

پژوهش‌نامه‌ی آموزش زبان فارسی به غیر فارسی‌زبانان

(علمی-پژوهشی)

سال سوم، شماره‌ی سوم (پیاپی ۸)، پاییز و زمستان ۱۳۹۳

## تدوین فهرست واژگان پایه برای زبان فارسی: رویکردی تلفیقی

سامان عبادی

استادیار گروه آموزش زبان انگلیسی - دانشگاه رازی

امیررضا وکیلی فرد

استادیار گروه آموزش زبان فارسی - دانشگاه بین‌المللی امام خمینی<sup>(ره)</sup>

خسرو بهراملو

دانشجوی دکتری آموزش زبان انگلیسی - دانشگاه رازی

### چکیده

بخشی از سرمایه‌ی زمانی فراگیران زبان فارسی به یادگیری واژگان فارسی اختصاص داده می‌شود. برای این که این سرمایه‌گذاری، بیشترین بازگشت سرمایه را در پی داشته باشد، لازم است محتوای آموزشی به گونه‌ای طراحی شود که مبتنی بر واژگان پرکاربرد موجود در قالب فهرست واژگان پایه باشد. برای تدوین فهرست واژگان پایه فارسی، ضروری است مفهوم «واژه» مطابق با انگاره‌ی ذهنی گویشوران عادی این زبان تعریف شود تا بتوان به کمک آن به بسامدشماری مصادیق یک واژه در پیکره پرداخت. از سوی دیگر، علاوه بر بسامد باید به «پوشش متنی» و «گستره‌ی واژگان پایه» در پیکره‌ی مبدأ و پیکره‌های دیگر نیز توجه نمود. در این پژوهش، برداشت معتدلانه‌ای از «خانوار واژگانی» برای تعیین مصادیق یک واژه پیشنهاد شده است و براساس آن، پوشش متنی و گستره‌ی واژگان فرهنگ بسامدی بی‌جن‌خان و محسنی (۱۳۹۱) در یک پیکره‌ی ۳۴۰۰۰ کلمه‌ای مورد بررسی قرار گرفته است. نتایج این پژوهش نشان داد که شکل اصلاح‌شده‌ی فرهنگ بی‌جن‌خان و محسنی، ۸۴ درصد از کلمات پیکره‌ی ما را پوشش می‌دهد و واژگان‌ش از گستره‌ی مطلوبی برخوردارند. بنابراین، می‌توان با اصلاح این فرهنگ به تدوین فهرست واژگان پایه‌ی فارسی اقدام کرد. می‌توان از این فهرست در امر هدف‌گذاری برای آموزش زبان فارسی و سطح‌بندی فراگیران و محتوا استفاده نمود. این ابزار می‌تواند تهیه‌ی آزمون‌های دانش واژگانی و توانش فارسی را تسهیل نموده، در شناسایی واژگان علمی و تخصصی فارسی نیز به ما کمک کند.

**کلیدواژه‌ها:** سرواژگی، خانوار واژگانی، رویکرد تلفیقی، پوشش متنی، گستره‌ی واژگان

تاریخ پذیرش نهایی مقاله: ۱۳۹۳/۱۱/۲۸

samanebadi@gmail.com

تاریخ دریافت مقاله: ۱۳۹۳/۸/۲۶

رایانشانی نویسنده مسئول (سامان عبادی):

## ۱. مقدمه

یادگیری واژگان یک زبان، فرایند پیچیده‌ای است که در آن، فراگیر باید هم شکل و هم معانی مختلف یک کلمه را بیاموزد. تعیین نقطه‌ی شروع مناسب برای این فرایند از اهمیت بسزایی برای فراگیران غیربومی برخوردار است. از طرف دیگر، با توجه به این که فراگیران زبان فارسی وقت محدودی برای یادگیری این زبان در اختیار دارند، محتوای آموزشی باید به گونه‌ای طراحی شود که بیشترین بازگشت سرمایه را برای آنها در پی داشته باشد. یکی از حوزه‌هایی که این امر در آن از اهمیت بالایی برخوردار است، حوزه‌ی واژگان می‌باشد.

نیشن (۲۰۱۱) چنین بیان می‌کند که بررسی هزینه-فایده یکی از ملاحظات مهم در آموزش و یادگیری واژگان می‌باشد. هزینه‌ی مالی و زمانی که معلمان و فراگیران صرف آموزش و یادگیری واژگان می‌کنند، باید سود مناسبی برای آنها داشته باشد. در حوزه‌ی واژه‌آموزی، این سود در میزان کاربرد کلمه (پرکاربردی/پرتکراری) نمود می‌یابد. این بدین معنی است که هزینه‌ی مالی و زمانی باید صرف اکتساب واژگانی شود که در متون مختلف از بسامد بالایی برخوردارند. این جا است که استفاده از فرهنگ‌های لغات پایه می‌تواند راهگشا باشد.

این واژه‌نامه‌ها، کلمات بسیار پرتکرار یک زبان را بر پایه‌ی بسامد واژگان در متون متعددی که در زمینه‌های مختلف به تحریر درآمده‌اند، شناسایی و فهرست‌بندی می‌کنند (نیشن و ورینگ، ۱۹۹۷؛ نیشن، ۲۰۱۱b؛ بگلر و هانت، ۲۰۰۵؛ کارتر، ۱۹۹۸). فراگیران می‌توانند مستقیماً از این فرهنگ‌ها استفاده کنند و آموزگاران و مؤلفین کتاب‌های درسی نیز می‌توانند از آنها برای انتخاب مواد آموزشی مناسب برای گروه خاصی از دانش‌آموزان بهره‌گیرند. علاوه بر این، می‌توان از واژه‌نامه‌های لغات پایه برای تدوین فرهنگ‌های لغات تخصصی نظیر واژه‌نامه‌های علمی یا فنی نیز بهره‌برد. در این حالت، واژگان پایه به‌عنوان مبنایی برای شناسایی لغات تخصصی‌تر عمل می‌کنند (نیشن و هوانگ، ۱۹۹۵). با آن که از اوایل دهه‌ی ۵۰ تلاش‌هایی برای شناسایی لغات پایه‌ی زبان فارسی به عمل آمده است، هنوز فهرستی از واژگان پایه‌ی فارسی در دسترس نیست (باقری، ۱۳۸۹). این مقاله سعی دارد مشکلات موجود در این زمینه را بررسی نماید و راهکارهایی را برای رفع آنها ارائه نماید.

اگرچه پیکره‌های خوبی برای زبان فارسی تهیه شده است و براساس آنها، چند فرهنگ بسامدی تدوین شده است (مانند حسنی، ۱۳۸۴ و بی‌جن‌خان و محسنی، ۱۳۹۱)، نبود تعریف علمی و دقیق مفهوم «کلمه» که با توانش زبانی یک «فارسی‌زبان متوسط» متناسب باشد، مانع از تدوین فهرست لغات پایه‌ی فارسی براساس این فرهنگ‌های بسامدی شده است. بنابر یافته‌های برخی پژوهشگران، گویشوران عادی یک زبان نه با همه‌ی فرایندهای اشتقاقی زبان خود آشنا هستند (گویشور فرهیخته)، و نه کلاً با آن بیگانه‌اند (نوآموز)؛

بلکه دانش لغوی آنها چیزی بین این دو غایت است. یعنی گویشوران بومی فرایندهای اشتقاقی عمده در زبان خود را می‌شناسند و به کار می‌برند. به گفته‌ی این پژوهشگران، در تعریف مفهوم «واژه»، باید گنجینه‌ی واژگانی این نوع گویشوران را مبنا قرار داد (نگی و همکاران، ۱۹۸۹؛ باور و نیشن، ۱۹۹۳؛ کلزن و همکاران، ۲۰۰۳).

برای تعیین و شمارش مصادیق یک کلمه در پیکره، از دو مفهوم «خانوار واژگانی»<sup>۱</sup> و «سرواژه»<sup>۲</sup> استفاده می‌شود. در رویه‌ی «خانوار واژگانی»، یک کلمه، شامل یک بن‌واژه‌ی ساده یا مرکب، تمامی اشکال تصریفی آن، کلمات مشتق از آن بن‌واژه و اشکال تصریفی این مشتق‌ها است. در رویکرد «سرواژگی»، یک کلمه شامل یک بن‌واژه‌ی ساده، مرکب، مشتق، یا مرکب-مشتق به همراه اشکال تصریفی آن بن‌واژه است (نیشن و میرا، ۲۰۰۲؛ باور و نیشن، ۱۹۹۳).

تحقیقات نشان داده است که گویشوران متوسط یک زبان با فرایندهای اشتقاقی عمده در زبان خود آشنا هستند و مشتقات حاصل از این فرایندها را جزو بن‌واژه‌ی مبنا به حساب می‌آورند. این نوع گویشوران به دلیل عدم آشنایی با فرایندهای اشتقاقی غیرعمده، نمی‌توانند مشتقات حاصل از این فرایندها را به اجزای کوچکتر تجزیه کنند و در نتیجه آنها را بن‌واژه‌های مستقل تلقی می‌کنند (نگی و همکاران، ۱۹۸۹؛ باور و نیشن، ۱۹۹۳؛ کلزن و همکاران، ۲۰۰۳).

با در نظر داشتن این یافته و اعمال رویه‌های خانوار واژگانی و سرواژگی در مورد کلمات زبان فارسی، می‌توان انتظار داشت که رویکرد خانوار واژگانی به دلیل گستردگی حوزه‌ی شمول یک کلمه، با دانش لغوی فرهیختگان فارسی‌زبان تناسب داشته باشد، ولی رویکرد سرواژگی نمایانگر دانش لغوی فارسی‌زبانانی است که به دلیل سطح پایین توانش زبانی، شناختی از فرایند اشتقاق و ترکیب در زبان فارسی ندارند و هر کلمه‌ی ساده، مشتق، مرکب، یا مرکب-مشتقی را به همراه اشکال تصریفی‌اش، یک واژه‌ی مستقل می‌دانند.

