

پژوهش‌نامه‌ی آموزش زبان فارسی به غیرفارسی‌زبانان

(علمی-پژوهشی)

سال پنجم، شماره‌ی دوم (پیاپی ۱۲)، پاییز و زمستان ۱۳۹۵

واژه در کتاب‌های آموزش زبان فارسی به غیرفارسی‌زبانان:

پژوهشی پیکره‌بنیاد

کیومرث جهانگردی

دانشجوی دکتری زبان‌شناسی- پژوهشگاه علوم انسانی و مطالعات فرهنگی

دکتر مصطفی عاصی

استاد زبان‌شناسی- پژوهشگاه علوم انسانی و مطالعات فرهنگی

دکتر آریتا افراسی

دانشیار زبان‌شناسی- پژوهشگاه علوم انسانی و مطالعات فرهنگی

دکتر امیررضا وکیلی‌فرد

استادیار آموزش‌کاوی زبان فارسی- دانشگاه بین‌المللی امام خمینی^(۱)

چکیده

پژوهش حاضر با استفاده از روش‌ها و ابزارهای زبان‌شناسی پیکره‌ای، به سنجش میزان هم‌پوشانی و انطباق واژه‌های ارائه شده در کتاب‌های آموزش زبان فارسی به غیرفارسی‌زبانان، با پُربسامدترین واژه‌های زبان فارسی، می‌پردازد. بدین منظور، براساس متن‌های موجود در پایگاه داده‌های زبان فارسی، یک پیکره‌ی زبانی متوازن طراحی شد و پُربسامدترین واژه‌های آن به‌عنوان مبنای کار قرار گرفت. همچنین، سه عنوان از کتاب‌های پرکاربرد در مراکز آموزش زبان فارسی به غیرفارسی‌زبانان، انتخاب شدند. بر همین اساس، ۵۰۰ واژه‌ی پُربسامد هر یک از کتاب‌های مذکور، به پنج گروه هزارواژه‌ای دسته‌بندی شد و هر یک از این گروه‌ها به مثابه‌ی یکی از سطوح زبان‌آموزی تلقی گردید. در نهایت، گروه‌های پنج‌تایی فهرست واژگانی هر کدام از کتاب‌ها با گروه‌های متناظر با آنها در فهرست واژگانی پیکره‌ی زبانی مرجع/مبنا مورد مقایسه قرار گرفت. نتایج و یافته‌های پژوهش نشان داد که به لحاظ سطوح زبان‌آموزی، میزان هم‌پوشانی واژگانی هر یک از پنج گروه هزارواژه‌ای برگرفته از کتاب‌ها با گروه‌های متناظر آنها در پیکره‌ی مبنا بسیار پایین است؛ حتی در گروه اول که انتظار می‌رود به دلیل حضور تعداد زیادی از واژه‌های نقشی در آن، هم‌پوشانی واژگانی به سمت ۱۰۰٪ میل کند، در بالاترین میزان هم‌پوشانی (یعنی در کتاب رسولی/جامعه‌الزهره) این هم‌پوشانی حدود ۵۵٪ بود. در سایر سطوح واژه‌آموزی، مثلاً گروه پنجم، میزان هم‌پوشانی واژگانی با پیکره‌ی مبنا، حتی به زیر ۲٪ نیز تقلیل یافت. در نهایت، یافته‌های پژوهش، ما را به این جمع‌بندی رهنمون ساخت که کتاب‌های آموزش زبان فارسی به غیرفارسی‌زبانان در بعضی از سطوح زبان‌آموزی حتی تا ۲٪ نیز با پیکره‌ی مورد نظر این پژوهش هم‌پوشانی ندارند. همچنین، یافته‌ها نشان داد، چنانچه موضوع سطوح زبان‌آموزی را نیز کنار بگذاریم و صرفاً به پوشش پُربسامدترین واژه‌ها در کتاب‌ها بپردازیم، واژه‌های ارائه شده در کتاب‌ها کمتر از ۵۰٪ از هم‌پوشانی با پیکره‌ی مورد نظر این پژوهش برخوردار هستند.

کلیدواژه‌ها: واژه، پیکره‌ی زبانی، هم‌پوشانی واژگانی، منابع آموزش زبان فارسی

۱. مقدمه

هرچند زبان‌شناسی پیکره‌ای از آغاز پیوندی قوی با آموزش زبان داشته است و در این رابطه تأثیر سینکدر بر فرهنگ‌نگاری و کار پیشرو او در پژوهش‌های پیکره‌بنیاد (سینکدر، ۱۹۸۷، ۱۹۹۱ و ۲۰۰۴) را می‌توان نقطه‌ی شروع بسیاری از رویکردهای پیکره‌بنیاد در آموزش زبان به شمار آورد، سینکدر (۲۰۰۴: ۱) درباره‌ی تاریخچه‌ی زبان‌شناسی پیکره‌ای می‌نویسد: «پژوهشگران زبان و متخصصان آموزش زبان با آغوش باز از گنجینه‌ی سرشار داده‌های پیکره‌ای استقبال نکردند و تنها در چند سال اخیر افرادی که خود را زبان‌شناس کاربردی^۱ می‌دانند، پیکره‌های زبانی الکترونیکی را مورد توجه جدی قرار داده‌اند». بنابراین، پیکره‌های زبانی، تبدیل به بخشی از چشم‌انداز آموزش زبان شده‌اند و این حاصل تعامل آشکار بین آموزش زبان و زبان‌شناسی پیکره‌ای است.

در خصوص اهمیت پیکره‌های زبانی، آیمر (۲۰۰۹: ۱) می‌نویسد: «پیکره‌ها نگاه ما به زبان و کاربرد زبان را تغییر داده‌اند، ما این توقع را پیدا کرده‌ایم که از پیکره‌ها به‌عنوان خوراک کلاس‌های آموزش زبان استفاده کنیم و صرفاً پیکره‌های زبانی خام نیستند که مورد علاقه‌ی مدرسان قرار گرفته‌اند». همچنین، فاکچینی (۲۰۰۷) به اهمیت و ارزش پیکره‌های زبانی در فراگیری زبان دوم اشاره می‌کند.

از سوی دیگر، می‌توان ملاحظه کرد که در مراکز آموزش زبان فارسی به غیرفارسی‌زبانان، تعدادی واژه برای آموزش در سطوح مختلف زبان‌آموزی هدف‌گذاری می‌شود. این هدف‌گذاری بر حسب تجربه‌ی سال‌های قبل انجام می‌گیرد. پرسشی که در اینجا مطرح می‌شود این است که آیا تجربه می‌تواند برای انتخاب این واژه‌ها کافی باشد؟ چگونه می‌توان این انتخاب‌های تجربی را محک زد و راستی‌آزمایی نمود؟ یکی از راه‌های راستی‌آزمایی، استفاده از پیکره‌های واژگانی است. به عبارتی دیگر، می‌توان دریافت که این واژه‌های منتخب تا چه اندازه با پیکره یا فهرست‌های واژگانی هم‌پوشانی دارند؟ به‌طور دقیق، پژوهش حاضر تلاشی است در جهت پاسخ‌گویی به پرسش زیر:

واژه‌های ارائه شده در کتاب‌های آموزش زبان فارسی به غیرفارسی‌زبانان تا چه اندازه با پیکره‌ی واژگانی مورد نظر این پژوهش انطباق و هم‌پوشانی دارند؟

۲. پیشینه‌ی پژوهش

در سال‌های اخیر، منابع آموزش زبان فارسی به غیرفارسی‌زبانان مورد توجه قرار گرفته و از زوایای مختلفی نظیر دیدگاه‌های اصلی زبان‌شناسی، و انگاره‌ها و نظریه‌های مطرح در حوزه‌ی آموزش زبان دوم/خارجی بررسی شده‌اند. می‌توان گفت که در این میان، منابع مذکور بیش از هر چیز به لحاظ مفاهیم و اصول

^۱. applied linguist

زبان‌شناسی نقش‌گرا بررسی شده‌اند. به‌عنوان نمونه، غفاری‌مهر (۱۳۸۳) مجموعه‌ی آموزشی *آزفا* و همچنین *فارسی بیاموزیم* را از دیدگاه نقش‌گرایی مورد بررسی قرار داده است؛ همچنین، رضاخانی (۱۳۸۵) مجموعه‌ی *فارسی بیاموزیم* را از منظر فرانتش بینا فردی دستور نقش‌گرای هلیدی مورد مطالعه قرار داده است. از دیگر آثاری که کتاب‌های آموزش زبان فارسی به غیر فارسی‌زبانان را از منظر زبان‌شناسی نقش‌گرا مورد مطالعه قرار داده‌اند، می‌توان به داوری اردکانی و آقابراهیمی (۱۳۹۱)، رضاقلی‌فامیان (۱۳۹۲)، رضایی و علی‌پور (۱۳۹۱)، سعیدی و شرفی‌نژاد (۱۳۹۲) و تبریزمنش (۱۳۸۷) اشاره کرد.

به‌عنوان یک نمونه از پژوهش‌هایی که منابع آموزش زبان فارسی به غیر فارسی‌زبانان را براساس نظریه‌های مطرح در حوزه‌ی آموزش زبان واکاوی نموده‌اند، به پژوهش صحرائی و شهباز (۱۳۹۱) اکتفا می‌کنیم. این پژوهش نیز همان کتاب‌هایی را مورد تحلیل قرار داده‌اند که در پژوهش حاضر مورد بررسی قرار می‌گیرند. هدف از پژوهش آنها بررسی کتاب‌های آموزش زبان فارسی در زمینه‌ی مهارت خواندن بوده است و یافته‌هایشان نشان می‌دهد که کتاب‌های آموزش زبان فارسی به غیر فارسی‌زبانان توجه چندانی به یافته‌های نوین علم آموزش زبان و به‌ویژه انگاره‌ی اندرسون ندارند. همچنین، بررسی آنان، بیانگر آن است که اکثر کتاب‌های آموزشی، به اصل گسترش گنجینه‌ی واژه‌ها توجه فراوانی دارند.

وکیلی‌فرد (۱۳۸۷: ۲۱۶) با طرح این پرسش که «کدام زبان فارسی را به غیر فارسی‌زبانان آموزش دهیم؟»، به دنبال آن است که نشان دهد هر پاسخی به این پرسش تابع اهداف آموزشی است. وکیلی‌فرد (۱۳۸۷) در این‌باره می‌نویسد: «برای آموزش کارآمد و گسترش همه‌جانبه‌ی زبان فارسی، به‌ناچار باید به گرینش عناصر اساسی زبان دست زد و با پژوهشی در زبان فارسی گفتاری و جستجوی مفیدترین واژه‌های موجود، حدود مجموعه‌ی فارسی پایه را مشخص کرد. واژه‌نامه‌ها ده‌ها هزار واژه در اختیار ما می‌نهند، اما فارسی پایه از تعدادی اسم، صفت، فعل، واژه‌های دستوری، سایر واژه‌ها و نکاتی راجع به صرف و نحو تشکیل می‌شود».

اما از آنجا که موضوع پژوهش حاضر، بررسی میزان هم‌پوشانی واژه‌های منابع آموزش زبان فارسی به غیر فارسی‌زبانان با یک بیکره‌ی معین می‌باشد، در ادامه، به بیان پژوهش‌های می‌پردازیم که از این لحاظ با پژوهش حاضر تشابهی دارند.

