



فصلنامه علمی زبان پژوهی دانشگاه الزهراء (س)

سال پانزدهم، شماره ۴۶، بهار ۱۴۰۲

نوع مقاله: پژوهشی

صفحات ۱۴۷-۱۱۱

زیایایی صرفی پسوندهای اشتقاقی غیر فعلی فارسی: شواهدی از «پیکره وابستگی»^۱

آناهید دشتی^۲، ویدا شقاقی^۳

تاریخ دریافت: ۱۴۰۰/۱۰/۱۷

تاریخ پذیرش: ۱۴۰۱/۰۸/۰۳

چکیده

زیایایی صرفی، موضوعی بنیادی در واژه‌سازی اشتقاقی است که در سنجش و تحلیل، ابعاد پیچیده خود را به نمایش می‌گذارد. توجه به تفاوت بین سنجش زیایایی گذشته، میزان استفاده کاربران و زیایایی کنونی، موضوع در خور توجهی است. در بیشتر سنجش‌های زیایایی صرفی، شمارش ساده واژه‌های موجود (بالفعل) در فرهنگ‌های لغت و پیکره‌های زبانی، انجام می‌شود. در صورتی که سنجش زیایایی با تعیین بسامد واژه‌های دیگری همچون واژه‌های ممکن (بالقوه)، نوواژه‌ها و تک‌بسامدی‌ها نیز سر و کار دارد. زیایایی، هر چند با شمارش این بسامدها، مرتبط است اما شمارش صرف با مفهوم زیایایی صرفی، برابر نیست. در این مقاله، گروه‌بندی نوینی از همه سنجش‌های کمی زیایایی صرفی، بر مبنای مفاهیم آمار و احتمالات ارائه گردید که در چهار مدل کلی؛ تعیین نسبت، تعیین بسامد، تعیین نوواژه‌ها و تعیین تک‌بسامدی‌ها (برآورد احتمال)، گنجانده شد. سپس زیایایی صرفی ۵۰ پسونداشتقاقی غیر فعلی زبان فارسی بر اساس «پیکره وابستگی» ارزیابی شده و

^۱ شناسه دیجیتال (DOI): 10.22051/jlr.2022.38942.2129

شناسه دیجیتال (DOR): 20.1001.1.20088833.1402.15.1.5.8

^۲ بر خود لازم می‌دانیم از استاد محترم دکتر شهرام مدرس خیابانی که در بخش داده‌پردازی، زحمت به‌کارگیری واژه‌پرداز ورد اسمیت «Word Smith» را (به صورت نمونه) متقبل شدند، به طور ویژه سپاسگزاری و قدردانی نمایم. همچنین قدردان همکاری‌های ارزشمند خانم انیس معصومی، دانشجوی دکتری زبان‌شناسی هستیم و از ایشان سپاسگزاری می‌کنیم.

^۳ دانشجوی دکترا، گروه زبان‌شناسی همگانی، دانشگاه علامه طباطبائی، تهران، ایران؛

anahid_dashti@atu.ac.ir

^۳ استاد گروه زبان‌شناسی، دانشگاه علامه طباطبائی، تهران، ایران (نویسنده مسئول)؛ shaghaghi@atu.ac.ir

رتبه‌بندی پسوندها در پیوستار احتمالی زایایی، تحلیل شد. این پیکره دارای ۴۶۹۶۴۳ واژه است. یافته‌ها نشان می‌دهد، در این پیکره مطابق با چارچوب سنجش زایایی جهانی (Bauer, 2004) و (Baayen, 2009)، سه پسوند «-ی» و «-ن و -ا»، با مقادیر (۰.۱)، (۰.۱) و (۰.۴۷)، بالاترین احتمال زایایی را دارند. ۲۵ پسوند دارای پایین‌ترین احتمال (۰.۰۰۱) زایایی‌اند. ۲۲ پسوند دیگر با احتمالی بین (۰.۰۱) و (۰.۰۰۷)، در میانه پیوستار قرار گرفتند. استفاده از زایایی جهانی به دلیل محاسبه سهم عددی مجموع تک‌بسامدی‌های پیکره در تغییر احتمال زایایی، دارای پایایی‌اند. بدیهی است یافته‌های این پژوهش فقط در «پیکره وابستگی» دارای روایی است. برای تحلیل دقیق‌تر پیوستار احتمالی زایایی پسوندها، به بررسی‌های مبتنی بر پیکره‌های بزرگ‌تر نیازمندیم.