برای تهیه‌ی فهرست واژگان پایه‌ی یک زبان، توصیه می‌شود توانش زبانی گویشور متوسط آن زبان به‌عنوان مبنای شناسایی و تعیین مصادیق یک کلمه در نظر گرفته شود (نگی و همکاران، ۱۹۸۹؛ باور و نیشن، ۱۹۹۳؛ کلزن و همکاران، ۲۰۰۳). این مسأله مستلزم آن است که برای تعیین مصادیق یک کلمه در پیکره، رویه‌ای بینابین دو رویکرد موجود اتخاذ شود. در این حالت، یک کلمه علاوه بر شمول یک بن‌واژه‌ی ساده یا مرکب و همه‌ی اشکال تصریفی آن، کلمات مشتق از آن بن‌واژه (و تصریفات آنها) را نیز در برمی‌گیرد؛ ذکر این نکته ضروری است که کلمه، همه‌ی کلمات مشتق از آن بن‌واژه را شامل نمی‌شود بلکه فقط آن دسته از بن‌واژه‌هایی را در برمی‌گیرد که از افزودن پربسامدترین، باثبات‌ترین و زیابترین تکواژه‌های اشتقاقی به آن بن‌واژه به وجود می‌آیند (هرش و نیشن، ۱۹۹۲؛ باور و نیشن، ۱۹۹۳؛ کلزن و همکاران،

1. word family  
2. lemmatization

۲۰۰۳). در این صورت، مصادیق و اشکال یک کلمه، بی‌اندازه پرشمار (خانوار واژگانی) و یا بسیار کم‌شمار (سرواژه) نخواهد بود.

در تهیه‌ی فرهنگ‌های بسامدی فارسی، رویه‌ی میانه‌ای برای تعیین مصادیق یک کلمه ابداع نشده است و در بیشتر موارد، تکیه بر رویکرد سرواژگی بوده است (بی‌جن‌خان و محسنی، ۱۳۹۱). در برخی موارد، حتی این رویکرد نیز به درستی به‌کار نرفته است. اگر این چنین نبود، در فهرست حسنی (۱۳۸۴)، صرف‌های مختلف فعل «کردن»، هر یک به عنوان مدخلی مجزا با بسامدی مستقل محسوب نمی‌شدند.

گذشته از این، بنابر گفته‌های صاحب‌نظران عرصه‌ی یادگیری واژگان، در تدوین فهرست واژگان پایه، توجه هم‌زمان به سه مقوله‌ی «بسامد»، «پوشش متنی» و «گستره‌ی واژگان» ضروری است. «بسامد» به تعداد وقوع یک کلمه (و مصادیق آن) در پیکره اشاره دارد. «پوشش متنی»، درصدی از کل کلمات متون پیکره یا متون دیگر است که تحت پوشش بسامد یک یا چند کلمه‌ی خاص قرار می‌گیرد. معمولاً، متون هر پیکره براساس ژانر و حوزه‌ی علمی به ده‌ها گروه دسته‌بندی می‌شوند. «گستره»، به پربسامد بودن یک کلمه در همه یا بیشتر ژانرهای پیکره اشاره دارد (نیشن و هوانگ، ۱۹۹۵؛ نیشن و ورینگ، ۱۹۹۷). در زبان فارسی، فرهنگ‌های بسامدی (حسنی، ۱۳۸۴ و بی‌جن‌خان و محسنی، ۱۳۹۱) عمدتاً براساس بسامد تهیه شده‌اند و پوشش متنی و گستره‌ی واژگان از نظرها دور مانده است. با این حال، در تهیه‌ی فهرست لغات پایه‌ی یک زبان باید به هر سه مقوله‌ی بسامد، پوشش متنی و گستره‌ی واژگان توجه داشت.

پژوهش حاضر تلاش کرده است راهکارهایی را برای رفع این مشکلات ارائه نماید و زمینه را برای پژوهش‌های آینده هموار سازد. در این راستا، برای تعیین و شمارش مصادیق یک کلمه در پیکره، پیشنهاد می‌شود از رویکردی که مابین رویه‌های خانوار واژگانی و سرواژگی است، استفاده شود (هرش و نیشن، ۱۹۹۲؛ باور و نیشن، ۱۹۹۳؛ کلارن و همکاران، ۲۰۰۳). در این رویکرد، زیرمجموعه‌ی یک بن‌واژه را اینها تشکیل می‌دهند: همه‌ی کلمات مشتقی که از اضافه شدن پربسامدترین، باثبات‌ترین و زیایاترین تکواژهای اشتقاقی به یک بن‌واژه‌ی ساده یا مرکب به وجود می‌آیند و همچنین اشکال تصریفی آنها. پربسامدترین، باثبات‌ترین و زیایاترین تکواژهای اشتقاقی زبان فارسی را می‌توان با مراجعه به فهرست واژگان زایای فارسی که توسط اسلامی و همکاران (۱۳۸۳) تهیه شده است، شناسایی کرد. البته، برای این منظور از نتایج پژوهش بی‌جن‌خان (۱۳۸۶) نیز می‌توان کمک گرفت.

پس از تهیه‌ی فهرستی جامع از تکواژهای اشتقاقی فارسی، می‌توان بسامد هر تکواژ را در فهرست اسلامی و همکاران (۱۳۸۳) به دست آورد و سپس پربسامدترین تکواژها را از نظر ثبات (معنایی، نقشی، و شکلی)<sup>۱</sup> و

1. regularity (of meaning, function, and form)

زیایی<sup>۱</sup> مورد بررسی قرار داد. اگر یک تکواژ اشتقاقی در همه‌ی کلمات مشتقی که در ساختارشان به کار می‌رود، دارای معنا، نقش، یا شکل ثابتی باشد، دارای ثبات معنایی، نقشی یا شکلی است. وقتی یک تکواژ اشتقاقی توانایی افزوده شدن به بن‌واژه‌های متعدد و ساخت مشتق‌های جدید را داشته باشد، زیا به حساب می‌آید (باور و نیشن، ۱۹۹۳). با این کار، آن دسته از تکواژهای اشتقاقی که دارای هر سه مؤلفه‌ی پرسامدی، باثباتی، و زیایی هستند شناسایی می‌شوند. آن‌گاه، می‌توان این تکواژها را به بن‌واژه‌های ساده و مرکب در فهرست بسامدی بی‌جن‌خان و محسنی (۱۳۹۱) اضافه کرد و کلمات مشتقی به وجود آورد. از افزودن این کلمات مشتق و اشکال تصریفی‌شان به زیرمجموعه‌ی یک بن‌واژه‌ی ساده یا مرکب، بر تعداد مدخل‌های فهرست بی‌جن‌خان و محسنی (۱۳۹۱) افزوده نمی‌شود، ولی تعداد کلماتی که زیرمجموعه‌ی هر مدخل هستند افزایش می‌یابد. بدین ترتیب، فهرست اصلاح شده‌ای به دست می‌آید که انتظار می‌رود هر مدخل آن با دانش واژگانی یک گویشور متوسط زبان فارسی مطابقت داشته باشد.

با این حال، هنوز این فهرست اصلاح شده را نمی‌توان فهرست لغات پایه به شمار آورد؛ زیرا پوشش متنی و گستره‌ی واژگان آن آزموده نشده است. برای این منظور، می‌توان چند پیکره‌ی کوچک در ژانرهای مختلف را انتخاب کرد و بسامد و پوشش متنی هر واژه از این فهرست را در هر ژانر جداگانه محاسبه نمود. اگر مجموع پوشش متنی کلمات این فهرست در هر پیکره بالاتر از ۸۰ درصد باشد، فهرست واژگان پایه، یکی دیگر از آزمون‌هایش را با موفقیت پشت سر گذاشته است (هرش و نیشن، ۱۹۹۲؛ نیشن و هوانگ، ۱۹۹۵؛ نیشن و ورینگ، ۱۹۹۷؛ نیشن، ۲۰۱۵). اگر کلمات این فهرست در همه یا بیشتر ژانر-پیکره‌ها از بسامد بالایی برخوردار باشند، واژگان آن دارای گستره نیز می‌باشند. پوشش متنی بالا و گستره‌ی مطلوب برای لغات موجود در این فهرست، این اطمینان را به وجود می‌آورد که فهرست اصلاح شده در واقع، همان فهرست لغات پایه برای زبان فارسی است.

## ۲. تعریف کلمه

برای تهیه‌ی فهرست واژگان پایه‌ی یک زبان، نیاز است مفهوم «کلمه» متناسب با ویژگی‌های خاص آن زبان تعریف و تبیین شود. در بسیاری از موارد، آنچه را که ما «کلمه» می‌خوانیم، خود از عناصری تشکیل شده است که هر یک از آنها، بخشی از معنای کل کلمه را با خود حمل می‌کنند. به کوچکترین واحدهای زبان‌شناختی معنادار، از نظر ساخت‌واژه<sup>۲</sup> تکواژ<sup>۳</sup> می‌گویند که به دو دسته‌ی آزاد و وابسته تقسیم می‌شوند.