احمدی‌قادر (۱۳۹۰) آموزش مستقیم واژه در متن را بررسی می‌کند و تأثیر دو رویکرد یادگیری مستقیم و تصادفی در یادگیری واژه را مقایسه می‌کند. یافته‌های بررسی وی نشان می‌دهد که آموزش مستقیم واژه در تمام سطوح دانش واژگانی بسیار مؤثرتر از موقعیت یادگیری تصادفی از طریق خواندن است. شاید بتوان گفت این یافته‌ها با شیوه‌های افرادی مانند نیشن (۲۰۰۶ و ۲۰۱۴) که به آموزش واژه از طریق خواندن می‌پردازند، در مغایرت باشد.

پژوهش عبادی، وکیلی فرد و بهراملو (۱۳۹۳) را می‌توان تلاشی دانست در جهت پاسخ به چگونگی تدوین واژگان پایه‌ی زبان فارسی، با بهره‌گیری از یک روش تلفیقی که شیوه‌های پیکره‌بنیاد را نیز شامل می‌شود. عبادی و همکارانش، با تعیین پوشش متنی و گستره‌ی واژگان فرهنگ بسامدی بی‌جن‌خان و محسنی (۱۳۹۱)، در یک پیکره‌ی ۳۴۰۰۰ واژه‌ای، نشان می‌دهند که این فرهنگ چه میزان از واژگان پیکره را پوشش می‌دهد. یافته‌های آنان نشان داد که فرهنگ بسامدی مذکور، ۸۴٪ از کلمات پیکره را پوشش می‌دهد؛ بنابراین، واژگان این فرهنگ از گستره‌ی قابل‌قبولی برخوردار است. یکی از شباهت‌های اصلی پژوهش حاضر با پژوهش عبادی و همکارانش، مبنا قرار دادن واژه‌های پُرکاربرد زبان فارسی، به‌عنوان یکی از معیارهای یادگیری واژگان است. به هر حال، این پژوهش به بررسی کتاب‌های درسی آموزش زبان فارسی نمی‌پردازد و آنان این موضوع را در یک فرهنگ بسامدی مورد توجه قرار داده‌اند.

پژوهش دیگری که به لحاظ هدف و بهره‌گیری از زبان‌شناسی پیکره‌ای به پژوهش حاضر شباهت داشت، پژوهش بی‌جن‌خان، نصری و جلابی (۱۳۹۳) است. این پژوهش به لحاظ روش‌شناسی یک تفاوت بسیار مهم با پژوهش حاضر دارد و آن این است که در پژوهش مذکور، به جای واژه‌های عینی، واحد واژگانی کانونی که دربرگیرنده‌ی چند صورت‌واژه بوده و می‌توان آن را یک واحد ذهنی نامید، مورد محاسبه و شمارش قرار می‌گیرد، اما در پژوهش حاضر، یک به یک ۵۰۰۰ واژه‌ی پُرسامد کتاب‌های مورد بررسی، با پُرسامدترین واژه‌های پیکره‌ی مبنا مورد مقایسه قرار می‌گیرد و نتایج آن، وضعیت کتاب‌های فعلی آموزش زبان فارسی به غیرفارسی‌زبانان را در ارائه و معرفی پُرسامدترین پُرکاربردترین واژه‌های زبان فارسی به فارسی‌آموزان نشان می‌دهد. به هر حال، بی‌جن‌خان و همکارانش براساس یافته‌های پژوهش خود این گونه استدلال می‌کنند که براساس قانون زیف، تعداد واژه‌های پُرسامد فهرست‌های بسامدی اندک است و انتظار می‌رود بر حسب احتمال وقوع‌شان در هر ماده‌ی آزمون به کار بروند و باید در ارزیابی مهارت واژگانی مورد نظر قرار بگیرند. ولی با توجه به این که فراوانی ۹۹٪ واژه‌ها در پیکره‌های رسمی و محاوره‌ای به ترتیب زیر ۱۰۰۴ و ۶۸۳ است و این واژه‌ها در زمره‌ی واژه‌های با بسامد متوسط و کم قرار می‌گیرند، آگاهی بیشتر زبان‌آموزان نسبت به واژه‌های کم‌بسامد موقعیت‌بنیاد، در پرسش‌های آزمون، نشانگر مهارت واژگانی بیشتر است.

از میان مطالعاتی که بر واژه‌آموزی در زبان انگلیسی تمرکز کرده است، میلتون (۲۰۰۹)، یک بررسی جامع و همه‌جانبه است که در کنار سایر موارد، بر نقش کتاب‌های درسی در آموزش و یادگیری واژه نیز متمرکز می‌شود. مرتبط‌ترین یافته‌ی میلتون با پژوهش حاضر، بخشی از نتیجه‌گیری او مبنی بر این است که بسیاری از کارهایی که در خصوص واژه‌آموزی انجام شده است، توجه چندانی به واژه‌های پُرسامد نکرده‌اند.

جیمینز و منسب (۲۰۰۸)، در بررسی خود دریافتند که چنانچه مجموع کل واژه‌های کتاب‌های درسی II واژه باشد، صرفاً کمتر از ۵۰٪ از این تعداد واژه، واژه‌های پُرسامد زبان مورد بررسی را شامل می‌شود.

کریادو و سانچز (۲۰۱۲) با هدف تعیین فرصت‌های لازم برای یادگیری واژه‌های ارائه شده در کلاس‌های آموزش زبان، از منظر بسامد اقلام واژگانی و تکرار واژه‌ها، به بررسی و تحلیل محتوای دو عنوان از کتاب‌های آموزش زبان پرداختند. آنها با ایجاد پیکره‌های زبانی از دو کتاب درسی مورد نظر و تهیه‌ی فهرست‌های واژگانی بسامدی از این دو پیکره، در فهرست‌های ایجاد شده محاسبات واژگانی انجام دادند. یافته‌های این پژوهش نشان داد که همبستگی بین فراگیری اقلام واژگانی و بسامد رخداد آنها در کتاب‌های درسی و در مجموع، همبستگی بین رخداد اقلام واژگانی مورد نظر در کتاب‌های درسی و بیان آن توسط مدرسان ضعیف است و به لحاظ آماری، همبستگی آن با واژه‌آموزی معنی‌دار نیست.

در این میان، مطالعات و آثار نیشن (۲۰۰۶ و ۲۰۱۴) از اهمیت بیشتری برخوردار است؛ به طوری که روش‌شناسی این آثار که در بخش مبانی نظری مورد بحث و بررسی قرار خواهد گرفت، در انجام پژوهش حاضر مؤثر بود.

۳. مباحث نظری

نیشن (۲۰۱۴: ۱) معتقد است «از کارهای اولیه‌ی افرادی مانند وست (۱۹۵۵) و بعدها در نتیجه‌ی فرضیه‌ی درونداد^۱ کرشن (۱۹۸۵)، جنبش قوی و فزاینده‌ای برای تشویق به استفاده از برنامه‌های فشرده‌ی خواندن در توسعه‌ی آموزش زبان خارجی به راه افتاد» به گفته‌ی وی، غیر از یک بررسی (کاب، ۲۰۰۷)، هیچ مطالعه‌ی پیکره‌بنیاد دیگری درباره‌ی یادگیری تعداد بسیار زیادی واژه از طریق خواندن انجام نشده است. وی با طرح این سؤال که «آیا یادگیری میزان کافی از واژه‌ها، صرفاً از طریق خواندن انجام نشده است. وی با طرح که پاسخ مناسبی برای یک سؤال مهم‌تر بیابد و آن این که «میزان درونداد واژه‌هایی که باید برای یادگیری پُربسامدترین واژه‌های زبان (۹۰۰۰ واژه‌ی پُربسامد زبان انگلیسی) در متون مختلف توسط زبان‌آموز مورد خواندن قرار بگیرد، چه قدر است؟». نیشن (۲۰۱۴: ۳) به مناقشاتی نیز که در این زمینه وجود دارد اشاره می‌کند (برای نمونه، کاب، ۲۰۰۷ و ۲۰۰۸ با مک‌کویلان و کرشن، ۲۰۰۸ در خصوص این که آیا امکان یادگیری واژگان کافی صرفاً از طریق درونداد خواندن وجود دارد یا خیر اختلاف نظر دارند).

در پژوهش بی‌جن‌خان و همکاران (۱۳۹۳)، دسترسی به واحدهای واژگانی کانونی شاخص ارزیابی مهارت واژگانی در نظر گرفته شده است. آنها بر این باور هستند که واژگان بسامدی به علت عدم وابستگی به منبع و سیاق، بسیاری از واژه‌های کانونی را پوشش می‌دهند. بدین لحاظ، نظر آنان مؤید سنجه‌ی مورد استفاده در پژوهش حاضر است. نکته‌ی مهمی که لازم است در رابطه با آن پژوهش مورد بررسی جدی قرار گیرد،

^۱. input hypothesis

چگونگی دسترسی به آن چیزی است که بی‌جن‌خان و همکارانش، ظاهراً با وام‌گیری از کراسلی و همکاران (۲۰۱۰) از آن با عنوان دسترسی به «واحد‌های واژگانی کانونی» یاد کرده‌اند.

نیشن (۲۰۱۴: ۴) برای انجام پژوهش بر مبنای خانواده‌های واژگانی^۱ مزیت‌ها و کاستی‌هایی را برمی‌شمارد. به اعتقاد وی، دسترسی زبان‌آموز به یکی از این واژه‌های هم‌خانواده را می‌توان به معنی دانش او از این واژه و سایر صورت‌واژه‌های آن دانست. به نظر می‌رسد بتوانیم آنچه را که بی‌جن‌خان و همکاران (۱۳۹۳) واحد‌های واژگانی کانونی نامیده‌اند و آن را به‌عنوان واحد بررسی خود در نظر گرفته‌اند، معادل هم‌خانواده‌های واژه که نیشن مورد بررسی قرار داده است، تلقی کنیم. با وجود این، شاید به دلیل مشکلاتی که در سنجش عملی دسترسی به واحد‌های واژگانی کانونی پیش می‌آید، نتوان با اطمینان کامل بر نتایج حاصل از این شیوه صحه گذاشت. برخی از این دلایل به شرح زیر هستند:

الف) واحد‌های واژگانی کانونی بیش از حد انتزاعی هستند و این میزان از انتزاع عملاً پژوهشگر را در محاسبه‌ی دقیق تعداد این واحد با مشکل مواجه می‌کند.

ب) واحد واژگانی کانونی، یک واحد ذهنی است و برای هر گونه بررسی دقیق علمی لازم است شرایطی را در نظر بگیریم که بتوان هر واژه‌ی انتزاعی^۲ و صورت‌واژه‌های^۳ وابسته به آن را (که به بیان نیشن، همان واژه‌های هم‌خانواده است) در همان شرایط آرمایی ذهنی و بافت‌زده^۴ بررسی کرده و مورد محاسبه یا شمارش قرار دهیم. در حالی که داده‌های واژگانی بسامدی مورد استفاده‌ی بی‌جن‌خان و همکاران (۱۳۹۳)، همگی بافت‌متکی^۵ بوده و فقط به دلایل پیکره‌ای از بافت خود منفک شده‌اند. اما در هنگام شمارش این داده‌ها، عملاً با آنان به گونه‌ای رفتار می‌شود که گویی ارتباطشان با آن واحد‌های واژگانی کانونی که براساس استدلال فوق، واحدهایی ذهنی هستند، قطع شده است.