واژه‌های کلیدی: زایایی صرفی، نوواژه، واژه تک‌بسامدی، پیکره وابستگی، زایایی جهانی، زبان فارسی.

۱. مقدمه

کاربران زبان فارسی همچون هر زبان دیگری که دارای ویژگی‌های رده پیوندی است؛ با استفاده از وندافزایی و داشتن شمار زیادی ریشه، پیشوند و پسوند و همچنین با کمک فرایندهای صرفی آن، می‌توانند در رویارویی با هر پدیده جدیدی، واژه‌های نو بسازند. اشتقاق، یکی از این فرایندهاست که طی آن، تکواژهای وابسته‌ای به نام وند به واژه‌های موجود^۱ در زبان پیوسته می‌شوند. زایایی، طیف گسترده‌ای از پدیده‌های زبانی را پوشش می‌دهد. بوی (Booii, 2005) بیان می‌کند؛ «زایایی به همان اندازه که با پدیده‌های نحوی ارتباط دارد؛ می‌تواند با پدیده‌های صرفی مرتبط باشد.» (Booii, 2005, p. 303). آرونوف (Aronoff, 1980) نیز ادعا می‌کند «زایایی غالباً موضوع پنهانی در پس‌زمینه است.» (Aronoff, 1980, p. 71). باور (Bauer, 2004) در مورد یکی از مشکلات تجربی مسئله زایایی در نظریه‌های صرفی بیان نموده است که «هنوز معیار قابل‌اعتمادی برای زایایی نداریم.» (Bauer, 2004, p. 214). زایایی تعریف‌های گوناگونی دارد. تعریف زایایی در نمایش رویکرد^۲ و ارائه روش^۳ اندازه‌گیری زایایی، بازتاب اساسی دارد. شواهد بسیاری مبنی بر تفاوت در میزان زایایی وندها وجود دارد. برای نمونه، در زبان فارسی، پسوند «-ی و -ا»^۴ راحت‌تر به واژه‌ها افزوده می‌شود تا پسوند «-واره، -گر و -بان». به بیانی، کاربران زبان فارسی، واژه‌های مرکب مشتق دروغ‌پراکنی و دروغ‌گویی را بیشتر از

¹ realized words

² approach

³ method

⁴ -i, -e

واژه‌های مشتق دروغ‌بان، دروغچی یا دروغ‌گر می‌پذیرند. برخلاف نظام تصریفی زبان، متغیرهای نظام اشتقاق مجموعه بازی را تشکیل می‌دهند. به گونه‌ای که در زبان فارسی، برای ساختن اسم از اسم، پسوندهای گوناگونی مانند «-زار»، «-گاه»، «-بان» و موارد مشابه وجود دارند (Abbassi, 2006).

اصلی‌ترین هدف این پژوهش، سنجش کمی زایایی پسوندهای اشتقاقی غیر فعلی در زبان فارسی است. به گونه‌ای که در آن، پیوستار احتمالی زایایی این پسوندها در «پیکره وابستگی» به دست داده می‌شود. در این مقاله، همچنین به معرفی و تحلیل فرمول‌های مختلف برای سنجش کمی زایایی در بررسی‌های پژوهشگران طی سی سال گذشته، می‌پردازیم. موضوع در خور توجه این تحلیل‌ها، دقت به بعدهای زایایی صرفی است که پیچیدگی خود را هنگام سنجش زایایی با هر رویکرد و روش، نشان می‌دهد. در جریان تحلیل، تفاوت‌های عمده بین زایایی گذشته، میزان استفاده کاربران و زایایی کنونی را نیز مشخص می‌کنیم.

