پیکره خطاهای فارسی‌آموزان^۱» را که شامل بیش از ۴۷۰۰ خطا از انواع مختلف است، ساختیم. مشاهده شد که مطابقت فعل و فاعل جزء خطاهای پرتکرار در میان فارسی‌آموزان می‌باشد. در این پژوهش یک روش داده‌محور برای اصلاح خطای تطابق فعل و فاعل در زبان فارسی ارائه می‌شود. در زبان فارسی بین فعل و فاعل از دو جهت تطابق وجود دارد: شخص و تعداد. برای تصحیح جمع یا مفرد بودن فعل، پس از تعریف ویژگی‌های مناسب، دو طبقه‌بند یادگیری ماشین بی‌نظمی بیشینه و درخت تصمیم را به کار می‌بریم. برای تصحیح فعل از نظر شخص {اول شخص، دوم شخص، سوم شخص} از روش مبتنی بر قاعده استفاده می‌کنیم. در نهایت برتری روش پیشنهادی خود را نسبت به روش پایه برچسب اکثریت و روش‌های رقیب ماشین ترجمه آماری و مبتنی بر شبکه عصبی نشان می‌دهیم.

^۱آدرس:

<http://ece.ut.ac.ir/documents/۷۶۶۸۷۴۱۱/۰/CPG.zip>

خطایابی نحوی، تشخیص و اصلاح خطاهای نحوی زبان فارسی، تشخیص و اصلاح خطای تطابق فعل و فاعل، استفاده از طبقه‌بندهای یادگیری ماشین برای خطایابی، طبقه‌بند بی‌نظمی بیشینه، درخت تصمیم

correcting verb grammatical errors, grammatical error detection and correction, machine learning classifiers, decision tree, Persian grammatical error correction