ج) به نظر می‌رسد واحد‌های واژگانی کانونی، واحد‌های مناسبی برای بررسی یا ارزیابی غنای واژگانی گویندگان زبان و نیز زبان‌آموزان باشند، اما در مورد متن‌ها و پیکره‌های زبانی، انجام بررسی‌های کمی با ملاک قراردادن صورت‌واژه‌هایی که در متن‌ها یا پیکره‌ی زبانی وجود دارد، نتایج دقیق‌تری به دست می‌دهد.

از آنجا که کنش زبانی بر توانش زبانی استوار است، می‌توان نتایج بررسی‌های مبتنی بر پیکره‌های زبانی کلان و بزرگ را که لغزش‌ها و خطاهای زبانی در آن به سمت صفر می‌روند، به «واژگان ذهنی»^۶ نیز تعمیم داد و با روش‌های آزمایشگاهی و آزمون‌های روان‌شناختی که در حوزه‌ی روان‌شناسی زبان انجام می‌گیرند،

1. word families

2. lexeme

3. word-forms

4. de-contextualized

5. context-based

6. lexicon

این نتایج را در بوته‌ی آزمایش قرار داد. بنابراین، چنانچه واژه‌ها را مستقل از صورت ذهنی آنها و به همان صورت که در بافت زبانی ظاهر شده‌اند، شمارش کنیم، نتایج حاصل از آن، حداقل به لحاظ کمی از دقت کافی برخوردار خواهد بود. بی‌جن‌خان و همکاران (۱۳۹۳: ۴۲) نیز به این مسأله واقف بوده‌اند و در توضیح این مسأله بیان می‌کنند که «این سؤال مطرح است که چون هر داده‌ی مستند زبانی به یک سیاق و یک موقعیت خاص ارتباطی [مانند . . .] وابسته است، چگونه می‌توان از واژگان بسامدی و ویژگی‌های آماری مستقل از سیاق به‌عنوان یک ابزار، در ساخت پرسش‌ها و ارزیابی آزمون استفاده کرد . . . این اشکال به نمایه‌ی فراوانی واژگانی لافر و نیشن (۱۹۹۵) نیز وارد است». همان‌طور که ملاحظه می‌شود، آنان در نهایت ادعان داشته‌اند که این اشکال در موارد دیگر نیز وجود داشته است. بنابراین، بهتر است برای بررسی میزان واژگان متن‌های آموزشی و یا حتی واژگان ذهنی زبان‌آموزان، تک‌تک واژه‌های داده‌های گردآوری‌شده محاسبه و شمارش شوند.

به هر حال، در پژوهش حاضر تا اندازه‌ای از روش‌های نیشن (۲۰۱۴) الگوبرداری شده است. آنچه پژوهش حاضر را به پژوهش نیشن (۲۰۱۴) شبیه می‌کند، این است که نیشن نیز بیشتر به کمیّت یا میزان درونداد مطالب خواندنی برای یادگیری تعداد مشخصی از واژگان می‌پردازد. پژوهش حاضر با این هدف انجام می‌گیرد که مشخص شود محتوای آموزشی کتاب‌های موجود در حوزه‌ی آموزش زبان فارسی به غیرفارسی‌زبانان تا چه اندازه نیازهای واژگانی زبان‌آموزان را پوشش می‌دهد، با این تفاوت که نیشن در پژوهش خود بر ۹۰۰۰ واژه‌ی پُربسامد متمرکز می‌شود و پژوهش حاضر بر ۵۰۰۰ واژه‌ی نخست^۱ در زبان‌آموزی متمرکز است.

پژوهش‌ها نشان می‌دهند که در زبان‌آموزی، تسلط زبان‌آموزان بر ۹۰۰۰ واژه‌ی نخست در واژه‌آموزی، علاوه بر اسم‌های خاصی که زبان‌آموزان با آنها مواجه می‌شوند، حدود ۹۸٪ از واژه‌های جاری را در گستره‌ی وسیعی از متن‌ها پوشش خواهد داد.^۲ از این‌رو، تمرکز بر ۵۰۰۰ واژه‌ی نخست واژه‌آموزی را می‌توان به‌طور منطقی، یک سطح حداقل متوسط^۳ در نظر گرفت و این موضوعی است که در پژوهش حاضر هدف‌گذاری شده است. بنابراین، معیار بسامد واژه‌ها در پیکره‌ی طبیعی زبان فارسی، مبنای سنجش میزان هم‌پوشانی واژه‌های ارائه شده در کتاب‌ها با واژه‌های پُربسامد زبان فارسی، برای تأمین نیازهای واژگانی فارسی‌آموزان قرار گرفت و محاسبه شد که ۵۰۰۰ واژه‌ی پُربسامد کتاب‌های مورد بررسی تا چه اندازه با پُربسامدترین واژه‌های زبان فارسی طبیعی هم‌پوشانی دارند. چگونگی انجام این محاسبات در بخش چهارم تشریح شده است.

^۱. the first 5000 words for vocabulary learning

^۲. به‌عنوان نمونه‌ای از این پژوهش‌ها ر.ک. نیشن (۲۰۰۶ و ۲۰۱۴)

^۳. intermediate

۴. روش پژوهش

برای گردآوری و تحلیل داده‌ها در این پژوهش، به‌طورکلی روش‌های تحلیل محتوا، مطالعات کتابخانه‌ای و مقایسه‌ی فهرست‌های واژگانی مورد استفاده قرار گرفت و به‌طور خاص، از روش‌ها و شیوه‌های زبان‌شناسی پیکره‌ای برای بررسی وضعیت واژگانی کتاب‌ها استفاده شد.

۱.۴. روش گردآوری داده‌ها

برای گردآوری داده‌ها، سه عنوان از کتاب‌های آموزش زبان فارسی به غیرفارسی‌زبانان که در مرکز آموزش زبان فارسی دانشگاه بین‌المللی امام خمینی^(۵)، بنیاد سعدی، و جامعه‌الزهرای کاربرد فراوانی دارند، انتخاب شدند. تعداد آثاری که برای آموزش زبان فارسی به نگارش درآمده است، فهرستی مشتمل بر حداقل ۱۵۰ عنوان کتاب است. بنابراین، با توجه به این که بررسی تمامی این آثار ممکن نبود، معیارهای زیر برای انتخاب سه عنوان از کتاب‌های مذکور، به‌عنوان نماینده‌ی سایر کتاب‌ها به‌کار بسته شد:

- کاربرد در مراکز آموزش زبان فارسی به غیرفارسی‌زبانان

- جدید بودن آثار به لحاظ سال تدوین و نشر

- حجم کتاب

- کتاب‌هایی که تا حد امکان معرف همه‌ی کتاب‌های آموزش زبان فارسی به غیرفارسی‌زبانان، در داخل ایران باشند.

در نهایت، با توجه به ملاک‌های فوق، آثار زیر جهت بررسی و تحلیل محتوا برگزیده شد:

- زبان فارسی (آموزش زبان و فرهنگ ایران به غیرفارسی‌زبانان)، تألیف صفارمقدم در چهار جلد که یکی

از کتاب‌های اصلی مورد استفاده در بنیاد سعدی و دانشگاه شهید بهشتی است.

- آموزش زبان فارسی، تألیف سید فاطمه رسولی در چهار جلد که در مجتمع آموزشی جامعه‌الزهرای مورد

استفاده قرار می‌گیرد.

- فارسی بیاموزیم، تألیف ذوالفقاری، غفاری و محمودی بختیاری در سه جلد که در مرکز آموزش زبان

فارسی دانشگاه بین‌المللی امام خمینی^(۵) به‌عنوان کتاب درسی به‌کار می‌رود.^۱

ابتدا متن تمام این آثار تایپ شد و با رعایت معیارهای زبان‌شناسی پیکره‌ای، یک پیکره‌ی زبانی رایانشی

برای هر یک از کتاب‌های مورد بررسی و همچنین، یک پیکره‌ی آموزشی، متشکل از پیکره‌های ایجاد شده

برای همه‌ی کتاب‌های مورد بررسی طراحی و ایجاد گردید. آن‌گاه با استفاده از این پیکره‌های زبانی و نیز با

^۱ «فارسی بیاموزیم» مجموعه‌ای پنج جلدی است که دو جلد آخر آن برای فارسی‌آموزی در سطوح پیشرفته است. با توجه به این که هدف پژوهش حاضر، بررسی واژه‌آموزی تا سطح متوسط می‌باشد، صرفاً سه کتاب اول این مجموعه مورد بررسی قرار گرفت.

به‌کارگیری جدیدترین نرم‌افزارهای تحلیل پیکره‌های زبانی (ر.ک. بخش ۵) فهرست‌های واژگانی-بسامدی^۱ برای دستیابی به داده‌های لازم ایجاد شد. از سوی دیگر، با استفاده از متن‌های موجود در پایگاه داده‌های زبان فارسی^۲، یک پیکره‌ی متوازن^۳ زبانی طراحی و ایجاد شد تا داده‌های لازم برای اخذ یک فهرست واژگانی-بسامدی، به‌عنوان مبنا و مرجع بررسی هر یک از کتاب‌های مورد بررسی به دست آید.

۲.۴. طراحی و ایجاد پیکره‌ی مبنا

یکی از مهمترین و پُرچالش‌ترین مراحل انجام پژوهش، طراحی و ایجاد پیکره‌های مناسب و سازگار^۴ با شیوه‌های تحلیل پیکره‌های زبانی بود. در این پژوهش، برای تصمیم‌گیری درباره‌ی چگونگی طراحی و ایجاد یک پیکره‌ی متوازن، روش و پیشنهاد آتکینز و راندل (۲۰۰۸: ۶۹ - ۶۸) اساس کار قرار گرفت. با این تفاوت که چون پیکره‌ی آنان یک پیکره‌ی عمومی بود که به منظور فرهنگ‌نگاری^۵ طراحی و ایجاد شده بود، لازم بود در پیکره‌ی این پژوهش، نسبت متون گفتاری به نوشتاری برای اهداف زبان‌آموزی تغییر کند و در واقع، افزایش یابد؛ زیرا در آموزش زبان، یکی از مهارت‌های اساسی زبان‌آموزی گفتار است و چون این مهارت در تعامل مستقیم با مهارت درک شنیداری نیز می‌باشد، اهمیت وجود آن در پیکره‌های زبان‌آموزی دوچندان می‌گردد. بنابراین، می‌توان نسبت متون گفتاری را برای پیکره‌های زبانی‌آموزی تا دو برابر نیز افزایش داد. برای این کار، پایگاه داده‌های زبان فارسی (۱۳۷۶)^۶ به‌عنوان درون‌داد این پیکره‌ی مرجع برگزیده شد.