پیش از پرداختن به بعدهای گوناگون سنجش زایایی^۱، به مفهوم زایایی صرفی^۲ می‌پردازیم. زایایی صرفی از جنبه‌های بسیاری تعریف شده است. شولتینک (Schultink, 1961)، می‌نویسد «زایایی صرفی، امکان استفاده کاربران از قاعده‌های زبان به طور ناخواسته برای ابداع تعداد نامحدودی صورت جدید است که در پس تناظرهای صورت-معنی واژه‌های شناخته شده قرار دارد» (Schultink, 1961, p.113). در این تعریف، به سه بُعد مهم غیر عمدی بودن^۳، قاعده‌مند بودن^۴ و نامحدود بودن^۵ اشاره شده که به هم وابسته‌اند. آثار نظری و توصیفی در مورد زایایی صرفی بیشتر بر آنچه که بااین (Baayen, 1992) آن را بُعد کیفی زایایی می‌نامد، متمرکزند و در پیوند با محدودیت‌های دستوری هر فرایند صرفی، مورد بررسی قرار می‌گیرند (Baayen, 1992). بسیاری از تعریف‌های دیگر، زایایی را مفهومی می‌انگارند که مستلزم فرض وجود پیوستار است. برای نمونه، به باور انشن و آرونوف (Anshen & Aronoff, 1998)، «زایایی به معنای استفاده از نند معینی در تولید واژه‌های جدید است (Anshen & Aronoff, 1998, p. 242-243). با این نگاه، زایایی پیوستاری احتمالی است که امکان استفاده واژه‌های ممکن^۶ را پیش‌بینی می‌کند، که در دو انتهای پیوستار، الگوهای کاملاً غیر زایا و الگوهای بسیار زایا وجود دارد. شماری از موارد بینابینی نیز در محدوده بین آن دو یافت می‌شود. پلنگ

¹ productivity measurement
² morphological productivity
³ unintentionality
⁴ regularity
⁵ unlimitedness
⁶ potential word

(Plag, 2006) نیز معتقدست «زایایی یک فرایند صرفی می‌تواند به عنوان «امکان» کلی آن برای ساخت واژه‌های جدید و به عنوان میزان استفاده کاربران از این امکان، تعریف شود.» (Plag, 2006, p. 127). به آن معنا که کاربران زبان فارسی به هنگام نیاز می‌توانند از «امکان» زبان خود استفاده نمایند و واژه‌های «مرگ آسا، افزونه، پول‌شویی، شبح‌وار و رنگ‌واره» را با استفاده از پسوندهای موجود زبان بسازند و مفهوم موردنظرشان را با ساختن آن واژه‌ها به دیگران منتقل نمایند. بنابراین «امکان» یکی از ویژگی‌های اصلی تعیین‌کننده زایایی صرفی است. زایایی می‌تواند با به‌کارگیری بیش از یک الگوی واژه‌سازی، رخدادی همزمان و قاعده‌مند برای تولید واژه جدید باشد (Safari & Niknasab, 2020, p. 17).

در بخش دوم این مقاله، پیشینه زایایی صرفی و در بخش سوم، چارچوب و شیوه‌های کمی^۱ سنجش زایایی صرفی، بیان می‌شود. در بخش چهارم و پنجم، به ترتیب به بحث پیرامون سنجش زایایی صرفی در «پیکره وابستگی» و نتیجه‌گیری می‌پردازیم.

۲. پیشینه پژوهش

پدیده زایایی صرفی مدت‌هاست که در متن‌های زبان‌شناسی به صورت نظری مورد بحث بوده و «به عنوان موضوعی غیر قابل حل، در صرف اشتقاقی دنبال شده‌است» (Aronoff, 1976, p. 36). ولی، ایده زایایی صرفی قدیمی‌تر از پژوهش‌های آرونوف است. نخستین اظهارنظر در مورد جنبه‌ای از توصیف زایایی به وسیله ویلمنس (Willmanns, 1899)، بیان گردیده که انواع واژه‌های مشتق را با استفاده از پایداری و تداوم ساخت توصیف کرد^۲ (Willmanns, 1899). ایده مشابه دیگر را کروسینگا (Kruisinga, 1932)، با طرح اصطلاحات پسوندهای زنده و پسوندهای مرده ارائه کرده‌است (Kruisinga, 1932, p. 22). ایده‌ها و تعریف‌های ارائه‌شده، ابعادی دارند که می‌تواند به عنوان ابعاد کیفی زایایی صرفی مطرح شود. پلگ (Plag, 1999)، هی (Hay, 2003) و باور (Bauer, 2001)، پژوهش‌های کیفی^۳ گوناگونی درباره زبان انگلیسی انجام داده‌اند. در حالی که بااین (Baayen, 1992)، در محاسبه زایایی، به احتمال یافتن کلمات جدیدی در هر متن توجه کرده است که توسط پسوندهای اشتقاقی ساخته می‌شوند. بنابراین، بااین (همان) روش‌های اندازه‌گیری زایایی صرفی به شیوه کمی را گسترش می‌دهد.