1. productivity  
2. morphology  
3. morpheme

تکواژه‌های آزاد<sup>۱</sup> دارای معنای مستقل هستند و می‌توانند به‌عنوان کلمات مستقل عمل نمایند مانند «کتاب، خانه، از، به». تکواژه‌های آزاد دو نوع دارند: واژگانی<sup>۲</sup> و دستوری/نقشی<sup>۳</sup>. تکواژه‌های واژگانی شامل اسم، صفت، فعل، و قید می‌شوند و عمده‌ی بار معنایی جملات را بر دوش می‌کشند مانند «کتاب و خانه». اهالی زبان هیچ محدودیتی در ابداع این گونه تکواژه‌ها ندارند و تعداد آنها می‌تواند افزایش یابد. تکواژه‌های دستوری شامل حروف ربط، اضافه، تعریف و ضمائر می‌شوند که بین محتوای معنایی کلمات و جملات ارتباط برقرار می‌کنند. تعداد این تکواژه‌ها محدود می‌باشد و تقریباً هرگز افزایش نمی‌یابد مانند «از، به». تکواژه‌های وابسته<sup>۴</sup> نمی‌توانند به تنهایی و مستقل به‌عنوان کلمه به‌کار برده شوند و باید به تکواژه‌های آزاد یا کلمات دیگر افزوده شوند مانند «دار، ای، ها، تر». تعداد این تکواژه‌ها نیز محدود می‌باشد. در زبان فارسی، تکواژه‌های وابسته به صورت پیشوند، پسوند یا میان‌وند به تکواژه‌های آزاد و کلمات دیگر افزوده می‌شوند. این تکواژه‌ها بر دو نوع هستند: اشتقاقی<sup>۵</sup> و تصریفی<sup>۶</sup>. با افزودن تکواژه‌های اشتقاقی به کلمات دیگر، اغلب رده‌ی دستوری آن کلمات تغییر می‌کند و کلمات جدیدی به وجود می‌آیند مانند «ای» در کلمات «رسا و خانگی». با افزودن تکواژه‌های تصریفی به کلمات دیگر، تغییری در رده‌ی دستوری آن کلمات به وجود نمی‌آید و کلمات جدیدی شکل نمی‌گیرند، بلکه جنبه‌هایی از نقش دستوری آن کلمات معلوم می‌گردد. این جنبه‌ها می‌تواند شامل صرف جمع، زمان، شخص و قیاس صفت باشد مانند «ها، می، ام، تر» در کلمات «کتابخانه‌ها، می‌رود، رفتم، خوب‌تر» (براساس یول، ۱۹۹۶).

فرایندهای واژه‌سازی نقش بسزایی در زایایی و پویایی یک زبان دارند. ترکیب، اشتقاق، و ترکیب-اشتقاق از مهمترین فرایندهای واژه‌سازی به شمار می‌آیند. در فرایند ترکیب، از کنار هم گذاشتن دو یا چند تکواژ آزاد یک بن‌واژه‌ی مرکب به وجود می‌آید مثل «کتابخانه و بادب». در فرایند اشتقاق، با افزودن یک یا چند تکواژ اشتقاقی به یک تکواژ آزاد، بن‌واژه‌ی مشتق شکل می‌گیرد مانند «پوشش» و «زندگی». لازم به ذکر است که ما هر یک از بن‌های ماضی و مضارع یک فعل را تکواژ آزاد در نظر گرفته‌ایم، چون می‌توانند تکواژ اشتقاقی بگیرند (براساس یول، ۱۹۹۶). در فرایند ترکیب-اشتقاق، کلمه‌ای به وجود می‌آید که هم مرکب است و هم مشتق. در این حالت، معمولاً به یک بن‌واژه‌ی مرکب، یک یا چند تکواژ اشتقاقی افزوده می‌شود تا یک بن‌واژه‌ی مرکب-مشتق به وجود آید (مک آرتور، ۲۰۱۵)؛ مانند صفت «کتابخانه‌ای» و اسم «سرزندگی».

1. free morpheme

2. lexical morpheme

3. functional morpheme

4. bound morpheme

5. derivational morpheme

6. inflectional morpheme

برای تعیین و شمارش مصادیق یک کلمه در تهیه‌ی واژگان پایه برای زبان انگلیسی از مفهوم خانوار واژگانی استفاده می‌شود. یک خانوار واژگانی شامل یک بن‌واژه‌ی ساده یا مرکب، اشکال تصریفی آن بن‌واژه، کلمات مشتق از آن بن‌واژه و اشکال تصریفی آن مشتق‌ها است. در این رویکرد، فرض بر این است که دانستن معنای بن‌واژه و تکواژهای اشتقاقی و تصریفی عامل بر آن، خواننده را قادر می‌سازد تا معنی همه کلمات مشتق و تصریفی حاصل را به راحتی استخراج کند و در نتیجه، همه‌ی آنها را به‌عنوان مصادیق آن بن‌واژه در ذهن خود ثبت نماید (نیشن و میرا، ۲۰۰۲؛ باور و نیشن، ۱۹۹۳). شواهدی وجود دارد که نشان می‌دهد خانوار واژگانی، واحدی مهم در ساختار «واژگان ذهنی»<sup>۱</sup> گویشوران مادری یک زبان است. البته این بدین معنی نیست که گویشوران عادی یک زبان با همه‌ی فرایندهای اشتقاقی آن زبان آشنایی کامل دارند (نگی و همکاران، ۱۹۸۹). از این‌رو، برژینا و گابلاسوا (۲۰۱۳) پیشنهاد کرده‌اند که از مفهوم سرواژگی برای تعیین و شمارش مصادیق یک کلمه در پیکره استفاده شود. یک سرواژه شامل یک بن‌واژه‌ی ساده، مرکب، مشتق، یا مرکب-مشتق به همراه همه‌ی اشکالی است که از افزودن تکواژهای تصریفی به آن به‌وجود آمده‌اند و همگی در یک رده‌ی دستوری قرار دارند. در واقع، در رویکرد سرواژگی، بن‌واژه‌های مشتق و مرکب-مشتق از زیرمجموعه‌ی یک خانوار واژگانی حذف می‌شوند و هر یک به همراه تصریفات خود به‌عنوان یک مدخل یا سرواژه‌ی مجزا به حساب می‌آیند (نیشن و میرا، ۲۰۰۲؛ باور و نیشن، ۱۹۹۳).

در حالی که رویکرد خانوار واژگانی، کلمات مشتق زیادی را تحت پوشش یک مدخل قرار می‌دهد و با دانش ساخت‌واژی گویشوران عادی یک زبان تطابق ندارد، رویکرد سرواژگی کل کلمات مشتق را بدون هیچ ملاحظه‌ای از زیرمجموعه‌ی یک خانوار واژگانی خارج می‌کند و به این ترتیب، آشنایی گویشوران عادی یک زبان با فرایند اشتقاق را صفر تلقی می‌کند. در واژگان ذهنی گویشوران عادی یک زبان، بین کلماتی که از نظر اشتقاقی با هم مرتبط هستند روابط نزدیکی وجود دارد؛ هرچند ممکن است این روابط به گستردگی و قوت روابطی که در واژگان ذهنی گویشوران فرهیخته‌ی آن زبان وجود دارد، نباشد. این بدین معنی است که رویکرد خانوار واژگانی رویه‌ای حداکثری است و دانش ساخت‌واژی گویشوران عادی یک زبان را دست‌بالا برآورد می‌کند، ولی رویکرد سرواژگی به‌عنوان رویه‌ای حداقلی، دانش ساخت‌واژی این گویشوران را دست‌کم می‌گیرد. این جا است که اتخاذ «رویکرد بینابینی» ضروری به نظر می‌رسد (نگی و همکاران، ۱۹۸۹؛ باور و نیشن، ۱۹۹۳؛ کلازن و همکاران، ۲۰۰۳).

باور و نیشن (۱۹۹۳) بر این عقیده‌اند که تکواژهای اشتقاقی و تصریفی را می‌توان براساس معیارهایی چون بسامد، ثبات در شکل و نقش، زایایی، و قابل‌پیش‌بینی بودن معنا در هفت سطح طبقه‌بندی کرد. به کمک این طبقه‌بندی می‌توان تعیین کرد گویشورانی که در سطوح مختلف دانش ساخت‌واژی قرار دارند، چه

1. mental lexicon

کلماتی را جزو یک خانوار واژگانی حساب می‌کنند. فراگیران بسیار مبتدی که در سطح یک قرار دارند، هر یک از کلمات مشتق از یک بن‌واژه و حتی اشکال تصریفی خود بن‌واژه را به‌عنوان یک کلمه‌ی کاملاً جدید در نظر می‌گیرند و قادر به درک رابطه‌ی بین این کلمات و بن‌واژه نیستند. فراگیران سطح دو، اشکال تصریفی یک بن‌واژه را می‌شناسند و آنها را جزو آن بن‌واژه به شمار می‌آورند. گویشورانی که در سطح سه قرار دارند، علاوه بر تکواژهای تصریفی مؤثر بر یک بن‌واژه، پربسامدترین، باثبات‌ترین، و زیاترین تکواژهای اشتقاقی مؤثر بر آن بن‌واژه را نیز می‌شناسند و کلمات حاصل از افزودن آنها به بن‌واژه را نیز جزو آن بن‌واژه محسوب می‌کنند. محاسبات باور و نیشن (۱۹۹۳) نشان می‌دهد که این سطح از آشنایی با تکواژهای اشتقاقی و تصریفی زبان انگلیسی، تعداد خانوارهای واژگانی یک پیکره‌ی یک میلیون لغتی را از ۶۱۰۰۰ به ۳۴۰۰۰ کاهش می‌دهد؛ در حالی که افزودن دیگر تکواژهای اشتقاقی سطوح چهار تا هفت، تأثیر چشم‌گیری بر کاهش تعداد خانوارهای واژگانی این پیکره ندارد. بر پایه‌ی این محاسبات، چنین به نظر می‌رسد که این سطح از آشنایی با تکواژهای اشتقاقی و تصریفی برای استفاده در تعیین و شمارش مصادیق یک کلمه در پیکره، مناسب است؛ زیرا این سطح چیزی بین رویکرد خانوار واژگانی و سرواژگی است و با توجه به مطالب گفته شده برای استفاده در تدوین واژه‌نامه‌ی پایه، مطلوب به نظر می‌رسد.

نتایج تحقیق کلان و همکاران (۲۰۰۳) نشان می‌دهد که هر یک از کلمات مشتق از یک بن‌واژه دارای مدخل مجزایی در نظام زبانی هستند و تنها بخشی از این کلمات مشتق که حاصل اعمال فرایندهای اشتقاقی گویا و زایا هستند، به صورت تجزیه‌پذیر ذخیره شده‌اند و می‌توانند جزو زیرمجموعه‌ی بن‌واژه‌ی مادر قرار گیرند.