عاصی (۱۳۷۲: ۱۷) بر این باور است که نمونه‌های مننی^۷ برگزیده‌ای^۸ که درون‌داد پایگاه داده‌های زبان فارسی را تشکیل می‌دهند، نماینده‌های واقعی زبان فارسی معاصر به شمار می‌روند. بنابراین، نسبت به طراحی و ایجاد پیکره‌ای برگزیده از متن‌های موجود در پایگاه داده‌ها اقدام شد. برای اطمینان از متوازن بودن پیکره‌ی برگزیده، فهرست متون برگزیده‌ی پایگاه داده‌های زبان فارسی مبنای انتخاب متونی شد که قرار بود در پیکره‌ی مرجع قرار داده شود. در این فهرست سعی شد از ژانرهای مختلف زبان فارسی به نسبت لازم در پیکره‌ی مرجع/مبنا قرار داده شده و تا حد امکان پیکره‌ای ایجاد گردد که اولاً نماینده‌ی زبان طبیعی فارسی امروز باشد و ثانیاً، در زبان‌آموزی نیز از اثربخشی لازم برخوردار باشد. به این معنی که به‌طور نمونه، متن‌های تاریخی که دربرگیرنده‌ی واژه‌های مهجور^۹ و صورت‌های واژگانی کهنه^۹ هستند و نیز متن‌های تخصصی و فوق

1. lexical-frequency list

2. www.pldb.ihcs.ac.ir

3. balanced corpus

4. compatible

5. lexicography

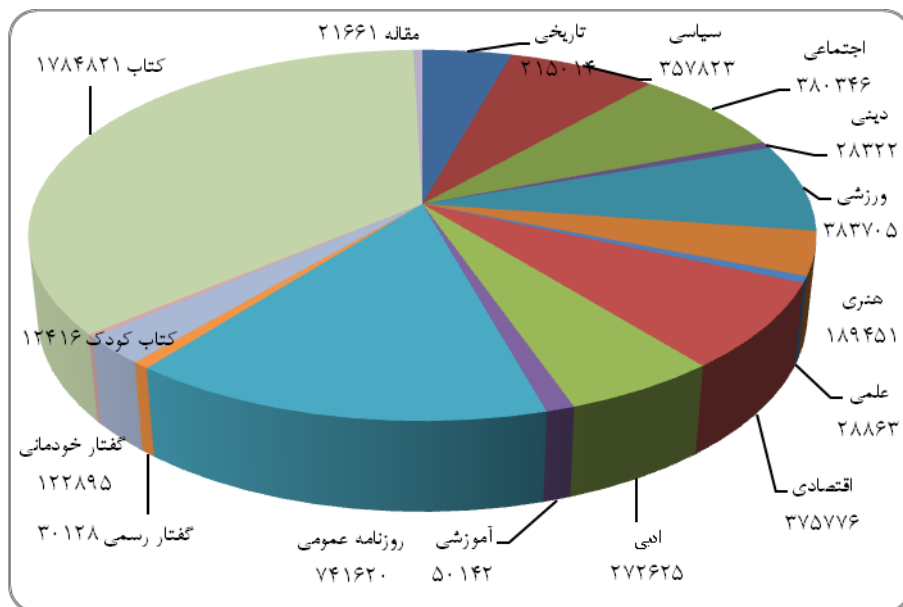
6. Persian Language Data Base/PLDB, ©Assi (1994)

7. selected textual samples

8. obsolete

9. archaic lexical form

تخصصی به کلی کنار گذاشته شدند. در عین حال، سعی شد با توجه به این که برخی متن‌های تخصصی یا به عبارت دقیق‌تر، نیمه‌تخصصی وارد زندگی مردم عادی شده‌اند و در زبان‌آموزی حائز اهمیت هستند، با نسبتی منطقی و به‌طور متنوع از حوزه‌های مختلفی نظیر بخش‌هایی از روزنامه‌ی ایران اجتماعی، ادبی، اقتصادی، مذهبی، هنری، ورزشی و مانند آن نیز در پیکره‌ی مرجع قرار داده شود.^۱ نمودار (۱) نسبت و حجم واژگانی هر یک از متن‌هایی را که در پیکره‌ی مرجع/مبنا قرار داده شده، نشان می‌دهد.



نمودار ۱. ترکیب متنی، واژگانی و حجم هر یک از متن‌های پیکره‌ی مبنا (اعداد مندرج در برش‌های نمودار، تعداد واژه‌های آن را نشان می‌دهد)

در نهایت می‌توان ادعا کرد که پیکره‌ی مبنا، به‌عنوان یک پیکره‌ی متوازن و با تعداد ۵۰۳۰۱۳۸ واژه، از حجم و بزرگی مناسب و همچنین ترکیب متنی لازم برخوردار است تا بتوان از آن به‌عنوان نماینده‌ی فارسی طبیعی امروز برای سنجش واژگانی هر پیکره‌ی دیگری استفاده کرد.

۳.۴. پیکره‌ی واژگانی

برای ایجاد پیکره‌ی واژگانی ابتدا متن‌های مورد نظر حروف‌نگاری شدند. پس از حروف‌نگاری با هدف یکسان‌سازی و یک‌دست شدن متن‌ها، فایل‌ها بررسی و مورد اصلاح قرار گرفتند و در نهایت، برای آماده‌سازی و پالایش متون، متن هر سه کتاب حداقل سه دور، مرور و اصلاح شد. دور اول برای حذف جدول‌ها و

^۱. برای آگاهی از نوع یا قالب متن‌های مندرج در پیکره‌ی مرجع، تعداد هر قالب و تعداد واژه‌های آن ر.ک. پیوست همین مقاله

فضاهای خالی اضافی قرار گرفته در متن که موجب سنگینی متن و عدم امکان پردازش و تحلیل ماشینی آنها می‌شد. دور دوم به دو منظور انجام شد: (۱) برای اطمینان از عدم وجود غلط‌های املایی یا هر گونه اشتباه تایپی، (۲) برای یک‌دست‌سازی متن هر یک از کتاب‌ها. دور آخر به این منظور انجام گرفت که پیکره‌ای مناسب تحلیل ایجاد شود و همچنین، همه‌ی افعال مرکب، اسم‌های مرکب، قیدها و کلمات پیوندی/رابط^۱ چندجزئی صرفاً با نیم‌فاصله (شبه‌فاصله)^۲ در کنار هم قرار داده شوند. به‌طور نمونه، فعل‌های مرکبی مانند «مورد توجه قرار گرفت» و «فرار کرد» به صورت «مورد توجه قرار گرفت» و «فرار کرد»؛ اسم‌های مرکبی که با شکل‌های مختلفی مانند «کتابخانه» یا «کتاب‌خانه» در متن هر یک از کتاب‌ها وجود داشتند، به‌صورت یک‌دست و به شکل «کتاب‌خانه» در همه‌ی کتاب‌ها حروف‌نگاری شدند و قیدهای چندجزئی و کلمات پیوندی/رابطی نظیر «با توجه به این که» و «بنا بر این» و مانند آن، به‌صورت یک‌دست در همه‌ی متن‌ها به شکل «باتوجه‌به‌این‌که» و «بنابراین» تنظیم و حروف‌چینی شدند تا ابزارهای مورد استفاده بتوانند آنها را یک قلم واژگانی^۳ در نظر گرفته و فهرست آنها را ارائه کنند.

در نهایت، محتوای همه‌ی جلد‌های هر یک از کتاب‌های مورد بررسی به صورت یک متن کامل و به عبارت فنی، یک پیکره‌ی زبانی تنظیم شد. همچنین، این سه متن کاملی که از هر یک از کتاب‌ها به‌دست آمده بود با هم ترکیب شدند تا در نهایت یک پیکره‌ی بزرگ‌تر به دست آید. این پیکره حاصل ترکیب سه عنوان از کتاب‌های آموزش زبان فارسی به غیرفارسی‌زبانان بود و با دلایلی که در بخش (۴،۱) گفته شد، می‌توان آنها را نماینده‌ی کتاب‌های این حوزه تلقی کرد. به این ترتیب، اطلاق اصطلاح «پیکره‌ی زبان‌آموزی»^۴ به پیکره‌ی ایجاد شده اشتباه نخواهد بود. به هر حال، در پایان، چهار پیکره‌ی زبانی حاصل شد که عبارت بودند از:

- (۱) پیکره‌ی زبانی کتاب ذوالفقاری و همکاران، شامل محتوای هر سه جلد این کتاب
- (۲) پیکره‌ی زبانی کتاب صفار مقدم، شامل محتوای هر چهار جلد این کتاب
- (۳) پیکره‌ی زبانی کتاب رسولی (جامعه‌الزهران)، شامل محتوای هر چهار جلد این کتاب
- (۴) پیکره‌ی شامل محتوای کتاب‌های مورد بررسی، یعنی ترکیب تمام متن‌های فوق که آن را پیکره‌ی زبان‌آموزی نامیدیم.

در واقع می‌توان پیکره‌ی زبان‌آموزی را که شامل محتوای آموزشی کتاب‌های آموزش زبان فارسی به غیرفارسی‌زبانان است، پیکره‌ی دانست که حداقل در داخل ایران نماینده‌ی واقعی از محتوای همه‌ی کتاب‌های این حوزه محسوب می‌شود، جدول (۱) دربرگیرنده‌ی اطلاعات این پیکره است.

1. linking words

2. zero-width-non-joiner character (pseudo-space)

3. lexical item

4. language learning corpus

جدول ۱. آمار محتوای متنی پیکره‌ی زبان‌آموزی

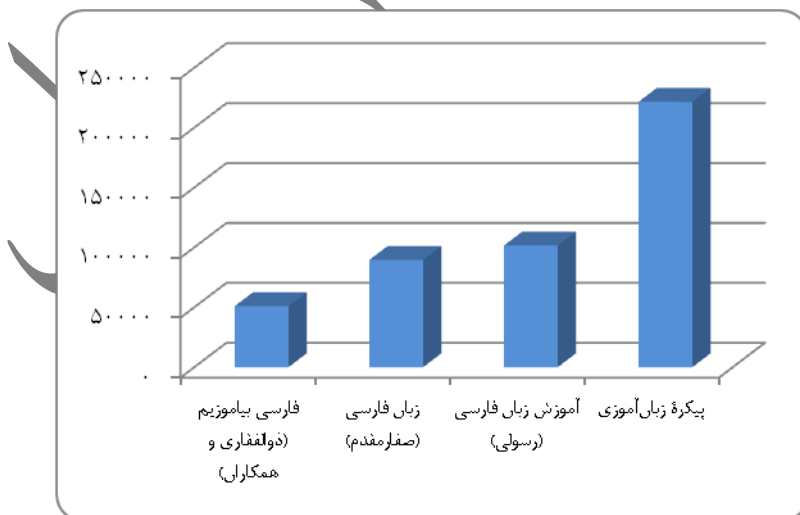
۱۲۰۲	تعداد صفحه
۲۲۱۵۱۴	تعداد واژه
۱۰۵۵۱۰۳	تعداد نویسه (بدون احتساب فاصله‌ی بین نویسه‌ها)
۱۳۰۱۰۳۰	تعداد نویسه (با احتساب فاصله)
۳۷۰۹۵	تعداد سطر

آنچه از محتوای پیکره‌ی زبان‌آموزی در تحلیل‌های پژوهش مورد استفاده قرار گرفت، آمار مربوط به واژه‌های آن بود. بنابراین، آمار واژه‌های هر یک از کتاب‌های مورد بررسی و همچنین، پیکره‌ی زبان‌آموزی در جدول (۲) ارائه شده است.