^۱ (Quantitative manners) رویکردی قیاسی که با تبدیل اطلاعات پژوهش به نمادهای ریاضی و به‌کارگیری قاعده‌ها، به آزمودن نظریه‌های زبان می‌پردازد.

^۲ by vitality and persistency

^۳ رویکردی استقرایی که با استفاده از فنون مشاهده، مصاحبه، بررسی سند و تحلیل محتوا، به تفسیر فرایند تولید و درک واژه‌ها در راستای تعمیم نظریه‌های زبان می‌پردازد.

روش‌های آماری که به وسیله باین و همکارانش ارائه شده‌است، بهره‌گیری از پیکره‌ها برای سنجش احتمال رویارویی با واژه جدید ساخته‌شده با هر مقوله صرفی را، فراهم می‌سازد (Plag, 2003, p. 65). هرچند فرمول‌های باین (Baayen, 1992) مورد انتقاد قرار گرفته‌است، ولی ون مرلی (Van Marle, 1992)، می‌نویسد «کسی که به طور جدی درگیر مطالعه زبانی صرفی باشد نمی‌تواند مطالعات باین را نادیده بگیرد» (Van Marle, 1992, p. 152). نکته مهم آنکه، پلگ (Plag, 1999, p. 117) و بولوسکی (Bolozky, 1999, p. 194) بر این باورند که استفاده از مقایسه یافته‌های چندشاخه^۱، می‌تواند ارزیابی قابل اعتمادی از زبانی صرفی باشد. اگرچه بولوسکی (Bolozky, 1999) نشان می‌دهد یافته‌های به‌دست آمده از مقایسه‌های گوناگون، یک‌دیگر را تقویت می‌کنند ولی او همچنین می‌نویسد «به نظر نمی‌رسد سنجش دقیق زبانی صرفی، هدفی واقع‌بینانه باشد» (Bolozky, 1999, p. 3). هدف ما نیز در این پژوهش، رسیدن به یک فرمول قطعی برای سنجش زبانی صرفی در زبان فارسی نیست. ولی هر پژوهشی در این زمینه، می‌تواند در سنجش کمی پدیده چندلایه و پیچیده‌ای مانند زبانی، افق تازه‌ای بگشاید. در این جا لازم است به نمونه‌ای از کارهای انجام‌شده در این زمینه در زبان فارسی اشاره شود. برای نمونه، می‌توان به رساله دکتری عباسی (Abbassi, 2006) با عنوان «زبانی در فرایند اشتقاق زبان فارسی»، اشاره نمود. پیکره مورد بررسی در این پژوهش، مشتمل بر متن‌های ادبی معاصر ۳۵ کتاب داستانی تألیف و ترجمه بوده که از پیکره زبانی گردآوری شده در پژوهشگاه علوم انسانی و مطالعات فرهنگی استخراج گردیده‌است. برای به دست آوردن میزان تأثیر ۲۷ فرایند اشتقاقی زبان فارسی (شامل ۲۰ پسوند و ۷ پیشوند)، در رشد همه واژه‌ها، پیکره مورد مطالعه (با ۱۱۲۶۴۲۶ واژه)، از فرمول باور (Bauer, 2001; quoted in Abbassi, 2006)، بهره گرفته شده‌است. بر مبنای این فرمول، میزان نواژه‌های ساخته‌شده در یک فرایند اشتقاقی معین در مقایسه با نواژه‌های کل پیکره، از تقسیم تعداد واژه‌های تک‌بسامدی ساخته‌شده در آن فرایند بر تعداد کل واژه‌های تک‌بسامدی در پیکره مورد بررسی، محاسبه گردیده‌است. در این پژوهش، وندهای «-ی، بی-، -نده، -انه و نا-» دارای بالاترین رتبه زبانی و وندهای «آبر-، -ینه، -مان، -ار و پاد-» دارای پایین‌ترین رتبه زبانی هستند.

