

چکیده:

در این پژوهش، برای اولین بار مدلی ترکیبی برای تراز بندی جملات، جهت ساخت پیکره‌های موازی انگلیسی-فارسی ارائه شده است. در حالت کلی چارچوب روش پیشنهادی، غیر وابسته به زبان‌های مبدأ و مقصد بوده و از آن می‌توان برای تولید پیکره‌های موازی، برای هر جفت زبان دیگر، نیز استفاده کرد. نتایج بدست آمده از پیاده‌سازی‌ها نشان می‌دهد که بکار بردن ویژگی‌های زبانی و غیر زبانی، عملکرد سیستم را تا حد قابل قبولی بهبود خواهد بخشید. همچنین در این تحقیق میزان تأثیر استفاده از شباهت‌های طولی، ترجمه تحت‌اللفظی و نقش دستوری کلمات، به صورت مستقل یا ترکیبی، برای عملیات تراز بندی مورد بررسی قرار گرفته است. بکارگیری طبقه بندهای سلسله‌مراتبی در تشخیص نوع تراز بندی به عنوان یکی از شاخص‌های اصلی سیستم پیشنهادی محسوب می‌شود که موجب بالا رفتن دقت و سرعت عملیات تولید پیکره نسبت به سایر مدل‌ها می‌شود. از ویژگی‌های دیگر این روش، قابلیت توسعه پذیری آن است. به این ترتیب که می‌توان با گنجاندن خصوصیات (در بدنه بردارهای ورودی) که ممکن است در آینده برای تشخیص بهتر نوع تراز بندی مورد توجه قرار گیرند کیفیت سیستم تراز بندی را ارتقاء بخشید. با این حال، چالش اساسی این رهیافت و بسیاری از روش‌های پیشین، وجود سلايق متنوع در ترجمه متون است. این مسئله در برخی موارد موجب تولید جملاتی می‌گردد که مشابهت آن‌ها با متن اصلی در حد معنا بوده و تنها قابل درک و تشخیص برای انسان می‌باشد. در چنین وضعیتی کار استخراج جفت عبارات معادل، بسیار دشوار است. علاوه بر این در مورد متونی که مقید به رعایت ساختار دستوری زبان فارسی و انگلیسی نمی‌باشند عملیات تشخیص جملات هم تراز ممکن است به خوبی انجام نپذیرد.

کلمات کلیدی:

تراز بندی جملات، پیکره موازی، ترجمه آماری