جدول ۲. آمار واژه‌های کتاب‌ها و پیکره‌ی ایجاد شده از آنها

ردیف	عنوان کتاب یا پیکره	تعداد واژه‌ها
۱	فارسی بیاموزیم (ذوالفقاری و همکاران، ۱۳۸۷)	۵۲۲۲۶
۲	آموزش زبان فارسی (رسولی، ۱۳۹۱)	۱۰۳۰۶۷
۳	زبان فارسی (صفا مقدم، ۱۳۸۶)	۹۲۸۴۰
۴	پیکره‌ی زبان‌آموزی	۲۲۶۴۲۳

با استفاده از نمودار (۲) می‌توان میزان واژه‌های هر یک از کتاب‌ها را با یکدیگر و همچنین با واژه‌های پیکره‌ی زبان‌آموزی ایجاد شده از خود این کتاب‌ها مقایسه کرد.



نمودار ۲. مقایسه‌ی تعداد واژه‌های کتاب‌ها و پیکره‌ی زبان‌آموزی با یکدیگر (محور عمودی، تعداد واژه‌ها را نشان می‌دهد)

۵. ابزار پژوهش

برای انجام تحلیل‌های لازم بر روی داده‌های گردآوری شده، از دو ابزار مختلف استفاده شد. اولین ابزار مورد استفاده در پژوهش، نرم‌افزاری است که حداقل در سطح کشور از قدرتمندترین نرم‌افزارهای تحلیل متن محسوب می‌شود. این نرم‌افزار در پایگاه داده‌های زبان فارسی قرار دارد و به لحاظ کارکردی و همچنین ساخت، مشابه برنامه‌ای است که در دانشگاه آکسفورد و با عنوان «برنامه‌ی واژه‌نمای^۱ آکسفورد» (ا.سی.پی.)^۲ مورد استفاده قرار می‌گیرد. در واقع، این نرم‌افزار شامل یک سامانه‌ی جستجوگر پیکره یا برنامه‌ی رایانه‌ای است که داده‌های پیکره را مورد پردازش قرار می‌دهد. بخش مهمی از تحلیل‌های پژوهش، به‌ویژه استخراج متن‌های دروندادی پیکره‌ی مبنا و اخذ فهرست واژگانی از آن توسط این نرم‌افزار انجام گرفت. دومین نرم‌افزار مورد استفاده در پژوهش، نسخه‌ی هفت نرم‌افزار وُرداسمیت (اسکات، ۲۰۱۵)^۳ بود که برای تنظیم فهرست واژگانی کتاب‌ها و پیکره‌ی زبان‌آموزی مورد استفاده قرار گرفت.

۶. تحلیل داده‌ها

۱.۶. راستی‌آزمایی داده‌های خروجی نرم‌افزارها

برای تحلیل داده‌های پژوهش، لازم بود تحلیل‌های انجام‌گرفته توسط نرم‌افزارها راستی‌آزمایی شوند. آزمودن درستی تحلیل داده‌ها در پژوهش حاضر به دو روش انجام شد: اول، چند متن بسیار کوتاه تنظیم شد. سپس با استفاده از نرم‌افزارهای مورد اشاره، فهرست‌های واژگانی و واژه‌نما از این متن‌ها استخراج شد. آن‌گاه برای تعیین درستی این فهرست‌ها، خروجی‌های انجام‌گرفته به صورت شمارش دستی کنترل شدند. خوشبختانه تمام موارد آزمایشی، درستی و دقت تحلیل‌های انجام‌شده توسط نرم‌افزار پایگاه داده‌های زبان فارسی و همچنین نرم‌افزار وُرداسمیت را تأیید کردند. دوم، با استفاده از هر دو نرم‌افزار پایگاه داده‌های زبان فارسی و وُرداسمیت از پیکره‌های مورد بررسی فهرست واژگانی گرفته شد که خوشبختانه خروجی هر دو یکسان بود. به این ترتیب، ضمن راستی‌آزمایی تحلیل‌های انجام‌شده مشخص گردید که هر دو نرم‌افزار در مورد زبان فارسی^۴ به درستی عمل می‌کنند.

^۱. concordance

^۲. The Oxford Concordance Program (OCP)

^۳. WordSmith Tools: Windows software for finding word patterns©M. Scott (2015)

^۴. نرم‌افزار وُرداسمیت در ابتدا با خط فارسی ناسازگار بود که طی مکاتبه‌های متعددی که با مایک اسکات انجام گرفت، ضمن

سازگار شدن این نرم‌افزار با خط فارسی، زمینه‌ی انتشار نسخه جدیدی از این نرم‌افزار نیز فراهم شد.

۶.۲. تحلیل داده‌ها

داده‌های پژوهش در دو بخش تحلیل شدند.

۶.۲.۱. استخراج فهرست‌های واژگانی-بسامدی از پیکره‌ها

در مجموع، پنج فهرست واژگانی زیر با استفاده از نرم‌افزار پایگاه داده‌های زبان فارسی و ابزارهای موجود در نرم‌افزار ورداسمیت به دست آمد که اساس تحلیل‌ها و یافته‌های پژوهش حاضر قرار گرفت:

الف) فهرست واژگانی مبتنی بر پیکره‌ی زبانی کتاب صفارمقدم

ب) فهرست واژگانی مبتنی بر پیکره‌ی زبانی کتاب ذوالفقاری و همکاران

ج) فهرست واژگانی مبتنی بر پیکره‌ی زبانی کتاب رسولی (جامعه‌الزهره‌ا)

د) فهرست واژگانی مبتنی بر پیکره‌ی زبانی زبان‌آموزی (ترکیب داده‌های سه پیکره‌ی الف، ب و ج)

ه) فهرست واژگانی ۶۰۰۰ واژه‌ای مبتنی بر پیکره‌ی مبنا (ایجاد شده براساس متن‌های برگزیده‌ای از پایگاه داده‌های زبان فارسی و انتخاب ۶۰۰۰ واژه‌ی آغازین این فهرست واژگانی که براساس بسامد وقوع واژه‌ها در پیکره‌ی مرتب‌شده بودند)

۶.۲.۲. مقایسه‌ی فهرست‌های واژگانی

برای نیل به هدف پژوهش، براساس پیکره‌ی مبنا تحلیل‌های مقایسه‌ای انجام گرفت. هدف از مقایسه‌ی واژه‌های کتاب‌های مورد پژوهش و پیکره‌ی مبنا، تشخیص برتری‌ها یا کاستی‌های کتاب‌ها در حوزه‌ی واژه بود. همچنین، با این مقایسه مشخص می‌شود که کدام یک از کتاب‌ها نسبت به دیگر کتاب‌ها، واژه‌های پُربسامد بیشتری را پوشش می‌دهند. برای انجام تحلیل‌های مقایسه‌ای، ابتدا فهرست بسامدی ۵۰۰۰ واژه‌ای هر یک از کتاب‌ها که براساس بالاترین میزان بسامد وقوع یا رخداد واژه‌های تشکیل‌دهنده‌ی هر پیکره مرتب شده بود، به پنج گروه ۱۰۰۰ واژه‌ای تقسیم شد. هر یک از این پنج گروه که براساس بسامد وقوع در پیکره‌های ایجادشده مرتب شده بودند، به‌عنوان سطوح اولویت‌بندی‌شده‌ی واژه‌آموزی در نظر گرفته شد.

در مرحله‌ی بعدی، میزان همپوشانی گروه‌های واژگانی هر یک از فهرست‌های مورد اشاره در بخش (۶)، (۲۰۱) با گروه‌های واژگانی متناظر با آنها در فهرست واژگانی پیکره‌ی مبنا تعیین شد. برای این منظور، هر یک از پنج گروه ۱۰۰۰ واژه‌ای دسته‌بندی‌شده (سطوح ۱ تا ۵) با گروه‌های متناظر خود در فهرست واژگانی برگرفته از پیکره‌ی مبنا مقایسه شد. در این مسیر، پس از آن که بخشی از این مقایسه‌ها انجام گرفت، مشخص شد که فهرست واژگانی برگرفته از پیکره‌ی مبنا دارای برخی مشکلات است. اگرچه این مشکلات محدود بودند، به هر حال می‌توانست بر دقت تحلیل‌ها اثرگذار باشد. با بررسی بیشتر مشخص شد که بخش فراوانی از این مشکلات در ۱۰۰۰ واژه‌ی ابتدایی فهرست واژگانی مبنا ظاهر شده‌اند. این مشکلات عبارت

بودند از وجود اعداد و برخی واژه‌ها و حروف لاتین و همچنین وجود خطاهای نوشتاری در متون پیکره‌ی مبنا (نظیر عدم یک‌دستی در رعایت نیم‌فاصله یا عدم دقت و جامعیت لازم در تعیین کلمات یا فعل‌های مرکب).

در ابتدا تصور بر این بود که فهرست‌های واژگانی هر یک از کتاب‌های مورد بررسی و نیز پیکره‌ی زبان‌آموزی به راحتی با فهرست واژگانی طبقه‌بندی‌شده‌ی پیکره‌ی مرجع مورد مقایسه قرار می‌گیرد و در نهایت میزان هم‌پوشانی (مطابقت) این فهرست‌ها تعیین خواهد شد. متأسفانه در عمل، یعنی زمانی که عملاً این تحلیل‌ها و مقایسه‌ها آغاز گردید، روند تحلیل‌ها و مقایسه‌ها به دلیل محدودیت‌های ذکر شده، بسیار تغییر کرد و پیچیده‌تر شد. در واقع، امکان مقایسه‌ی هر دو گروه هزارواژه‌ای متناظر از فهرست‌های واژگانی کتاب‌ها و فهرست واژگانی مرجع که تعداد واژه‌های آن به دلیل مشکلات مربوط به متون دروندادی پیکره‌ی مبنا قدری کمتر از ۱۰۰۰ واژه می‌شد، وجود نداشت. برای مرتفع کردن چنین محدودیتی و همچنین برای پرهیز از هر گونه سطحی‌کاری غیرعلمی، با دو هدف زیر، تغییر اندکی در گروه‌بندی پیکره‌ی مبنا انجام شد:

(۱) تعیین میزان اثربخشی هر یک از کتاب‌های مورد بررسی بر واژه‌آموزی

(۲) مرتفع کردن مشکلات و خلأهای موجود در فهرست واژگانی برگرفته از پیکره‌ی مبنا

این تغییرات به این صورت انجام پذیرفت که به جای یک فهرست واژگانی ۵۰۰۰ واژه‌ای، یک فهرست واژگانی ۶۰۰۰ واژه‌ای از پیکره‌ی مبنا ایجاد و به شش گروه تقسیم شد. گروه اول پیکره‌ی مبنا که بیشترین بخش از مشکلات مذکور نیز در آن مشاهده می‌شد، ۱۰۵۰ واژه، گروه دوم ۱۰۳۰ واژه و گروه‌های سوم، چهارم و پنجم هر کدام ۱۰۲۰ واژه را دربرمی‌گرفتند. بنابراین، هنگام مقایسه‌ی گروه‌های واژگانی مذکور با یکدیگر، عملاً زمینه‌ای فراهم شد تا موارد خطا، از گروه‌های ۱۰۰۰ واژه‌ای پیکره‌ی مرجع حذف و نتایج حاصل از مقایسه‌ها به درستی تقریباً صددرصدی نزدیک شود. این گروه‌بندی جدید در جدول (۳) نشان داده شده است.