همچنین عباسی (Abbassi, 2006) در مقاله‌ای با عنوان «محدودیت‌های صرفی و نحوی در زبانی فرایند اشتقاق در زبان فارسی»، می‌نویسد میان زبانی و تعداد محدودیت‌های حاکم بر عملکرد زبانی فرایندهای صرفی، رابطه معکوس وجود دارد. به این ترتیب که با افزایش تعداد

¹ many-pronged

محدودیت‌ها، زایایی صرفی کاهش می‌یابد. محدودیت‌های صرفی شامل طبقه صرفی پایه، ساخت صرفی پایه و وند موجود در آخرین لایه پایه هستند.

در ادامه، به‌طور گسترده بحث خواهد شد که تخمین زایایی با روش‌های مختلف و داده‌های متفاوت، بازتابِ گوناگونی از ابعاد گوناگون زایایی، را نشان می‌دهد.

۲.۱. رویکردهای مطالعه زایایی صرفی

به‌طور کلی رویکردهای مختلف در مطالعات زایایی صرفی را می‌توان در چهار رویکرد مبتنی بر فرهنگ لغت، رویکرد مبتنی بر پیکره‌ها، رویکرد طراحی آزمون و رویکرد شبیه‌سازی گنجانند (Bolzky, 1999; Plag, 1999). در این پژوهش، سنجش زایایی صرفی بر اساس رویکرد مبتنی بر پیکره‌ها انجام می‌شود. امروزه زبان‌شناسی پیکره‌ای برای اندازه‌گیری زایایی صرفی به‌طور فزاینده‌ای رایج شده‌است. همان‌گونه که باین و لیبر (Baayen & Lieber, 1991) بیان کرده‌اند، پیکره‌ها نسبت به فرهنگ لغت، دارای سه ویژگی مثبت‌اند؛ نخست آنکه پیکره‌ها معمولاً در بردارنده‌ی واژه‌های شفاف‌تری هستند که به‌طور معمول در فرهنگ‌های لغت نادیده گرفته می‌شوند. دوم آنکه پیکره‌ها فقط شامل واژه‌هایی‌اند که در گفتار و نوشتار طبیعی کاربران به کار می‌روند، ولی فرهنگ لغت واژه‌هایی را که در زبان روزمره رایج نیستند، فهرست می‌کند. ویژگی سوم نیز به این نکته برمی‌گردد که پیکره‌ها، اطلاعاتی در باره‌ی بسامد واژه‌ها ارائه می‌دهند (Baayen & Lieber, 1991, p. 803). عباسی (Abbassi, 2006)، در مقاله‌ی در باره‌ی زایایی ساخت‌وازی و مدخل‌گزینی در فرهنگ‌های عمومی یک‌زبانه انجام داده‌است. بر مبنای یک اصل صرفی در زایایی، پیش‌بینی می‌شود که با افزایش کاربرد یک فرایند صرفی، درصد نوواژه‌های ثبت‌نشده در فرهنگ لغت و در نتیجه درصد واژه‌های تک‌بسامدی ثبت‌نشده نیز بالا رود. ولی عباسی (همان) نتیجه گرفته‌است که این اصل رعایت نشده‌است. به گونه‌ای که آن دسته از صورت‌های ساخته‌شده با پسوند «-زار»، که در پیکره زبانی موجود بوده‌اند در فرهنگ سخن ثبت نشده‌اند. بیشتر موارد مشابه مربوط به پسوند «-ی» بوده‌است که زایاترین وند زبان فارسی است (Abbassi, 2006).

روش‌های گوناگونی وجود دارد که از پیکره‌ها برای سنجش زایایی استفاده می‌شود. ساده‌ترین روش‌ها با تعیین بسامد نوع‌واژه‌ها^۱ و موردواژه‌های واژه‌ها است

¹ types

منظور از (type)، تعداد واژه‌های ساخته‌شده با یک الگو یا وند در پیکره است (Shaghghi, 2016). مانند یک واژه «دانشمند» در یک متن که با فرایند پسوندافزایی ساخته شده‌است (نوع‌واژه دانشمند=۱).