در زبان فارسی، رویه‌ی بینابینی برای تعیین مصادیق یک کلمه در پیکره وجود ندارد. در تهیه‌ی فرهنگ‌های بسامدی فارسی، عمدتاً از رویکرد سرواژگی برای تعیین و شمارش کلمات استفاده شده است (رجایی و علیزاده، ۱۳۸۹؛ بی‌جن‌خان و محسنی، ۱۳۹۱). در فرهنگ بسامدی حسنی (۱۳۸۴)، حتی این رویکرد نیز به شکل مطلوب اجرا نشده است و هر یک از اشکال تصریفی یک بن‌واژه در مدخلی جدا آمده است؛ به گونه‌ای که فهرست حاصل شامل ۸۴۳۸ مدخل است. بنابراین، برای استخراج واژگان پایه‌ی فارسی باید ابتدا رویه‌ی میانه‌ای متناظر با آنچه که در بند بالا گفتیم، اتخاذ شود. آنچه در بند زیر می‌آید، روش پیشنهادی ما برای رفع این مشکل است.

براساس یک پیکره‌ی ده میلیون کلمه‌ای، اسلامی و همکاران (۱۳۸۳) موفق شدند پس از حذف اشکال تصریفی کلمات، ۴۴۰۰۰ بن‌واژه را برای زبان فارسی شناسایی کنند. آنها توانستند با استفاده از یک فرهنگ لغت به‌عنوان مکمل، فهرست خود را به ۵۵۰۰۰ بن‌واژه افزایش دهند که تقریباً معادل تعداد بن‌واژه‌های انگلیسی در مطالعه‌ی باور و نیشن (۱۹۹۳) است. در اقدامی مشابه، بی‌جن‌خان (۱۳۸۶) از یک پیکره‌ی ده



میلیون لغتی حدود ۶۰۰۰۰ بن‌واژه استخراج کرد. اگر فهرست جامعی از تکواژهای اشتقاقی زبان فارسی تهیه شود، می‌توان بسامد هر یک از این تکواژها را در بن‌واژه‌های فهرست اسلامی و همکاران (۱۳۸۳) یا بی‌جن‌خان (۱۳۸۶) محاسبه کرد. در رویکردی متفاوت و البته بسیار وقت‌گیر، عرفانیان قنسولی و شریفی (۱۳۹۰) سعی کردند با بررسی فراوانی تکواژهای اشتقاقی موجود در لغات محتوایی ۲۰ صفحه متن، زیاترین تکواژهای اشتقاقی را در زبان فارسی شناسایی کنند. پس از شناسایی پربسامدترین تکواژهای اشتقاقی، باید ثبات شکلی، نقشی و معنایی و همچنین زایایی این تکواژها را نیز به صورت دستی بررسی کرد. با این کار، تکواژهای پربسامدی که باثبات و زایا هستند استخراج می‌شوند.

حال با استفاده از این تکواژهای پربسامد، باثبات و زایا می‌توان فرهنگ بسامدی بی‌جن‌خان و محسنی (۱۳۹۱) را مورد بازبینی قرار داد. برای این کار، به هر یک از ۱۸۶۸ مدخل این فرهنگ مراجعه می‌کنیم و این نکته را بررسی خواهیم کرد که کدام یک از تکواژهای پربسامد، باثبات و زایا را می‌توان به بن‌واژه‌ی آن مدخل خاص اضافه کرد و از آن کلمه‌ی مشتق ساخت. آنگاه کلمات مشتق از آن بن‌واژه را به همراه اشکال تصریفی‌شان به زیرمجموعه‌ی آن مدخل اضافه می‌کنیم. از آن‌جا که تعریف بی‌جن‌خان و محسنی (۱۳۹۱) از کلمه مبتنی بر مفهوم سرواژه بوده است، مشتقات برخی بن‌واژه‌ها به‌عنوان مدخل مستقل در این فرهنگ آمده‌اند. اگر این کلمات مشتق با افزودن تکواژهای پربسامد، باثبات و زایا به بن‌واژه‌های ساده یا مرکب به وجود آمده باشند، باید آنها را حذف کنیم و به زیرمجموعه‌ی بن‌واژه‌ی ساده یا مرکب بیفزاییم.

در فرهنگ بی‌جن‌خان و محسنی (۱۳۹۱)، اشکال تصریفی بن ماضی و بن مضارع یک فعل، زیر یک مدخل واحد آمده‌اند. به نظر ما این رویه مناسب نیست؛ زیرا بن ماضی و بن مضارع یک فعل از نظر شکلی با هم فرق دارند و هر یک می‌توانند تکواژهای اشتقاقی خاص خود را بگیرند و بن‌واژه‌های مشتق متفاوت به وجود آورند. بنابراین، آنها تکواژهای آزاد به شمار می‌روند و باید هر یک به‌عنوان یک بن‌واژه‌ی ساده، مدخلی خاص داشته باشند که علاوه بر اشکال تصریفی آنها، شامل مشتقات آنها نیز باشد. دلیل دیگر ما برای این پیشنهاد آن است که اگر بن ماضی و مضارع یک فعل، زیر یک مدخل باشند و بخواهیم علاوه بر اشکال تصریفی، مشتقات هر دو بن را نیز اضافه کنیم، مدخل مربوطه بیش از اندازه حجیم می‌شود و ممکن است با واژگان ذهنی گویشوران عادی فارسی مطابقت نداشته باشد.

با اعمال تغییرات بالا در فرهنگ بی‌جن‌خان و محسنی (۱۳۹۱)، انتظار می‌رود تعداد مدخل‌ها افزایش محسوسی نکند؛ زیرا حدوداً تنها ۶۰ فعل در این فرهنگ وجود دارد. با اختصاص مدخل مستقل به بن‌های مضارع، ۶۰ مدخل به این فرهنگ افزوده می‌شود. در عوض، در حجم برخی از مدخل‌های مربوط به بن‌واژه‌های ساده یا مرکب تغییر عمده‌ای اتفاق می‌افتد. البته، مدخل‌های این فهرست اصلاح‌شده، نه بسیار حجیم خواهند شد و نه بسیار تُنک. تصور بر این است که حجم متوسطی که برای هر مدخل به دست

می‌آید، با دانش واژگانی گویشوران عادی زبان فارسی تطابق داشته باشد. بنابراین، این فهرست جدید را می‌توان به‌عنوان فهرست لغات پایه‌ی زبان فارسی به شمار آورد.

### ۳. بسامد، پوشش متنی<sup>۱</sup> و گستره‌ی<sup>۲</sup> واژگان پایه

نیشن و هوانگ (۱۹۹۵) و نیشن و ورینگ (۱۹۹۷) معتقد هستند که پس از این که فرهنگ بسامدی یک زبان تهیه شد و تغییرات لازم مطابق با سطح سوم طبقه‌بندی باور و نیشن (۱۹۹۳) بر آن اعمال شد، باید پوشش متنی و گستره‌ی واژگان این فهرست اصلاح‌شده را در متون پیکره‌ی مبدأ و متون دیگر نیز بررسی کرد. اگر این فهرست، پوشش متنی مناسب داشت و از گستره‌ی واژگانی مطلوبی برخوردار بود، آن‌گاه می‌توان آن را فهرست واژگان پایه نامید. در زبان فارسی، بیشتر پژوهشگران تلاش کرده‌اند تا فرهنگ‌های بسامدی مناسبی در عرصه‌ی لغت تدوین کنند و از پوشش متنی و گستره‌ی واژگان پربسامد در پیکره‌ی مبدأ و متون دیگر غافل مانده‌اند. البته تهیه‌ی فرهنگ‌های بسامدی، از این جهت که بستر لازم را برای تدوین واژگان پایه‌ی یک زبان فراهم می‌کند، اقدامی پسندیده و ستودنی است.

برای استخراج فرهنگ بسامدی، پربسامدترین کلمات یک پیکره را انتخاب می‌کنند. درصدی از کل کلمات پیکره‌ی مبدأ که تحت پوشش بسامد کلمات پرتکرار قرار می‌گیرد، پوشش متنی آن کلمات را تشکیل می‌دهد. معمولاً، برای آن که بتوان یک فرهنگ بسامدی را به‌عنوان فهرست واژگان پایه در نظر گرفت، باید تغییراتی را در آن مطابق با سطح سوم طبقه‌بندی باور و نیشن (۱۹۹۳) ایجاد کرد و سپس بسامد واژگان آن را در متونی غیر از متون پیکره‌ی مبدأ محاسبه کرد. اگر این فهرست، ۸۰ الی ۸۵ درصد از کلمات متون دیگر را هم پوشش دهد، می‌توان گفت که فهرست از اعتبار و صحت قابل‌قبولی برخوردار است و اطلاق فهرست واژگان پایه به آن قوت بیشتری می‌گیرد (هرش و نیشن، ۱۹۹۲؛ نیشن و هوانگ، ۱۹۹۵؛ نیشن و ورینگ، ۱۹۹۷؛ نیشن، ۲۰۱۵).