جدول ۳. گروه‌بندی واژه‌های فهرست واژگانی پیکره‌ی مبنا

رتبه‌ی بسامدی واژه	سطح	ردیف
رتبه‌ی ۱ تا ۱۰۵۰	اول	۱
رتبه‌ی ۱۰۵۱ تا ۲۰۸۰	دوم	۲
رتبه‌ی ۲۰۸۱ تا ۳۱۰۰	سوم	۳
رتبه‌ی ۳۱۰۱ تا ۴۱۲۰	چهارم	۴
رتبه‌ی ۴۱۲۱ تا ۵۱۲۰	پنجم	۵
رتبه‌ی ۵۱۲۱ تا ۶۰۰۰	ششم	۶

به این ترتیب، تعداد ۱۴۰ واژه (۵۰ واژه برای گروه اول، ۳۰ واژه برای گروه دوم و ۲۰ واژه برای هر یک از گروه‌های بعدی) از مجموع ۱۰۰۰ واژه‌ی گروه ششم کاسته می‌شود که موارد خطا در گروه‌های قبلی را مرتفع می‌کند. بنابراین، گروه ششم به جای ۱۰۰۰ واژه، یک گروه ۸۶۰ واژه‌ای را تشکیل می‌دهد که نتایج

مقایسه‌ی فهرست واژگانی-بسامدی هر یک از کتاب‌ها و نیز پیکره‌ی زبان‌آموزی نه صرفاً با این گروه و آن هم در اولویت‌بندی واژه‌ها، بلکه با کل شش گروه، به منظور بررسی وضعیت کلی محتوای آموزش زبان فارسی به غیرفارسی‌زبانان در معرفی پُربسامدترین/پُرکاربردترین واژه‌های زبان فارسی به فارسی‌آموزان، در ردیف ششم جدول (۵) ارائه شده است.

در مقایسه‌ی فهرست‌ها، ابتدا هر یک از گروه‌های ۱۰۰۰ واژه‌ای کتاب‌ها، بدون کم و کاست و هیچ گونه مداخله‌ای، با گروه‌های واژگانی طبقه‌بندی‌شده‌ی پیکره‌ی مبنا مورد مقایسه قرار گرفت. به این ترتیب مشخص گردید که:

(۱) میزان هم‌پوشانی واژگانی گروه‌های ۱ تا ۵ کتاب‌ها و پیکره‌ی زبان‌آموزی با گروه‌های ۱ تا ۶ پیکره‌ی مبنا تا چه اندازه است.

(۲) میزان هم‌پوشانی گروه‌های ۱ تا ۵ فهرست واژگانی کتاب‌ها و پیکره‌ی زبان‌آموزی با کل فهرست ۶۰۰۰ واژه‌ای پیکره‌ی مبنا تا چه اندازه است. به عبارت دیگر، اگر واژه‌ای از گروه اول کتاب‌ها یا پیکره‌ی زبان‌آموزی در گروه اول مناظر با آن در پیکره‌ی مرجع وجود نداشت، اما در سایر گروه‌های شش‌گانه‌ی پیکره‌ی مرجع یافت می‌شد، باز هم نسبت به ثبت آن اقدام گردید تا مشخص شود که اگر این واژه به لحاظ اولویت‌بندی در واژه‌آموزی مورد توجه مؤلف قرار نگرفته است، آیا آن واژه حداقل به‌عنوان واژه‌ای ضروری و بسیار پُرکاربرد، در کتاب‌های آموزش زبان فارسی به فارسی‌آموزان، حداقل با بسامد کمتر معرفی شده است یا خیر؟

جدول (۴) نشان می‌دهد که هر یک از کتاب‌های مورد بررسی و نیز پیکره‌ی زبانی حاصل از ترکیب آنها (پیکره‌ی زبان‌آموزی)، به لحاظ سطوح زبان‌آموزی، تا چه میزان با فهرست واژگانی پیکره‌ی زبانی مبنا هم‌پوشانی یا مطابقت دارند.

جدول ۴. میزان و درصد هم‌پوشانی واژه‌های هر یک از گروه‌های پنج‌گانه‌ی کتاب‌ها و نیز پیکره‌ی زبان‌آموزی با گروه مناظر با آنها در پیکره‌ی مبنا

ردیف	گروه	ذوالفقاری و همکاران		رسولی (جامعه‌الزهرا)		صفا مقدم		پیکره‌ی زبان‌آموزی	
		تعداد	درصد	تعداد	درصد	تعداد	درصد	تعداد	درصد
۱	گروه اول	۴۴۱	۴۴/۱	۵۷۱	۵۷/۱	۴۲۱	۴۲/۱	۴۷۸	۴۷/۸
۲	گروه دوم	۲۷۷	۲۷/۷	۲۱۶	۲۱/۶	۳۶۱	۳۶/۱	۱۷۷	۱۷/۷
۳	گروه سوم	۶۳	۶/۳	۱۰۹	۱۰/۹	۶۰	۶	۹۷	۹/۷
۴	گروه چهارم	۳۸	۳/۸	۵۸	۵/۸	۴۶	۴/۶	۶۵	۶/۵
۵	گروه پنجم	۲۶	۲/۶	۵۲	۵/۲	۱۹	۱/۹	۵۱	۵/۱

داده‌های جدول (۴) نشان می‌دهد که در هیچ یک از گروه‌های واژگانی، هم‌پوشانی یا مطابقت واژگانی قابل قبولی میان واژه‌های گروه‌های پنج‌گانه‌ی کتاب‌ها و پیکره‌ی مبنا وجود ندارد. این هم‌پوشانی واژگانی، حتی در گروه اول که به دلیل قرار داشتن واژه‌های نقشی^۱ در آن، انتظار می‌رود به سمت ۱۰۰٪ میل کند، در بالاترین میزان مطابقت به حدود ۵۵٪ می‌رسد. همان‌طور که ملاحظه می‌شود، این میزان در سایر سطوح واژه‌آموزی حتی به زیر ۵٪ نیز تقلیل می‌یابد. به‌طور نمونه، اگر به سطح ۵ واژه‌آموزی (گروه پنجم جدول فوق) توجه کنیم، خواهیم دید که کتاب جامعه‌الزهره (رسولی)، از هزار واژه‌ای که در این گروه وجود دارد، صرفاً ۵۲ مورد، یعنی حدود ۵٪ از واژه‌های پُرسامدِ گروه ۵ در پیکره‌ی مبنا را پوشش می‌دهد. این در حالی است که هم‌پوشانی سایر کتاب‌ها و پیکره‌ی زبان‌آموزی در گروه پنج، حتی کمتر از کتاب رسولی است و به‌عنوان نمونه در کتاب صفارمقدم تا کمتر از ۲٪ نیز تقلیل می‌یابد. به عبارت دیگر، این کتاب از ۱۰۰۰ واژه‌ی پُرسامدِ گروه ۵ پیکره‌ی مبنا، فقط ۱۹ واژه‌ی آن را به زبان‌آموزان معرفی کرده است.

شاید در یک نگاه ساده‌انگارانه تصور شود که کتاب رسولی (جامعه‌الزهره) نسبت به سایر کتاب‌های مورد بررسی، حداقل قدری بهتر عمل کرده است، اما وقتی به حجم واژگانی این کتاب نسبت به سایر کتاب‌ها توجه می‌کنیم، خواهیم دید که علت این اندک عملکرد بهتر شاید صرفاً به این دلیل است که حجم این کتاب بیشتر از سایر کتاب‌ها است و در نتیجه، به‌طور ناخواسته واژه‌های بیشتری را پوشش خواهد داد. بنابراین، یافته‌ها در این مرحله نشان می‌دهند که کتاب‌های مورد بررسی، به‌عنوان نماینده‌ی کتاب‌های آموزش زبان فارسی به غیرفارسی‌زبانان و نیز پیکره‌ی زبان‌آموزی که حاصل ترکیب محتوای همه‌ی کتاب‌ها است، به لحاظ سطوح یادگیری واژه‌آموزی، به خوبی طراحی و برنامه‌ریزی نشده‌اند و نیاز به بازنگری جدی دارند. نمودار (۳) کمک می‌کند که در یک نگاه، برداشتی صحیح و کلی از این موضوع داشته باشیم.

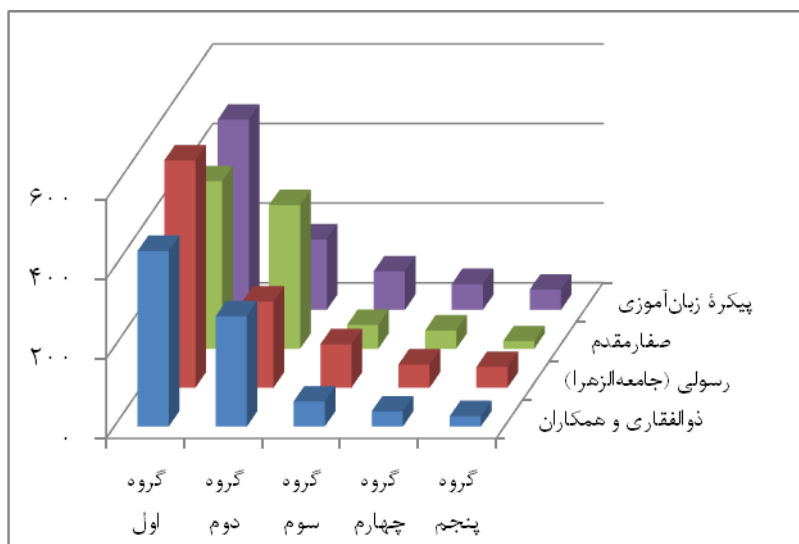
همان‌طور که در نمودار (۳) ملاحظه می‌شود، تعداد واژه‌های هم‌پوشان هیچ‌کدام از گروه‌های ۱۰۰۰ واژه‌ای اول تا پنجم کتاب‌ها و نیز پیکره‌ی زبان‌آموزی حتی به ۶۰۰ واژه، یعنی پوشش ۶۰ درصدی هر یک از گروه‌های ۱۰۰۰ واژه‌ای پیکره‌ی مبنا هم نمی‌رسد. در واقع، میزان هم‌پوشانی واژه‌های هر فهرست با فهرست پیکره‌ی مبنا، از حدود ۵۵۰ واژه شروع شده و تا ۱۹ واژه‌ی هم‌پوشان، یعنی کمتر از ۲٪ نیز تقلیل می‌یابد.

برای تعیین آن که کتاب‌های مورد بررسی، به لحاظ واژه‌آموزی حداقل چه تعداد از ۵۰۰۰ واژه‌ی پُرسامد زبان فارسی را پوشش داده و برای واژه‌آموزی هدف‌گذاری و ارائه کرده‌اند، هر یک از گروه‌های واژگانی طبقه‌بندی‌شده نه فقط با گروه متناظر با آن در فهرست واژگانی پیکره‌ی مبنا، بلکه با همه‌ی ۶۰۰۰ واژه‌ای

^۱ functional word

واژه‌هایی مانند حروف اضافه، حروف ربط، قیده‌های ربط و... که دارای محتوای معنایی مشخص نیستند، بلکه یک نقش دستوری را در زبان بر عهده دارند. در فهرست‌های واژگانی مربوط به زبان طبیعی، اکثر واژه‌های نقشی در گروه اول فهرست‌های بسامدی قرار می‌گیرند.