(McEnery & Wilson, 2001). در ادامه، ارتباط بین روش‌های مطرح‌شده در پژوهش‌های مختلف طی سه دهه، بیان می‌شود. همچنین مشاهده خواهد شد که در هر یک از روش‌های سنجش کمی، بعد متفاوتی از زایایی برجسته می‌شود. در پژوهش حاضر، برای سنجش زایایی از چهار روش بسامدی و دو روش برآورد احتمالی بهره گرفته شده است. روش‌های بسامدی از جنبه آماری دارای پایایی نیستند ولی به سبب بهره‌گیری از آن‌ها در پژوهش‌های زبان فارسی و نکته‌های مهمی که پیرامون تحلیل زایایی صرفی در زبان فارسی وجود دارد؛ توجه به یافته‌های ناپایای آن‌ها ضروری است. زایایی کنونی پسوند‌های مورد بررسی، مطابق با روش‌های سنجش احتمالی با عنوان‌های زایایی دقیق و زایایی جهانی باور و باین (Baayen, 2009; Bauer, 2004)، مورد بررسی و تحلیل قرار می‌گیرد. این سنجش کم دربرگیرنده دیدگاه‌های باین و لیبر (Baayen & Lieber, 1991)، باین (Baayen, 1992) و باور (Bauer, 2001)، است. پایایی آماری روش‌های سنجش احتمالی مورد استفاده در هر بخش، تحلیل شده است.

۲.۲. شیوه‌های سنجش کمی زایایی صرفی

زبان‌شناسان تاکنون با استفاده از روش‌های مختلف به سنجش کمی و تحلیل زایایی صرفی پرداخته‌اند که می‌توان آن‌ها را در چهار مدل کلی تعیین نسبت، تعیین بسامد، تعیین سهم عددی نوواژه‌ها^۲ و تعیین سهم عددی تک‌بسامدی‌ها^۳ یا برآورد احتمال قرار داد. نکته مهم آن است که هر چند تلاش نظریه پردازان در راستای محاسبه زایایی صرفی بوده ولی با تحلیل دقیق‌تر، روشن می‌شود که آن‌ها جنبه‌های متفاوتی از پدیده زایایی را اندازه‌گیری نموده‌اند.

۲.۲.۱. مدل‌های تعیین نسبت

در این مدل‌ها با مقایسه واژه‌های مرتبط، نسبت^۴ آن‌ها سنجش می‌شود. در نسبت، رابطه بین دو عدد سنجیده می‌شود که مقدار عدد نخست (صورت کسر) از مقدار عدد دوم (مخرج)، کاملاً جداست. این

¹ tokens

منظور از (token)، هر وقوع یک واژه به عنوان یک موردواژه از آن واژه در پیکره به شمار می‌آید (Shaghghi, 2016). مانند ۲۰ بار رخداد واژه «دانشمند» در یک متن که با همان فرایند پسوندافزایی ساخته شده است (موردواژه دانشمند=۲۰).

² neologisms

³ the count of hapax legomena

⁴ ratio

(ratio) یا نسبت، رابطه میان دو کمیت هم‌جنس است که به صورت حاصل تقسیم یکی بر دیگری بیان می‌شود مانند نسبت تک‌بسامدی‌های دارای پسوند (مند) به تک‌بسامدی‌های دارای پسوند (ور).

مدل شامل دو روش است:

الف. نسبت واژه‌های ممکن یک قاعده به همه قاعده‌ها

واژه‌های ممکن، کلماتی‌اند که می‌توانند مطابق فرایند ویژه‌ای ساخته شوند. سنجش این مفهوم بسیار مشکل است، زیرا مستلزم تشخیص درست همه‌ی محدودیت‌های حاکم بر یک فرایند صرفی است. هر چند ایده‌ی این روش در زایایی وجود دارد ولی از جنبه‌ی عملی، فرمول روشنی برای سنجش آن ارائه نمی‌شود.

ب. نسبت واژه‌های موجود واژه‌های ممکن

باور (Bauer, 2001, p. 27; quoted in Aronoff, 1976, p. 36)، می‌نویسد؛ می‌توان با محاسبه‌ی نسبت واژه‌های موجود ساخته‌شده با یک قاعده صرفی به تعداد واژه‌های ممکن قابل تولید هر قاعده، زایایی را اندازه‌گیری کرد. نسبت واژه‌های موجود به واژه‌های ممکن، مستلزم آن است که بتوان تعداد واژه‌های ممکن را شمرد که عملی نیست. افزون بر آن، اگر پایه‌هایی باشند که به صورت زایا شکل بگیرند، واژه‌های ممکن به صورت مجموعه‌ای باز در نمی‌آید که سنجش نسبت واژه‌های موجود به واژه‌های ممکن، معنادار نخواهد بود (Haspelmath & Sims, 2010, p. 130).