«گستره‌ی واژه» در واقع، پربسامد بودن یک کلمه در همه یا بیشتر ژانرهای پیکره‌ی مبدأ می‌باشد. از آن‌جا که فرهنگ‌های بسامدی و به دنبال آنها فهرست واژگان پایه از یک پیکره‌ی خاص استخراج می‌شوند، ممکن است از ویژگی‌های خاص آن پیکره تأثیر پذیرفته باشند و کلمات‌شان نماینده‌ی مناسبی از کلمات متون عمومی یک زبان نباشند. این مسأله به دسته‌بندی متون پیکره مربوط می‌شود. معمولاً متون یک پیکره براساس ژانر نوشتاری به چندین گروه از جمله خبری، ورزشی، علمی و... تقسیم می‌شوند. ممکن است کلمه‌ای که پربسامد بوده و جزو کلمات پایه آمده است، جزو کلمات کلیدی یک ژانر خاص باشد و در آن

1. text coverage

2. range

ژانر به صورتی غیرمعمول از بسامد بسیار بالایی برخوردار باشد. در این حالت، بسامد بالای آن کلمه در کل پیکره مدیون بسامد غیرمعمول آن کلمه در یک ژانر خاص می‌باشد. حال ممکن است با حذف آن ژانر خاص از پیکره، کلمه‌ی مذکور بسامد بالای خود را در کل پیکره از دست بدهد و از فهرست واژگان پربسامد خارج شود. در نتیجه، واژگان فرهنگ‌های بسامدی را باید در متون هر یک از ژانرهای پیکره دوباره بسامدشماری کرد تا معلوم شود آیا این کلمات در همه یا بیشتر ژانرهای پیکره از بسامد بالا برخوردار هستند یا خیر. اگر این واژگان از بسامد بالایی در همه یا بیشتر ژانرهای پیکره برخوردار باشند، می‌توان گفت که دارای گستره‌ی مطلوبی هستند. برای افزایش اعتبار فهرست اصلاح شده‌ای که از فرهنگ بسامدی حاصل شده است و به‌نوعی پیش‌نویس فهرست واژگان پایه است، می‌توان گستره‌ی کلمات آن را در پیکره‌ی مبدأ یا متونی غیر از پیکره‌ی مبدأ مورد آزمون قرار داد. اگر واژگان این فهرست از بسامد بالایی در همه یا بیشتر این متون برخوردار باشند، دارای گستره‌ی مطلوب می‌باشند و این فهرست، آزمون نهایی خود را با موفقیت گذرانده است و می‌توان با اطمینان کافی آن را «فهرست واژگان پایه» نامید (نیشن و هوانگ، ۱۹۹۵؛ نیشن و ورینگ، ۱۹۹۷).

در میان فهرست‌های بسامدی که در دسترس ما بودند، رجایی و عزیزاده (۱۳۸۹) فقط به بسامد اکتفا کرده بودند و در مورد پوشش متنی واژگان پربسامد و گستره‌ی این واژگان سخنی نگفته بودند. حسنی (۱۳۸۴) نیز در گزارش خود پوشش متنی واژگان پربسامد را محاسبه نکرده بود و واژه‌های دارای بسامد بیش از ۱۰ بار را در فهرست خود آورده بود. چون تعداد کلمات پیکره‌ی حسنی ذکر شده است، می‌توان پوشش متنی فرهنگ بسامدی او را محاسبه کرد. پوشش متنی این فرهنگ در متون خارج از پیکره محاسبه نشده است؛ البته او بسیار گذرا به این حقیقت اشاره کرده بود که بسامد وقوع برخی کلمات در متن یا متون خاصی بسیار بالا بوده است و آنها را در فهرست خود با حروف ریزتر و تیره‌تر و داخل دو گیومه مشخص کرده بود. او با این کار، بسامد بالا و غیرطبیعی برخی کلمات را در یک یا چند ژانر خاص از ۸۰ کلان‌متن<sup>۱</sup> پیکره نشان داده بود. این نوع پرتکراری یک کلمه در یک ژانر خاص ممکن است آن کلمه را در کل پیکره نیز پربسامد جلوه دهد و در نهایت، آن را به گزاف در فهرست واژگان پایه بگنجانند. از این اقدام چنین برمی‌آید که او حداقل به کلماتی که احتمالاً فاقد گستره‌ی مطلوب هستند، مشکوک شده است و آنها را در فهرست خود نشانه‌دار کرده است.

بی‌جن‌خان و محسنی (۱۳۹۱) نیز پوشش متنی فرهنگ بسامدی خود را گزارش نکرده‌اند و کلمات دارای بسامد بیش از ۵۰۰ بار را در فهرست خود آورده‌اند. البته از آنجا که تعداد کلمات پیکره (۹۸۸۰۴۰۰) و تعداد مصادیق ۱۸۶۸ واژه‌ی این فرهنگ در کل پیکره (۶۸۰۷۰۶۶) آمده است می‌توان

۱. حسنی (۱۳۸۴) از اصطلاح «کلان‌متن» به جای «ژانر» استفاده کرده است.

پوشش متنی این فرهنگ را محاسبه کرد (۶۸،۸۹ درصد). پوشش متنی این فرهنگ در متون خارج از پیکره مورد بررسی قرار نگرفته است. این فرهنگ، بسامد هر کلمه در کل پیکره را ذکر کرده است ولی بسامد هر کلمه در ژانرهای مختلف پیکره را جداگانه ذکر نکرده است تا بتوان گستره‌ی کلمات را مورد آزمایش قرار داد. از آن‌جا که در متن گزارش نیز گستره‌ی کلمات ذکر نشده است، می‌توان چنین استنباط کرد که به گستره‌ی واژگان در تهیه‌ی این فرهنگ توجهی نشده است. برای رفع این مشکل باید چاره‌ای اندیشید. بنابراین، به نظر نویسندگان مقاله‌ی حاضر، فرایند زیر می‌تواند این مشکل را رفع کند.

پس از آن که تغییرات لازم در فرهنگ بی‌جن‌خان و محسنی (۱۳۹۱) به عمل آمد و مطابق با سطح سوم طبقه‌بندی باور و نیشن (۱۹۹۳)، مشتقات بن‌واژه‌های ساده و مرکب در زیر آنها افزوده شد، در مرحله‌ی بعد باید پوشش متنی و گستره‌ی واژگان این فهرست جدید را مورد آزمایش قرار داد. به این منظور، می‌توان چند پیکره‌ی کوچک - که هر یک متعلق به یک ژانر نوشتاری مختلف هستند - انتخاب کرد و بسامد واژگان پایه را در هر یک از آنها به دست آورد. پوشش متنی واژگان پایه، به درصدی از کل کلمات یک پیکره‌ی متنی گفته می‌شود که تحت پوشش این واژگان پایه قرار می‌گیرند. برای محاسبه‌ی پوشش متنی واژگان پایه در یک پیکره‌ی خاص، بسامد هر کلمه‌ی پایه در آن پیکره را در عدد ۱۰۰ ضرب می‌کنند و حاصل را بر تعداد کلمات پیکره تقسیم می‌کنند و نتیجه را در ستونی جدا مقابل آن کلمه ثبت می‌کنند. عدد به دست آمده نشان می‌دهد چند درصد از کلمات آن پیکره را مصادیق یک کلمه‌ی خاص پوشش می‌دهد. سپس با تجمیع پوشش متنی همه‌ی کلمات فهرست، پوشش متنی واژگان پایه در آن پیکره‌ی به‌خصوص به دست می‌آید. انتظار می‌رود پوشش متنی فهرست اصلاح‌شده به ۸۰ تا ۸۵ درصد کلمات پیکره‌های مختلف برسد (هرش و نیشن، ۱۹۹۲؛ نیشن و هوانگ، ۱۹۹۵؛ نیشن و ورینگ، ۱۹۹۷؛ نیشن، ۲۰۱۸). برای بررسی گستره‌ی واژگان پایه، باید بسامد هر کلمه‌ی پایه در یک متن را با بسامد همان کلمه در متون دیگر مقایسه کرد. با در نظر گرفتن تعداد کلمات هر پیکره، اگر یک کلمه‌ی پایه در همه یا بیشتر پیکره‌ها بسامد بالایی داشت، می‌توان گفت که آن کلمه دارای گستره است. پس از طی موفقیت‌آمیز این مرحله، فهرست اصلاحی را می‌توان رسماً «فهرست لغات پایه‌ی فارسی» نام نهاد.

#### ۴. روش پژوهش

همان‌طور که در قسمت‌های بالا گفتیم، برای تهیه‌ی فهرست واژگان پایه‌ی فارسی، لازم است پرسامدترین، باثبات‌ترین، و زایاترین تکواژهای اشتقاقی فارسی شناسایی شوند. به کمک این تکواژها، می‌توان مشتقات گویای بن‌واژه‌های ساده و مرکب را شناسایی کرد و به ذیل آن بن‌واژه‌ها انتقال داد (زیر آن بن‌واژه‌ها نوشت). با اعمال این فرایند بر مدخل‌های فرهنگ بسامدی بی‌جن‌خان و محسنی (۱۳۹۱)، می‌توان فهرست واژگان

پایه‌ی فارسی را تهیه کرد. از آن‌جا که تاکنون پربسامدترین، باثبات‌ترین و زیاترین تکواژهای اشتقاقی فارسی شناسایی نشده‌اند و از سوی دیگر، پژوهشگران حاضر نیز فاقد دانش فنی و امکانات مالی لازم برای انجام این کار بودند، راه‌حل میانبری برای عملیاتی کردن این بازبینی به‌کار گرفته شد که در این قسمت به صورت مفصل در مورد آن توضیح داده می‌شود.