که پُربسامدترین واژه‌ها در فهرست واژگانی پیکره‌ی مبنا محسوب می‌شوند، مقایسه شدند. نتایج این بررسی مقایسه‌ای در جدول (۵) زیر نشان داده شده است.

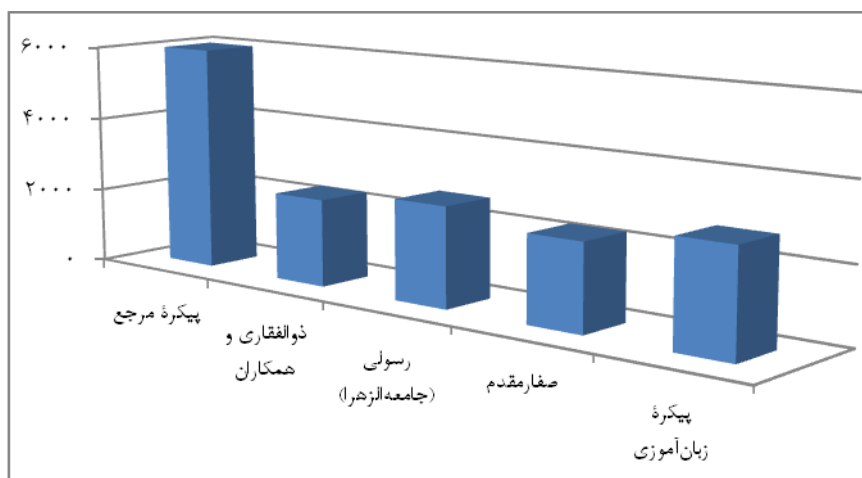


نمودار ۳. مقایسه‌ی میزان همپوشانی هر یک از گروه‌های ۱۰۰۰ واژه‌ای کتاب‌ها و پیکره‌ی زبان‌آموزی با پیکره‌ی مبنا (محور عمودی تعداد واژه‌های همپوشان با گروه‌های ۱۰۰۰ واژه‌ای پیکره‌ی مبنا را نشان می‌دهد)

جدول ۵. میزان و درصد همپوشانی واژه‌های کتاب‌ها و پیکره‌ی زبان‌آموزی با پیکره‌ی مبنا (بدون احتساب سطوح یادگیری در واژه‌آموزی)

ردیف	سطح یادگیری	ذوالفقاری و همکاران		رسولی (جامعه‌الزهرای)		سفارمقدم		پیکره‌ی زبان‌آموزی	
		تعداد	درصد	تعداد	درصد	تعداد	درصد	تعداد	درصد
۱	گروه اول	۸۴۰	۸۴	۹۲۳	۹۲/۳	۷۳۶	۷۲/۶	۸۸۱	۸۸/۱
۲	گروه دوم	۶۲۲	۶۲/۲	۵۸۹	۵۸/۹	۶۷۰	۶۷	۶۴۰	۶۴
۳	گروه سوم	۲۹۳	۲۹/۳	۴۰۹	۴۰/۹	۳۰۵	۳۰/۵	۴۲۱	۴۲/۱
۴	گروه چهارم	۲۳۶	۲۳/۶	۲۹۹	۲۹/۹	۲۵۶	۲۵/۶	۳۵۷	۳۵/۷
۵	گروه پنجم	۱۸۷	۱۸/۷	۲۵۰	۲۵	۲۰۶	۲۰/۶	۲۳۷	۲۳/۷
۶	گروه ششم	۱۴۵	۱۴/۵	۱۷۴	۱۷/۴	۱۰۹	۱۰/۹	۲۱۰	۲۱
۷	مجموع ^۱	۲۳۲۳	۳۹/۷	۲۶۴۴	۴۵/۲	۲۲۸۲	۳۹	۲۷۴۶	۴۶/۹

^۱ با توجه به این که ۱۴۰ واژه از ۶۰۰۰ واژه‌ی پیکره‌ی مبنا به جای موارد خطا قرار داده شد، درصدهای مربوط به ردیف ۷ به جای آن که بر مبنای ۶۰۰۰ واژه محاسبه گردد، بر مبنای ۵۸۶۰ واژه محاسبه شده است.



نمودار ۴. میزان همپوشانی واژگانی کتاب‌ها و پیکره‌ی زبان‌آموزی با واژگان پیکره‌ی مبنا

ردیف هفتم جدول (۵)، یعنی مجموع هم‌پوشانی واژگانی کتاب‌ها و پیکره‌ی زبان‌آموزی با پیکره‌ی مبنا، مشخص می‌کند که هیچ یک از کتاب‌های آموزش زبان فارسی به غیرفارسی‌زبانان حتی نیمی از واژه‌های پُرکاربرد زبان فارسی (۳۰۰۰ واژه از ۶۰۰۰ واژه) را نیز ارائه نکرده‌اند. جالب آن است که پیکره‌ی زبان‌آموزی که واژه‌های تشکیل‌دهنده‌ی آن حاصل واژه‌های همه‌ی کتاب‌های مورد بررسی می‌باشند نیز در این مورد از وضعیت مشابهی برخوردار است. به عبارت دقیق‌تر، این کتاب‌ها و پیکره‌ی زبان‌آموزی حاصل از آنها، بیش از ۵۰٪ از واژه‌های پُرکاربرد زبان فارسی را مورد بی‌توجهی قرار داده‌اند. نمودار (۴) به سادگی روشن‌گر این کاستی کتاب‌ها و به‌طور کلی محتوای آموزشی این حوزه است.

۷. نتیجه‌گیری و بحث

در این پژوهش، با هدف بررسی وضعیت واژه در کتاب‌های آموزش زبان فارسی به غیرفارسی‌زبانان و طرح این پرسش که این کتاب‌ها تا چه اندازه نیازهای واژگانی زبان‌آموزان را پوشش می‌دهند، محتوای آموزشی کتاب‌های ذوالفقاری و همکاران (۱۳۸۷) صفا مقدم (۱۳۸۶) و رسولی (۱۳۹۱)، به‌عنوان نماینده‌ی کتاب‌های مورد استفاده در داخل ایران (برای آموزش زبان فارسی به غیرفارسی‌زبانان) تحلیل شد. به این ترتیب که برای هر کدام یک پیکره‌ی زبانی ایجاد شد. آن‌گاه از ترکیب این سه پیکره نیز یک پیکره‌ی زبان‌آموزی تشکیل شد. فهرست واژگانی-بسامدی هر یک از این چهار پیکره‌ی زبانی تنظیم و ۵۰۰۰ واژه‌ی پربسامد هر فهرست به پنج گروه ۱۰۰۰ واژه‌ای تقسیم شد. این گروه‌های ۱۰۰۰ واژه‌ای پنج‌گانه، به‌عنوان سطوح یک تا پنج واژه‌آموزی در نظر گرفته شد. سپس بر پایه‌ی متون برگزیده‌ای از پایگاه داده‌های زبان فارسی و با رعایت معیارهای طراحی و ایجاد یک پیکره‌ی متوازن، پیکره‌ی دیگری به‌عنوان پیکره‌ی مبنا ایجاد و فرض شد که

پُرسامدترین واژه‌های این پیکره‌ی مبنا، تأمین‌کننده‌ی نیازهای واژگانی فارسی‌آموزان هستند. بنابراین، برای رسیدن به پاسخ پرسش پژوهش، فهرست‌های واژگانی هر یک از کتاب‌ها و نیز پیکره‌ی زبان‌آموزی با فهرست واژگانی بسامدی پیکره‌ی مبنا مورد مقایسه قرار گرفت.

نتایج پژوهش نشان داد که به لحاظ سطوح واژه‌آموزی، هم‌پوشانی میان پنج گروه ۱۰۰۰ واژه‌ای کتاب‌ها و پیکره‌ی مبنا، در گروه اول که به دلیل وجود واژه‌های نقشی در آن، انتظار هم‌پوشانی بسیار بالایی وجود داشت، در بهترین حالت حدود ۵۰٪ بود. در گروه‌های دیگر، کتاب‌ها گاهی صرفاً ۲٪ از واژه‌های پُرسامد زبان فارسی را به زبان‌آموزان معرفی کرده بودند. همچنین، یافته‌ها نشان داد که بدون در نظر گرفتن سطوح واژه‌آموزی، این کتاب‌ها در تأمین نیازهای واژگانی زبان‌آموزان ناکارآمد خواهند بود، به‌طوری که هیچ کدام از آنها حتی نیمی از واژه‌های پُرکاربرد زبان فارسی را نیز پوشش نداده‌اند و می‌توان گفت با این که در یک دوره‌ی آموزشی، معمولاً فقط یکی از این کتاب‌ها با زبان‌آموز کار خواهد شد، ولی حتی مجموع این کتاب‌ها (پیکره‌ی زبان‌آموزی) نیز کمتر از ۵۰٪ نیازهای واژگانی او را تأمین می‌کند. بر این اساس، در پاسخ به پرسش پژوهش مبنی بر این که کتاب‌های آموزش زبان فارسی به غیرفارسی‌زبانان تا چه اندازه نیازهای واژگانی زبان‌آموزان را پوشش می‌دهد، می‌توان گفت محتوای آموزشی موجود، چه از نظر سطوح واژه‌آموزی و چه حتی بدون در نظر گرفتن سطوح واژه‌آموزی و صرفاً از نظر معرفی واژه‌های پُرکاربرد زبان فارسی، در تأمین نیازهای واژگانی زبان‌آموزان وضعیت نامطلوبی دارند.

بنابراین، ضرورت دارد که برنامه‌ریزان و مؤلفان کتاب‌های حوزه‌ی آموزش زبان فارسی به غیرفارسی‌زبانان با توجه به این نیازهای واژگانی نسبت به بازنگری جدی کتاب‌های مذکور اقدام نمایند.

در پژوهش حاضر، تلاش شد با پرهیز از پیچیدگی‌های نظری نشان داده شود که چگونه می‌توان از طریق بررسی‌های پیکره‌ای و آماری بر روی محتوای واژگانی کتاب‌های آموزش زبان به افزایش کمی دانش واژگانی زبان‌آموزان کمک کرد.

یافته‌های این پژوهش می‌تواند در سطوح تدوین برنامه‌ی درسی، تولید محتوای آموزشی و همچنین فعالیت‌های کلاسی در حوزه‌ی آموزش زبان به کار بسته شود و از آنجا که در این زمینه برای زبان فارسی خلأ فراوانی نیز دیده می‌شود، پژوهش حاضر ضمن آن که راهنمایی خواهد بود برای متخصصان برنامه‌ی درسی در حوزه‌های مرتبط با آموزش زبان، مبنایی خواهد بود برای تولید محتوای آموزشی و به‌طور دقیق‌تر، تألیف کتاب‌های درسی، آموزشی و کمک‌آموزشی، در تمام حوزه‌هایی که به آموزش زبان و به‌ویژه آموزش زبان فارسی مربوط می‌شوند.