۲.۲.۲. مدل‌های تعیین بسامد

این مدل‌ها شامل روش‌هایی می‌شود که بسامد واژه‌ها را با پدیده‌ی زایایی مرتبط می‌داند. منظور از بسامد، تعداد بارهای رخداد یک واژه در پیکره‌ای معین است. در این مدل، سه روش بسامد نوع واژه^۱، بسامد مورد واژه^۲ و بسامد نسبی^۳ مورد توجه قرار گرفته‌است. باین (Baayen, 1992) به رابطه‌ی بین بسامد و زایایی پرداخته‌است. پلگ (Plag, 2003) نیز در مورد ارتباط زایایی با بسامد، بیان می‌کند که فرایندهای غیر زایای صرفی با تعداد بیشتر واژه‌های نسبتاً پُربسامد و تعداد کمتر واژه‌های کم‌بسامد، مشخص می‌شوند. در مقابل فرایندهای زایا، با تعداد زیاد واژه‌های کم‌بسامد و تعداد کمتر واژه‌های پُربسامد مورد انتظارند (Plag, 2003, p. 66). پس زایایی تا اندازه‌ی بسیاری، تابع بسامد واژه‌ها است. ولسی برخی پژوهشگران افزون بر بسامد، موارد دیگری مانند سادگی واژه، پوشش معنایی واژه و ویژگی‌های سبکی واژه را نیز مؤثر می‌دانند (Sahraei et al., 2019, p. 6).

¹ open-ended

² frequency of type

³ frequency of token

⁴ relative frequency

الف. بسامد نوع واژه (روش اول)

تعداد واژه‌های ساخته شده با الگویی معین را بسامد نوع واژه می‌نامند. هرگاه وند خاصی (مانند پسوند /-یای/ نسبت یا صفت ساز و /-یای مصدری/ عمل) به ترتیب برای ساخت نوع واژه‌های جدید (نسبتی مانند یمنی، چچنی، اینترنتی، و مصدری مانند پولشویی، رونمایی و موارد مشابه)، مورد استفاده کاربران زبان قرار می‌گیرد، گفته می‌شود که آن پسوند زیاست. در این روش، با شمارش تعداد نوع واژه‌های مختلف یافت شده با یک وند خاص در یک مقطع زمانی مشخص، زیایی آن وند تعیین می‌شود. به این روش، تعیین بسامد وند نیز گفته می‌شود. ولی مشکل جدی این سنجش آن است که فراوانی نوع واژه با زیایی صرفی برابر نیست. واژه‌های بسیاری با یک پسوند معین می‌توان یافت که برای تولید واژه‌های جدید مورد استفاده کاربران قرار نمی‌گیرند. پس پسوند باید در حال حاضر غیر زایا انگاشته شود ولی بسامد نوع واژه هر وند، واقعیاتی در مورد میزان زیایی آن در گذشته، بازنمایی می‌کند. بر مبنای این معیار، پسوند فارسی /-بان/ و پسوند انگلیسی /-ment/، بسامد نوع واژه بالایی دارند ولی زایا نیستند. بر عکس، واژه‌های زیادی با پسوند /-مند/ یا /-ese/ وجود ندارد، ولی کاربران از آن‌ها برای ساختن واژه‌های جدید بهره می‌گیرند (مانند توانمند و هدفمندسازی).