جهت بررسی پوشش متنی واژگان پایه، ۲۱ متن ۱۰۰۰ الی ۱۵۰۰ کلمه‌ای از تارنمای «پیکره‌ی فارسی (۱۳۹۳)»<sup>۱</sup> استخراج نموده و همه‌ی آنها را به صورت یک متن پیوسته در قالب نرم‌افزار واژه‌پرداز<sup>۲</sup> نگارش کردیم؛ اکنون، می‌توانیم این مجموعه را «پیکره‌ی کلی» بنامیم. این پیکره‌ی کلی طبق برآورد نرم‌افزار بسامدشمار، شامل ۳۴۱۵۴ کلمه بود که با کسر علائم نگارشی، اعداد رقیمی و حروف به ۳۳۴۶۴ کلمه رسید. متن این پیکره را به یک نرم‌افزار بسامدشمار دادیم تا بسامد واژگان متن به دست آید و به ترتیب الفبایی فهرست شود و در یک فایل ذخیره شود. سپس، از اولین مدخل فهرست الفبایی بی‌جن‌خان و محسنی (۱۳۹۱) که کلمه‌ی «آب» بود، شروع کردیم و با رجوع به فهرست الفبایی پیکره‌ی کلی، کلمه‌ی «آب» و اشکال تصریفی‌اش را پیدا کردیم و بسامد همه‌ی آنها را به صورت مجموع در روبه‌روی مدخل «آب» در فرهنگ بی‌جن‌خان و محسنی (۱۳۹۱) وارد نمودیم. در مورد مدخل‌های دیگر نیز همین کار را تکرار کردیم. چون فهرست منتج از پیکره‌ی کلی به صورت الفبایی بود، مشتقات یک بن‌واژه در زیر آن بن‌واژه قرار می‌گرفتند. اگر در بررسی کلمه‌ی مشتق از یک بن‌واژه، براساس شمّ زبانی خود بدین نتیجه می‌رسیدیم که خود تکواژ اشتقاقی موجود در آن مشتق، جزو پربسامدترین، باثبات‌ترین، و زیاترین تکواژهای فارسی است، بسامد آن مشتق و اشکال تصریفی‌اش را نیز به بسامد بن‌واژه‌ی مادر در فرهنگ اضافه می‌کردیم.<sup>۳</sup> بدین صورت، رویه‌ی میانه‌ای را که برای تعیین مصادیق یک کلمه در پیکره پیشنهاد کرده بودیم، عملیاتی نمودیم. با تجمیع بسامد به دست آمده برای کلمات این فرهنگ، پوشش متنی واژگان پایه را در پیکره‌ی کلی به دست آوردیم. برای این کار، بسامد به دست آمده برای کل کلمات فرهنگ را در ۱۰۰ ضرب کردیم و آن را بر تعداد کلمات پیکره که ۳۳۴۶۴ واژه بود، تقسیم کردیم.

برای بررسی گستره‌ی واژگان پایه، چهار پیکره‌ی کوچک‌تر را از دل پیکره‌ی کلی استخراج کردیم و نام آنها را پیکره‌ی ادبیات داستانی، ادبیات غیرداستانی، حقوقی، و متنوع نهادیم. هر یک از این پیکره‌ها شامل ۵۲۰۰ کلمه بود و از پنج متن ۱۰۰۰ الی ۱۵۰۰ کلمه‌ای تشکیل شده بود. برای بررسی گستره‌ی کلمات

<sup>۱</sup> بارگیری از [www.corpus.ir](http://www.corpus.ir) در تاریخ ۱۳۹۳/۱۰/۸

<sup>۲</sup> Microsoft Office Word

<sup>۳</sup> در بخش تعریف کلمه توضیح دادیم که «رویکرد تلفیقی» تاکنون در فارسی به‌کار بسته نشده است و از طرف دیگر، کاربری آن نیازمند دانش رایانه‌ای و یک تحقیق مفصل و جداگانه‌ای است که نتایج آن در این تحقیق نمی‌گنجد. در نتیجه، رویکرد تلفیقی را براساس شمّ زبانی خود عملیاتی نموده‌ایم.

پایه، می‌بایست بسامد هر کلمه‌ی پایه در یک پیکره را با بسامد همان کلمه در پیکره‌های دیگر مقایسه می‌کردیم. به دلیل محدودیت پژوهش، بسامد ۶۳ کلمه‌ی بالای فهرست بسامدی بی‌جن‌خان و محسنی (۱۳۹۱) را در چهار پیکره‌ی نامبرده به دست آوردیم و بسامد هر کلمه در یک پیکره را با بسامد همان کلمه در پیکره‌های دیگر مقایسه کردیم.

معیار خاصی برای تعیین گستره‌ی کلمات تعریف نشده است. نیشن و هوانگ (۱۹۹۵) گستره‌ی مطلوب برای یک کلمه‌ی پرسامد را حضور آن کلمه در ۱۰ ژانر از ۱۵ ژانر پیکره‌ی مبدأ (دو سوم ژانرهای پیکره‌ی مبدأ) تعریف کرده‌اند. همچنین، به نظر می‌رسد نیازی به برابر بودن بسامد یک کلمه در ژانرهای مختلف نیست و حضور نسبی کفایت می‌کند (نیشن و هوانگ، ۱۹۹۵؛ نیشن و ورینگ، ۱۹۹۷). با توجه به حجم بسیار کوچک پیکره‌ها، از حضور نسبی چشم‌پوشی کرده و صرف حضور را مبنا قرار دادیم. بنابراین، برای هر یک از واژگان منتخب، گستره‌ی مطلوب را حداقل یک بار حضور در سه پیکره از چهار پیکره‌ی بالا تعریف کردیم. بدین ترتیب، اگر کلمه‌ای در هیچ یک از پیکره‌ها حضور نداشت دارای گستره‌ی صفر می‌شد و در صورت حضور در یک، دو، سه، یا چهار پیکره به ترتیب با گستره‌ی یک، دو، سه، یا چهار رتبه‌بندی می‌شد.

## ۵. یافته‌های پژوهش

نتایج به دست آمده از بسامدشماری نشان داد که ۲۸۰۸۸ کلمه از کلمات پیکره‌ی کلی، تحت پوشش فهرست اصلاح‌شده‌ی بی‌جن‌خان و محسنی (۱۳۹۱) قرار دارد. به بیان دیگر، فهرست اصلاح‌شده‌ی بی‌جن‌خان و محسنی ۸۳،۹۳ درصد از کلمات پیکره‌ی کلی را تحت پوشش قرار می‌دهد. همان‌طور که در جدول (۱) می‌توان دید، از ۶۳ کلمه‌ی منتخب، ۵۷ کلمه دارای گستره‌ی مطلوب سه و چهار هستند (یعنی در سه یا چهار پیکره حداقل یک بار حضور داشته‌اند)؛ یک کلمه دارای گستره‌ی صفر است (یعنی در هیچ پیکره‌ای ذکر نشده است)؛ دو کلمه دارای گستره‌ی یک هستند و تنها در یک پیکره حضور داشته‌اند؛ سه کلمه نیز دارای گستره‌ی دو هستند که حضور آنها در دو پیکره را نشان می‌دهد. ردیف‌های مربوط به این شش کلمه‌ی دارای گستره‌ی نامطلوب در جدول (۱) با رنگ تیره مشخص شده‌اند.

جدول ۱. بسامد واژگان پایه در پیکره‌های ادبیات داستانی، ادبیات غیرداستانی، حقوقی و متنوع

رتبه‌ی بسامدی	واژه	ادبیات داستانی	ادبیات غیر داستانی	حقوقی	متنوع
۱	و	۲۳۱	۳۶۰	۴۳۹	۲۲۷
۲	در	۴۸	۱۳۸	۱۵۹	۱۲۱
۳	به	۱۲۳	۱۲۱	۱۵۴	۱۵۹
۴	از	۱۱۰	۱۵۴	۱۱۷	۱۱۲
۵	که	۱۴۲	۱۵۷	۷۹	۱۳۳
۶	بودن	۱۸۶	۲۰۰	۱۳۵	۱۳۲
۷	این	۶۴	۷۳	۴۳	۶۶
۸	را	۱۶۴	۱۲۲	۶۴	۹۵
۹	کردن	۱۳۰	۱۱۴	۶۳	۱۱۸
۱۰	شدن	۶۴	۶۳	۴۶	۱۱۵
۱۱	با	۵۴	۸۷	۴۷	۵۰
۱۲	آن	۳۲	۴۵	۵۳	۶۴
۱۳	داشتن	۵۱	۵۳	۳۶	۶۳
۱۴	برای	۲۴	۲۳	۳۴	۳۱
۱۵	خود	۳۳	۲۹	۲۱	۳۸
۱۶	دادن	۳۳	۲۰	۱۲	۲۴
۱۷	کشور	۰	۳	۳۵	۲
۱۸	بر	۶	۹	۳۵	۱۰
۱۹	گفتن	۴۷	۲۴	۱	۳۳
۲۰	سال	۶	۱۷	۸	۲۳
۲۱	تا	۳۳	۲۴	۶	۱۱
۲۲	گرفتن	۱۹	۱۴	۸	۱۲
۲۳	توانستن	۵	۱۱	۲۵	۱۱
۲۴	نیز	۰	۴	۳	۷
۲۵	دیگر	۲۰	۲۱	۱۵	۲۴
۲۶	هم	۶۵	۳۹	۲	۲۵
۲۷	مورد	۱	۶	۱۵	۴
۲۸	یا	۱۶	۲۳	۴۰	۱۳
۲۹	اسلامی	۰	۰	۷۵	۰
۳۰	بایستن	۹	۱۳	۲۹	۱۱
۳۱	کار	۲۵	۳۱	۱۷	۶
۳۲	ما	۱۷	۳۸	۱	۱۴

ادامه‌ی جدول ۱. بسامد واژگان پایه در پیکره‌های ادبیات داستانی، ادبیات غیرداستانی، حقوقی و متنوع