تشکر و قدردانی

از همکاری و توجه محبت‌آمیز دکتر مسعود قیومی، عضو هیأت علمی گروه زبان‌شناسی پژوهشگاه علوم انسانی و مطالعات فرهنگی تشکر و قدردانی می‌نماییم. ایشان در نوبت‌ها و روزهای مکرر و با شناخت و دسترسی کاملی که به پایگاه داده‌های زبان فارسی داشتند، همه‌ی متن‌های پایگاه داده‌ها را به آسانی برای انجام هر چه دقیق‌تر این پژوهش فراهم آوردند. همچنین، لازم است از خانم دکتر سیده فاطمه رسولی، مؤلف محترم مجموعه‌ی *آموزش زبان فارسی* جامعه‌الزهرا و آقای یوسفی که فایل‌های این کتاب را در اختیار پژوهشگران قرار دادند، تقدیر و تشکر می‌شود. آقای مجتبی احمدی، مدیرعامل مؤسسه‌ی فرهنگی و انتشاراتی ایماژ و خانم‌ها صادقیان و ذوالقدر نیز زحمت حروف‌نگاری مجدد متن سایر کتاب‌ها را پذیرفتند که بدین وسیله از ایشان تشکر و قدردانی می‌شود.

مراتب سپاس و قدردانی قلبی و بی‌پایان خود را تقدیم آقای دکتر مایک اسکات (Mike Scott)، پژوهشگر مدعو دانشگاه استون^۱ می‌نماییم. در زمانی که به علت تحریم‌های شدید اقتصادی علیه ایران امکان خرید و استفاده‌ی کامل از نرم‌افزار *ورداسمیت* در اختیار پژوهشگران داخل کشور وجود نداشت، ایشان سخاوتمندانه امکان یک اشتراک رایگان از نرم‌افزار خود را برای پژوهشگران فراهم آوردند. شایان ذکر است که نسخه‌ی شماره‌ی هفت این نرم‌افزار در تحلیل داده‌های زبان فارسی با مشکلات جدی مواجه بود. خوشبختانه طی مکاتبات متعددی که نویسنده‌ی مسئول مقاله، با دکتر اسکات، صاحب‌امتیاز حقیقی این نرم‌افزار انجام داد، این مشکلات مرتفع شد. در این مکاتبات، ویژگی‌های زبانی و همچنین نوشتاری زبان فارسی، طی ساعاتی از چند روز متوالی برای ایشان تشریح شد تا زمینه‌ی انتشار نسخه‌های جدیدی از این نرم‌افزار بسیار قدرتمند را فراهم آورد. در واقع، تعاملات و هم‌فکری‌های انجام گرفته، در ابتدا به انتشار نسخه‌ی ۷,۰۰,۰۱۰ و در نهایت، انتشار نسخه‌ی ۷,۰۰,۰۱۱ منجر شد که خوشبختانه با زبان فارسی سازگاری پیدا کرد. خرسندیم که در انتشار این نسخه‌ی جدید و خدمت به انجام پژوهش‌های دقیق‌تر بر روی زبان فارسی نقش کوچکی ایفا کردیم. در پایان، از داوران گمنام *پژوهش‌نامه‌ی آموزش زبان فارسی به غیر فارسی‌زبانان* که با ارائه‌ی پیشنهادهای ارزشمند خود بر غنای مقاله افزودند، تشکر می‌شود.

^۱. visiting research fellow at Aston university

منابع:

- احمدی‌قادر، ش. (۱۳۹۰). آموزش مستقیم واژه در متن: مقایسه‌ی تأثیر در رویکرد یادگیری مستقیم و تصادفی در یادگیری واژه. پایان‌نامه‌ی کارشناسی ارشد. دانشگاه علامه طباطبایی.
- بی‌جن‌خان، م.، نصری، ع. و جلایی، ش. (۱۳۹۳). نقش واژگان بسامدی در ارزیابی مهارت واژگانی فارسی‌آموزان. پژوهش‌نامه‌ی آموزش زبان فارسی به غیرفارسی‌زبانان، سال سوم، شماره ۲ (پیاپی ۷)، صص: ۲۵-۴۵.
- تبریزمنش، م. (۱۳۸۷). بررسی مجموعه‌ی «فارسی بیاموزیم: آموزش زبان فارسی» از منظر فرانتش تجربی، براساس چارچوب نظری دستور نقش‌گرای هلیدی. رساله‌ی کارشناسی ارشد. دانشگاه شیراز.
- داوری اردکانی، ن. و آقابراهیمی، ه. (۱۳۹۱). ارزیابی اصالت متون سه کتاب آموزش زبان فارسی به غیرفارسی‌زبانان از دیدگاه انسجام. پژوهش‌نامه‌ی آموزش زبان فارسی به غیرفارسی‌زبانان، سال اول، شماره ۲، صص: ۱۵۷-۱۶۸.
- ذوالفقاری، ح.، غفاری، م. و محمودی بختیاری، ب. (۱۳۸۷). فارسی بیاموزیم؛ دوره‌ی مقدماتی (جلد اول-سوم). چاپ سوم. تهران: انتشارات مدرسه.
- رسولی، س. ف. (۱۳۹۱). آموزش زبان فارسی (جلد اول-چهارم). قم: مرکز آموزش زبان فارسی جامعه‌الزهرا.
- رضاخانی، ف. (۱۳۸۵). بررسی مجموعه‌ی «فارسی بیاموزیم: آموزش زبان فارسی» از منظر فرانتش بینافردی و براساس چارچوب نظری دستور نقش‌گرای هلیدی. رساله‌ی کارشناسی ارشد. تهران: دانشگاه شهید بهشتی.
- رضاقلی‌فامیان، ع. (۱۳۹۲). آغاز و پایانه‌ی کلامی در درس‌نامه‌های آموزش زبان فارسی به غیرفارسی‌زبانان. پژوهش‌نامه‌ی آموزش زبان فارسی به غیرفارسی‌زبانان، سال دوم، شماره ۱، صص: ۹۶-۱۱۲.
- رضایی، ح. و علی‌پور، م. (۱۳۹۱). بررسی متون خوانداری مجموعه‌ی «فارسی بیاموزیم» براساس نقش‌های هفت‌گانه‌ی زبان از دیدگاه هلیدی. پژوهش‌نامه‌ی آموزش زبان فارسی به غیرفارسی‌زبانان، سال اول، شماره ۲، صص: ۱۶۳-۱۷۹.
- سعیدی، ز. و شرفی‌نژاد، ح. (۱۳۹۲). طراحی یک بسته‌ی الکترونیکی برای بهبود آموزش و سنجش واژگان پایه‌ی زبان فارسی. پژوهش‌نامه‌ی آموزش زبان فارسی به غیرفارسی‌زبانان، سال دوم، شماره ۱، صص: ۱۳۲-۱۵۰.
- صحرای، ر. و شهباز، م. (۱۳۹۱). تحلیل محتوایی منابع آموزشی زبان فارسی بر پایه‌ی انگاره‌ی نیل اندرسون. پژوهش‌نامه‌ی آموزش زبان فارسی به غیرفارسی‌زبانان، سال اول، شماره ۲، صص: ۱-۳۰.
- عاصی، س. م. (۱۳۷۲). طرح ایجاد پایگاه داده‌های زبان فارسی با کمک رایانه. در پژوهشگران، شماره ۸، تهران: مؤسسه‌ی مطالعات و تحقیقات فرهنگی، صص: ۱۴-۱۸.
- عبادی، س.، وکیلی‌فرد، ا. و بهراملو، خ. (۱۳۹۳). تدوین فهرست واژگان پایه برای زبان فارسی: رویکردی تلفیقی. پژوهش‌نامه‌ی آموزش زبان فارسی به غیرفارسی‌زبانان، سال سوم، شماره ۳ (پیاپی ۸)، صص: ۳-۲۳.
- غفاری‌مهر، م. (۱۳۸۳). بررسی دو مجموعه آموزشی آزفا و فارسی بیاموزیم از دیدگاه نقش‌گرای. رساله‌ی کارشناسی ارشد. تهران: دانشگاه شهید بهشتی.
- وکیلی‌فرد، ا. (۱۳۷۸). کدام زبان فارسی را به غیرفارسی‌زبانان آموزش دهیم؟ نامه‌ی پارسی، سال ۴، شماره ۳، صص: ۲۱۲-۲۱۹.

- Assi, M.** (1994). *Persian Language Data Base (PLDB)*©Assi. Retrieved from: [www.http://pldb.ihcs.ac.ir](http://pldb.ihcs.ac.ir).
- Atkins, B. T. S. & Rundell, M.** (2008). *The Oxford Guide to Practical Lexicography*. New York: Oxford University Press.
- Aijmer, K.** (2009). *Corpora and Language Teaching*. Studies in Corpus Linguistics 33. Amsterdam/Philadelphia: John Benjamins.
- Criado, R. & Sánchez, A.** (2012). Vocabulary Learning: The Word Frequency Factor in Textbooks and in the Teacher's Implementation of Activities in the Classroom. *22nd Annual Conference of the European Second Language Association*. Adam Mickiewicz University.
- Facchinetti, R.** (2007). *Corpus Linguistics: 25 Years on*. NY, Radopi.
- Jimenez, R. M. & Mancebo, R.** (2008). Vocabulary Input in EFL Textbooks. *RESLA*, 21: 147-165.
- Milton, J.** (2009). *Measuring Second Language Vocabulary Acquisition*. Bristol: Multilingual Matters.
- Nation, I. S. P.** (2006). How Large a Vocabulary Is Needed for Reading and Listening? *Canadian Modern Language Review*, 63, 59-82.
- Nation, I. S. P.** (2014). How Much Input Do You Need to Learn the Most Frequent 9,000 Words? *Reading in a Foreign Language*, 26(2), 1-16.
- Sinclair, J. M.** (1987). Introduction. In J. Sinclair (Ed.), *Looking Up*. London: Collins Publication and the University of Birmingham.
- Sinclair, J. M.** (1991). *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- Sinclair, J.** (2004). Introduction. In J. Sinclair (Ed.), *How to Use Corpora in Language Teaching* (pp. 1-10). Amsterdam, Netherlands: John Benjamins.
- Scott, M.** (2015). *WordSmith Tools: Windows Software for Finding Word Patterns* © M. Scott: Lexical Analysis Software Ltd.
Retrieved from: <http://www.lexically.net/wordsmith/>.

پیوست

نوع یا قالب متن‌های مندرج در پیکره‌ی مرجع، تعداد هر قالب و تعداد واژه‌های آن

ردیف	نوع یا قالب متن	تعداد متن	تعداد واژه‌ها (نمونه‌واژه‌ها)
۱	مقاله	۵	۲۱۶۶۱
۲	کتاب	۴۳	۱۷۸۴۸۲۱
۳	کتاب‌های ویژه‌ی کودکان	۱۰	۱۲۴۱۶
۴	متن‌های گفتاری خودمانی	۳۰	۱۲۲۸۹۵
۵	متن‌های گفتاری (علم و زندگی، خانه و خانواده، کندوکاو، محاوره)	۴	۳۰۱۲۸
۶	روزنامه‌ی عمومی	۱	۷۴۱۶۲۰
۷	آموزشی	۱	۵۰۱۴۲
۸	ادبی	۱	۲۷۲۶۲۵
۹	اقتصادی	۱	۳۷۵۷۷۶
۱۰	علمی	۱	۲۸۸۶۳
۱۱	هنری	۱	۱۸۹۴۵۱
۱۲	ورزشی	۱	۳۸۳۷۰۵
۱۳	دینی	۱	۲۸۳۲۲
۱۴	اجتماعی	۱	۳۸۰۳۴۶
۱۵	سیاسی	۱	۳۵۷۸۲۳
۱۶	تاریخی	۱	۲۱۵۰۱۴