

فردوس-لکس: پیکره واژگانی از منابع آموزشی فارسی برای غیرفارسی‌زبانان

مقدمه: مطالعات بسیاری بر نقش پیکره‌های آموزشی به خصوص در ارتباط با آموزش واژگان، تأکید داشته‌اند (Chen, 2011; Li, 2017)، اما پژوهش‌های پیکره‌بنیاد محدودی در ارتباط با واژگان در زمینه آموزش زبان فارسی به غیرفارسی‌زبانان انجام شده است و در هیچ‌کدام نیز به سطوح زبان‌آموزی توجهی نشده است. بنابراین، هدف از انجام این پژوهش، تدوین پیکره‌ای سطح‌بندی‌شده از واژگان براساس منابع به‌روز آموزشی است که در مراکز آموزش فارسی به غیرفارسی‌زبانان (آزفا) کاربرد دارند. بدین ترتیب، تدوین این پیکره براساس هم‌پوشانی‌ای که بین منابع مختلف در ارتباط با واژگان در سطوح مختلف آموزشی (پایه، مقدماتی، پیشرفته) وجود دارد، مطابق با اصول و مبانی ساخت پیکره صورت پذیرفته است.

روش‌شناسی: پژوهش حاضر که نوعی تحلیل محتوای کیفی با رویکرد پیکره‌آگاه محسوب می‌شود، با کمک ابزارهای رایانشی و برنامه‌نویسی پایتون در سه گام صورت پذیرفته است. گام اول آماده‌سازی داده‌گان شامل تایپ کردن ۱۴۵۸۵ واحد از سه سطح بود که از ۲۶ منبع آموزشی شامل مجموعه کتاب‌های پرفا، مینا، شیراز، پارسا، رسا، نگارا، خوانا و آموزش نوین زبان فارسی انتخاب شده بودند. سپس، پیش‌پردازش‌های رایانشی و افزودن برچسب‌زنی اجزاء کلام صورت پذیرفت که برای این منظور از ابزار دادماتولز (اعتزادی و همکاران، ۲۰۲۲) و ابزار استنزا (چی و همکاران، ۲۰۲۰) استفاده گردید. برای اطمینان از دقت برچسب‌زنی، ۵ درصد از واژگان نیمه اول هر سطح انتخاب شده و به صورت دستی بازنگری شدند. بدین ترتیب، خطاها اکثر در افعال مرکب (چندجزیی) مشاهده شدند که تعدادی از آن‌ها به اشتباه اسم تلقی شده بودند و تعدادی دیگر حرف اضافه بودند. در گام بعدی، با کمک برنامه‌نویسی پایتون، هم‌پوشانی واژگان در هر سطح و بین سطوح مقایسه گردید. در نهایت، پیکره پژوهش توسط برنامه‌نویسی پایتون به استاندارد XML نشانه‌گذاری شده و ماشین‌خوان گردید. این پیکره سه سطح مبتدی، میانه و پیشرفته دارد و شامل اطلاعاتی پیرامون واژگان است مانند حرف الفبا، واحد، بن‌واژه، دامنه هم‌پوشانی، برچسب جزء کلام و فراداده کتاب‌های منبع مربوط به واژگان.

یافته‌های اصلی: نتایج نشان داد نسبت نوع واژه به واحد از سطح مبتدی به میانه روند کاهشی و سپس از میانه به پیشرفته روند افزایشی داشته (در سطح مبتدی ۵۷٫۵، در سطح میانه ۴۷٫۹ و در سطح پیشرفته ۵۹٫۵ درصد) که این نتیجه احتمالاً بر تکرار بیشتر واژگان در سطح میانه نسبت به سایر سطوح دلالت دارد. علاوه‌براین، میزان نسبتاً اندک هم‌پوشانی واژگان در هر سطح و کاهش شدید آن در سطح پیشرفته (مبتدی حدود ۳۶، میانه حدود ۳۶٫۵ و پیشرفته حدود ۱۳ درصد) نشان می‌دهد که مؤلفان نسبت به تکرار و یادآوری واژگان، توجه ویژه‌ای مبذول نداشته‌اند. همچنین برخلاف نظر Cervetti et al. (2015) در تمامی سطوح، اسم غالب بوده است و فعل و صفت به ترتیب در رتبه‌های بعدی قرار

دارند. در سطح میانی، برخلاف انتظار، مقوله اسم در دامنه‌های بالای هم‌پوشانی غالب بود لذا سطح میانه به تکرار کمتر سایر مقوله‌ها پرداخته است. میانگین طول کلمات نیز در هر سطح روند افزایشی داشته (در سطح مبتدی ۴,۷۹، در سطح میانه ۵,۲۵ و در سطح پیشرفته ۵,۵۰ حرف) که می‌تواند دال بر کاربست واژگان از ساده به پیچیده در سطوح مختلف باشد.

بحث و نتیجه‌گیری: باتوجه به اهمیت نقش واژه‌ها در پیشبرد مهارت‌های زبانی و نقش مهم پیکره در آموزش زبان و یادگیری واژگان، در این پژوهش به تدوین پیکره‌ای از واژگان در سه سطح مبتدی، میانه و پیشرفته پرداخته شد. با توجه به یافته‌ها، برای بهبود تألیف منابع آموزشی پیشنهاد می‌شود که مؤلفان به افزایش تکرار واژگان و افزایش کاربست صفت و فعل در سطوح مختلف و افزایش تنوع افعال در سطح میانی توجه بیشتری کنند. همچنین از واژگان تخصصی‌تر در سطح پیشرفته به گونه‌ای استفاده گردد که امکان تکرار آن‌ها فراهم شود. پیشنهاد می‌شود پژوهش‌های آتی، متن کامل کتاب‌ها و همچنین صورت غیررسمی کلمات آنان را نیز در پیکره لحاظ کنند. همچنین، توجه به سایر مؤلفه‌هایی که در نظریه انتخاب واژگان (Nagy & Hiebert, 2011) و صورت بسط-یافته آن که توسط Cervetti et al. (2015) مطرح شد، از اهمیت بسیاری برخوردار است. این امر می‌تواند مؤلفان منابع آموزش زبان فارسی به غیرفارسی‌زبانان را به سوی معیاری سوق دهد که مطابق با آن به دسته‌بندی و گزینش صحیح واژه‌ها پردازند. در پایان، پیشنهاد می‌شود که پیکره پژوهش کنونی با داده‌های بیشتر از منابع متنوع‌تری گسترش یابد و همچنین نرم‌افزارهای پردازش واژگان فارسی توسعه داده شود تا بتوان با دقت بیشتری به تحلیل این زمینه پرداخت.