رتبه‌ی بسامدی	واژه	ادبیات داستانی	ادبیات غیر داستانی	حقوقی	متنوع
۳۳	وی	۰	۰	۰	۰
۳۴	هر	۲۱	۲۶	۴۳	۷
۳۵	او	۵۰	۲۵	۱۱	۱۶
۳۶	قرار	۰	۴	۶	۵
۳۷	مردم	۳	۱۱	۱۳	۳
۳۸	روز	۱۱	۱۰	۲	۶
۳۹	نظر	۲	۲	۱۱	۱
۴۰	سازمان	۰	۰	۲	۰
۴۱	آمدن	۲۴	۲۷	۳	۲۱
۴۲	اما	۳۱	۴	۱	۱۵
۴۳	عنوان	۰	۲	۳	۱
۴۴	من	۷۸	۸۴	۲	۵۰
۴۵	وجود	۴	۲	۰	۱۳
۴۶	رسیدن	۳	۵	۷	۶
۴۷	همه	۲۳	۲۲	۲۷	۱۲
۴۸	صورت	۹	۸	۳۰	۱۰
۴۹	دست	۲۵	۱۸	۲	۵
۵۰	شرکت	۰	۰	۹	۲
۵۱	یافتن	۰	۴	۱	۱
۵۲	شهر	۲	۹	۲	۵
۵۳	رفتن	۲۸	۱۵	۰	۲۰
۵۴	امر	۰	۲	۲۰	۰
۵۵	منطقه	۰	۳	۳	۲
۵۶	اگر	۹	۱۰	۱	۱۹
۵۷	گزارش	۰	۰	۶	۱
۵۸	پس	۵	۴	۱۱	۹
۵۹	بسیار	۱	۱۶	۱	۹
۶۰	نمودن	۴	۱	۱۷	۰
۶۱	دولت	۱	۰	۳۵	۱
۶۲	بیشتر	۴	۵	۰	۵
۶۳	بخش	۰	۴	۶	۴



## ۶. نتیجه‌گیری

با توجه به هزینه‌های مالی و زمانی که توسط آموزش‌یاران و فراگیران صرف آموزش و یادگیری زبان فارسی می‌شود، محتوای آموزشی باید به گونه‌ای طراحی شود که بیشترین بازگشت سرمایه را برای فراگیران و آموزش‌یاران در پی داشته باشد. یکی از راهکارهای موجود برای لحاظ کردن محاسبه‌ی هزینه-فایده در حوزه‌ی یادگیری واژگان، شناسایی لغات پرکاربرد فارسی براساس معیارهای بسامد، پوشش متنی و گستره‌ی واژگان است که می‌تواند در قالب فهرست واژگان پایه‌ی فارسی تبلور یابد. این تحقیق درصدد بود که با ارائه‌ی تعریفی دقیق و علمی از مفهوم «واژه»، پوشش متنی و گستره‌ی واژگان فرهنگ بسامدی بی‌جن‌خان و محسنی (۱۳۹۱) را در یک پیکره ۳۴۰۰۰ کلمه‌ای مورد بررسی قرار دهد و از این طریق، گامی در جهت تدوین فهرست واژگان پایه‌ی فارسی بردارد.

در این تحقیق، رویکردی تلفیقی برای تعریف مفهوم «واژه» و تعیین مصادیق آن در پیکره اتخاذ شد که بینابین رویکرد «سرواژگی» و «خانوار واژگانی» است. با اتخاذ این رویکرد به بسامدشماری واژگان فرهنگ بی‌جن‌خان و محسنی (۱۳۹۱) در پیکره‌ی خود پرداختیم تا دریابیم چند درصد از کلمات این پیکره، تحت پوشش کلمات این فرهنگ قرار دارند. برای بررسی گستره‌ی واژگان فرهنگ بی‌جن‌خان و محسنی (۱۳۹۱)، ۶۳ کلمه‌ی بالای فرهنگ را در هر یک از چهار ژانر زیرمجموعه‌ی پیکره، جداگانه بسامدشماری کردیم. سپس، بسامد هر کلمه در یک ژانر را با بسامد همان کلمه در ژانرهای دیگر مقایسه کردیم. گستره‌ی مطلوب برای هر واژه، حضور آن واژه در سه ژانر از چهار ژانر پیکره در نظر گرفته شد.

فهرست اصلاح‌شده‌ی بی‌جن‌خان و محسنی (۱۳۹۱)، ۸۳,۹۳ درصد کلمات پیکره‌ی کلی را پوشش می‌دهد. هرش و نیشن (۱۹۹۲) در مطالعه‌ی خود پوشش متنی ۲۰۰۰ واژه‌ی پایه‌ی انگلیسی را در سه رمان کوتاه مورد آزمون قرار دادند. این رمان‌ها بین ۲۷ تا ۳۵ هزار کلمه داشتند که از این نظر با تحقیق حاضر قابل مقایسه است. فهرست آنها نود درصد کلمات هر سه رمان را پوشش می‌داد. اگرچه این میزان بسیار بیشتر از پوشش متنی به دست آمده در این تحقیق است، باید متذکر شد که رمان‌های مورد استفاده‌ی هرش و نیشن برای رده‌ی سنی نوجوان نوشته شده بودند و به گفته‌ی خودشان بار لغوی کمتری نسبت به متون دیگر داشتند؛ این در حالی است که متون پیکره‌ی تحقیق حاضر از حداقل چهار ژانر متفاوت با مخاطب بالغ تشکیل شده است و بار لغوی آنها نیز بیشتر از این رمان‌ها بوده است. بنابراین، به نظر نویسندگان، این میزان پوشش متنی برای فهرست ما مطلوب می‌باشد؛ چون پوشش ۲۰۰۰ واژه‌ی پایه‌ی زبان انگلیسی در متونی که برای مخاطبان بالغ نوشته شده‌اند نیز بین ۸۰ تا ۸۵ درصد در نوسان بوده است (نیشن و هوانگ، ۱۹۹۵؛ نیشن و ورینگ، ۱۹۹۷؛ نیشن، ۲۰۰۱a).

البته، این بدین معنی نیست که فهرست مبتنی بر فرهنگ بی‌جن‌خان و محسنی (۱۳۹۱) بدون اشکال است. فرهنگ بی‌جن‌خان و محسنی با رویکرد «سرواژگی» مصادیق یک کلمه را تعیین و بسامدشماری کرده است و این بر صحت و دقت نتایج ما نیز تأثیر گذاشته است. اگر رویکرد خانوار واژگانی پیشنهادی ما در تهیه‌ی فرهنگ بی‌جن‌خان و محسنی مورد استفاده قرار گرفته بود، شاید این فرهنگ شکل دیگری می‌گرفت. در برخی موارد، ناگزیر بودیم از رویکرد میانه‌ای که برای تعیین مصادیق یک کلمه پیشنهاد داده بودیم تخطی کنیم. در فرهنگ بی‌جن‌خان و محسنی، برخی کلمات مشتق نیز مدخل داشتند ولی بن‌واژه‌ی مادر آن کلمات فاقد مدخل بود. در چنین مواردی مجبور بودیم طبق روال فرهنگ پیش برویم. مثلاً «آموختن» و بن مضارع آن، یعنی «آموز»، فاقد مدخل بود، ولی مشتق آن، یعنی «آموزش»، دارای مدخل بود. اگر «آموختن» صاحب مدخل در فرهنگ می‌بود، می‌توانستیم زیر بن مضارع آن، کلمات «آموزش»، «آموزگار»، «آموزه»، «آموزنده» و حتی «آموزشگاه» و «آموزشکده» را نیز بیاوریم که بی‌شک پوشش متنی را تحت تأثیر قرار می‌داد.

از ۶۳ کلمه‌ای که در بالای فهرست بسامدی بی‌جن‌خان و محسنی بودند ۵۷ کلمه در سه یا چهار پیکره از پیکره‌های ادبیات داستانی، ادبیات غیرداستانی، حقوقی، و متنوع حضور داشتند و بنابراین، طبق تعریف عملیاتی از گستره‌ی خوبی برخوردار بودند. کلمه‌ی «وی» گستره‌ی صفر داشت؛ این امر را می‌توان این‌گونه توجیه کرد که این واژه، ادبی است و بیشتر در متون نظم و نثر کهن به چشم می‌خورد و در فارسی امروز بسامد بالایی ندارد. پیکره‌های بالا همگی از متون معاصر تشکیل شده بودند.

کلمات «سازمان» و «اسلامی» گستره‌ی یک داشتند و تنها در پیکره‌ی حقوقی که بخشی از قانون اساسی جمهوری اسلامی ایران بود، وجود داشتند. «شرکت» و «گزارش» دارای گستره‌ی دو بودند و در متون حقوقی و متنوع حضور داشتند. همچنین، کلمه‌ی «امر» در پیکره‌های ادبیات غیرداستانی و حقوقی وجود داشت. همان‌طور که جدول (۲) نشان می‌دهد، وجه مشترک پنج کلمه‌ی بالا در این است که همگی در پیکره‌ی حقوقی نسبتاً حضور پررنگی دارند و تا حدودی نیز در پیکره‌ی متنوع که مشتمل بر متون درسی و رسمی است، وجود دارند. شاید بتوان چنین استنباط کرد که این کلمات واژه‌هایی رسمی هستند که در رسانه‌های نوشتاری و رسمی رواج بیشتری دارند و در تعاملات غیررسمی و روزمره که منبع الهام ادبیات داستانی هستند، چندان باز نمود ندارند.

جدول ۲. بسامد کلمات فاقد گستره‌ی مطلوب در پیکره‌های ادبیات داستانی، ادبیات غیرداستانی، حقوقی و متنوع

وی	متنوع	حقوقی	ادبیات غیرداستانی	ادبیات داستانی
۰	۰	۰	۰	۰
سازمان	۰	۲	۰	۰
اسلامی	۰	۷۵	۰	۰
شرکت	۲	۹	۰	۰
گزارش	۱	۶	۰	۰
امر	۰	۲۰	۲	۰

نیشن و هوانگ (۱۹۹۵) گستره‌ی مطلوب برای یک کلمه‌ی پر بسامد را حضور آن کلمه در ۱۰ ژانر از ۱۵ ژانر پیکره‌ی مبدأ یک فرهنگ بسامدی تعریف کرده‌اند. در تحقیق حاضر، گستره‌ی کلمات تنها در چهار ژانر و در متونی با حجم کم مورد بررسی قرار گرفت که تعمیم نتایج را بسیار مشکل می‌کند. اما همین ژانرهای اندک و متون محدود به روشنی نشان می‌دهند که بیشتر کلمات مورد بررسی از گستره‌ی مطلوبی در ژانر-پیکره‌های این پژوهش برخوردار بودند. مهم‌تر آن که بررسی‌ها نشان دادند که آزمودن گستره‌ی یک واژه می‌تواند اطلاعات مفیدی درباره‌ی موارد کاربرد آن کلمه در اختیار ما قرار بدهد که در صورت تکیه‌ی صرف بر بسامد، این نکته مورد توجه چندانی قرار نمی‌گیرد.

در مجموع چنین به نظر می‌رسد که اصلاح فرهنگ بی‌جن‌خان و محسنی (۱۳۹۱) براساس برداشت بینابینی از مفهوم خانوار واژگانی می‌تواند در جهت تدوین فهرست لغات پایه‌ی زبان فارسی راهگشا باشد؛ زیرا با اِعمال این رویکرد، این فرهنگ حدوداً ۸۴ درصد از متن یک پیکره را پوشش می‌دهد و واژگانش از گستره‌ی مطلوبی در ژانرهای این پیکره برخوردار می‌باشند. البته از آن‌جا که پیکره‌ی مورد استفاده در این تحقیق کوچک بود و به دلیل وجود محدودیت‌های پژوهشی، نویسندگان قادر نبودند تا گستره‌ی همه‌ی کلمات فرهنگ بی‌جن‌خان و محسنی (۱۳۹۱) را مورد بررسی قرار دهند، این امر، تعمیم نتایج تحقیق را مشکل می‌سازد. با وجود این، اگر تحقیقات بعدی که پوشش متنی و گستره‌ی واژگان فرهنگ بی‌جن‌خان و محسنی (۱۳۹۱) را در پیکره‌های بزرگ‌تر مورد بررسی قرار می‌دهند، به نتایجی مشابه نتایج ما دست یابند، می‌توان برآیند و وجه مشترک این پژوهش‌ها را به‌عنوان «فهرست واژگان پایه‌ی فارسی» مدون و منتشر کرد تا مورد استفاده‌ی سایر پژوهشگران و فعالان عرصه‌ی آموزش زبان فارسی قرار گیرد.

یکی از مهمترین کاربردهای این فهرست، استفاده در امر هدف‌گذاری برای آموزش زبان فارسی به‌ویژه آموزش واژگان فارسی است. برنامه‌ریزان آموزشی می‌توانند از پرتکرارترین واژگان این فهرست شروع کنند و تا کلمه‌ی معینی، مثلاً هفتصدمین کلمه را به‌عنوان یک سطح تعریف نمایند و به همین ترتیب تا دو هزارمین کلمه، سطوح مختلف را تعیین نمایند. البته در مورد زبان‌آموزان پیشرفته می‌توان از کلماتی که در

فهرست واژگان پایه قرار ندارند، ولی براساس بسامد مدون شده‌اند، استفاده کرد. با تعریف این سطوح، کار ارزیابی محتوای موجود، طراحی و تدوین محتوای جدید و تهیه‌ی آزمون‌های دانش واژگانی و آزمون‌های توانش فارسی قاعده‌مندتر می‌شود و معیاری عینی و استاندارد در این زمینه به دست می‌آید که بسیار علمی و مدون و به دور از هر گونه سوگیری و برداشت‌های شخصی خواهد بود.

یکی از دغدغه‌های فعلی در این عرصه، نبود معیاری عینی برای بررسی و تعیین سطح محتوای موجود و بررسی تناسب آنها با سطح فراگیران است. در زمینه‌ی تدوین محتوای جدید نیز همین مشکل مطرح است. فهرست واژگان پایه می‌تواند در تهیه‌ی متون داستانی و غیرداستانی سطح‌بندی شده برای خواندن و درک مطلب درون‌کلاسی و برون‌کلاسی و همچنین ساده‌سازی متون پیچیده برای سطوح پایین‌تر نیز بسیار مفید واقع شود. تعریف سطوح مختلف براساس واژگان هدف، طراحی آزمون‌های استاندارد دانش واژگانی را تسهیل می‌کند و گامی مؤثر در تهیه‌ی آزمون‌های استاندارد فارسی است.

فراگیرانی که قصد دارند تحصیلات دانشگاهی خود را در دانشگاه‌های فارسی‌زبان دنبال کنند باید واژگان علمی و تخصصی رایج زبان فارسی را یاد بگیرند. استفاده از فهرست واژگان پایه‌ی فارسی، روشی مؤثر در شناسایی این واژگان می‌باشد. بدین منظور باید واژگان پربسامد خارج از فهرست واژگان پایه را در متون رشته‌های مختلف علمی و دانشگاهی شناسایی کنیم. کلمات پربسامد و غیرپایه‌ای را که در میان متون رشته‌های مختلف علمی مشترک هستند، به‌عنوان واژه‌های دانشگاهی<sup>۱</sup> در نظر می‌گیریم و کلمات پربسامد مختص یک رشته را واژگان تخصصی<sup>۲</sup> آن حوزه تعریف می‌کنیم.

## منابع:

- اسلامی، م. و همکاران. (۱۳۸۳). *واژگان زبانی زبان فارسی*. مجموعه مقالات اولین کارگاه پژوهشی زبان فارسی و رایانه. تهران: دانشگاه تهران.
- باقری، ف. (۱۳۸۹). *بررسی میزان انطباق واژه‌های پایه در کتاب‌های آموزش زبان فارسی با نیازهای فارسی‌آموزان خارجی*. گلبانگ آشنایی: مجموعه مقالات هفتمین مجمع بین‌المللی استادان زبان و ادبیات فارسی (صص: ۱۸۴-۱۵۵). تهران: شورای گسترش زبان و ادبیات فارسی.
- بی‌جن‌خان، م. (۱۳۸۶). *گزارش طرح مطالعه و تحقیق جهت تدوین پژوهش‌نامه‌ی عملیاتی دادگان: پیاده‌سازی استاندارد ایگلز در پیکره‌ی متنی زبان فارسی معاصر*. تهران: معاونت پژوهشی و فناوری دانشگاه تهران. قرارداد شماره ۳۵۵۴/۱۹۰ مورخ ۱۳۸۵/۱۲/۱۸.
- بی‌جن‌خان، م. و محسنی، م. (۱۳۹۱). *فرهنگ بسامدی براساس پیکره‌ی متنی زبان فارسی امروز*. تهران: موسسه‌ی انتشارات دانشگاه تهران.

1. academic  
2. technical

حسینی، ح. (۱۳۸۴). واژه‌های پرکاربرد فارسی امروز (بر مبنای پیکره یک میلیون لغتی). تهران: کانون زبان ایران.  
رجائی، ح. و علیزاده، ع. (۱۳۸۹). تعیین واژگان پربسامد در رسانه‌های خبری فارسی‌زبان برای آموزش زبان فارسی. گلبانگ‌آشنایی: مجموعه مقالات هفتمین مجمع بین‌المللی استادان زبان و ادبیات فارسی (صص: ۴۹۷-۵۱۶). تهران: شورای گسترش زبان و ادبیات فارسی.  
عرفانیان قنسولی، ل. و شریفی، ش. (۱۳۹۰). بررسی وندهای اشتقاقی زایا و غیرزایا در فارسی معاصر. مجموعه مقالات سومین هم‌اندیشی صرف (صص: ۱۳۸-۱۲۳). تهران: انجمن زبان‌شناسی ایران.

- Bauer, L. & Nation, P.** (1993). Word Families. *International Journal of Lexicography*, 6(4): 253-279.
- Beglar, D. & Hunt, A.** (2005). Six Principles for Teaching Foreign Language Vocabulary: A Commentary on Laufer, Meara, and Nation's 'Ten Best Ideas'. *The Language Teacher*, 29(7): 7-10.
- Brezina, V. & Gablasova, D.** (2013). Is There a Core General Vocabulary? Introducing the New General Service List. *Applied Linguistics*, 1-23. Retrieved December 26, 2014, from <http://applied.oxfordjournals.org>.
- Carter, R.** (1998). *Vocabulary: Applied Linguistic Perspectives*. London: Routledge.
- Clahsen, H. et al.** (2003). Derivational morphology in the German mental lexicon: A dual mechanism account. In R. H. Baayen & R. Schreuder, (Eds.), *Morphological Structure in Language Processing* (pp. 125-155). Berlin: Mouton de Gruyter.
- Hirsh, D. & Nation, P.** (1992). What Vocabulary Size is Needed to Read Unsimplified Texts for Pleasure? *Reading in a Foreign Language*, 8(2): 689-696.
- McArthur, T.** (2015). Compound-Complex Word. *Concise Oxford Companion to the English Language*. Retrieved January 29, 2015, from <http://www.encyclopedia.com>.
- Nagy, W. et al.** (1989). Morphological Families in the Internal Lexicon. *Reading Research Quarterly*, 24: 262-281.
- Nation, P.** (2001a). How many high frequency words are there in English? In M. Gill, A. W. Johnson, L. M. Koski, R. D. Sell & B. Warvik, (Eds.), *Language, Learning, Literature: Studies Presented to Hakan Ringbom* (pp. 167-181). Abo Akademi University, Turku: English Department Publications.
- Nation, P.** (2001b). *Learning Vocabulary in Another Language*. Cambridge: Cambridge University Press.
- Nation, P. & Hwang, K.** (1995). Where Would General Service Vocabulary Stop and Special Purposes Vocabulary Begin? *System*, 23(1): 35-41.
- Nation, P. & Meara, P.** (2002). Vocabulary. In N. Schmit, (Ed.) *An Introduction to Applied Linguistics* (pp. 35-54). London: Arnold.
- Nation, P. & Waring, R.** (1997). Vocabulary size, text overage and word lists. In N. Schmitt & M. McCarthy, (Eds.), *Vocabulary: Description, Acquisition and Pedagogy* (pp. 6-19). Cambridge: Cambridge University Press.
- Yule, G.** (1996). *The Study of Language* (2<sup>nd</sup> edition). Cambridge: Cambridge University Press.