شریفی و عرفانیان قونسولی با بهره‌گیری از چهار کتاب نوشته شده در سال‌های (۱۹۸۳، ۱۹۸۴، ۱۹۹۰ و ۱۹۹۶) و استخراج ۲۸۰۰ واژه محتوایی (شامل اسم، فعل، صفت و قید)، از هر متن، میزان زیایی را با عنوان‌های زیاترین و غیر زیاترین ونداها به دست داده‌اند. این پژوهشگران، سه وند /-ی/ نسبت، /-ی/ مصدری و /-ه/ مصدری را به عنوان زیاترین ونداها و وندهای «ناک، مان، در، سان، با، ین، واده، فرا، ال، فر، وا، وند» را به عنوان وندهایی با کمترین زیایی در زبان فارسی معاصر معرفی نموده‌اند. همان گونه که مشاهده می‌شود پیکره مورد بررسی بسیار کوچک و تعیین میزان زیایی، بر مبنای شمارش ساده واژه‌ها بوده است (Sharifi & Erfanian-ghonsouli, 2011). مواجی و اسلامی نیز با در نظر گرفتن هر دو رویکرد هم‌زمانی و در زمانی و با ارائه اندازه‌گیری‌های آماری بر پیکره‌ای از شعر فارسی (پیکره گنجور)، به سنجش زیایی در فرایند اشتقاق پسوندی در نظم فارسی پرداخته‌اند. این پیکره بیش از یک میلیون مصراع از پنجاه و دو شاعر فارسی زبان از قرن دوازدهم تا قرن بیست و یکم میلادی را شامل می‌شود. آن‌ها با استفاده از روش شمارش تک‌بسامدی‌ها، میزان زیایی پسوند‌های اشتقاقی را با شاخص‌های عددی نشان داده‌اند. در این پژوهش، وجود ۸۳ پسوند زبان فارسی طی ده قرن نشان داده شده است (Mavaji & Eslami, 2013). همه یافته‌های به دست آمده منطبق بر متون نظم زبان فارسی است و اگر پیکره‌های دیگری مانند روزنامه‌ها، مجله‌های منشور و کتاب‌ها بررسی شود،

ممکن است یافته‌های متفاوتی به دست آید.

سنجش بسامد نوع‌واژه‌ها حتی در پیکره‌های بزرگ نیز معیار مناسبی برای تعیین زایایی کنونی به دست نمی‌دهد. زیرا زایایی صرفی به معنای قابلیت ساختن یا پدیدار شدن نوواژه‌هایی است که پیش از زمان کنونی، وجود نداشته‌اند (Booii, 2001, p. 677). باین (Baayen, 1993) با استفاده از فرمول‌های ریاضی، این روش را با توجه به اندازه مقوله صرفی طرح می‌کند. در این سنجش، وندی که به پایه‌های بیشتری پیوسته می‌شود و واژه‌های مرکب و ندادار بیشتری می‌سازد، زایاتر است. به بیان دیگر، قاعده این وندافزایی، دارای قابلیت زایایی موجود^۱ (تحقق یافته) است و باین (همان)، به آن، میزان استفاده^۲ گفته‌است که فقط زایایی صرفی گذشته را اندازه می‌گیرد. زایایی یک مقوله صرفی معین (C) با شمارش نوع‌واژه $\{V_{C,N}\}$ ، به شمار می‌آید و (N) اندازه پیکره است. باین (Baayen, 2009) در نقد این روش می‌نویسد که برای سنجش زایایی ممکن^۳ (نهفته)، می‌توان شمارش ساده‌ای از نوع‌واژه‌های پیکره داشت. ولی بسامد نوع‌واژه، فقط تقریب نخست زایایی یک قاعده را ارائه داده و نکته‌های مهمی را از دست می‌دهد^۴ (Baayen, 2009, p. 8). باید دقیقاً توجه داشت که این روش مورد انتقاد درستی قرار دارد زیرا همان گونه که اشاره شد ممکن است وندی در زمان گذشته، واژه‌های بسیاری ساخته باشد (مانند پسوندهای -کده، -گین، -وند و -بان)، ولی کاربران، دیگر نوواژه‌ای مطابق با آن الگو نسازند (Haspelmath & Sims, 2010, p. 129; Plag, 2003, p. 52). گفتنی است که شاهدهی نیز از نوواژه‌های ساخته شده با پسوندهایی همچون «-کده، -گین، -وند، -بان، -وش، -فام، -ناک، -زار، -گار، -سان، -گون و موارد مشابه»، در «پیکره وابستگی» به دست نیامد. البته همان گونه که پیش‌تر اشاره شد، بهترین شیوه مطالعه زایایی صرفی، کاربرد هم‌زمان شیوه‌های گوناگون سنجش کمی و بهره‌گیری از پیکره‌های مختلف است. در ادامه، خواهیم دید که اگر فقط به یک شیوه یا یک پیکره، بسنده نماییم، یافته‌های دارای روایی^۵ (اعتبار) به دست نخواهد آمد.

ب. بسامدِ موردواژه (P'=روش دوم)

کریر (Kreyer, 2015) استدلال می‌کند این روش می‌تواند در مورد تعداد تک‌بسامدی‌ها و

¹ realized productivity

² extent of use

³ potential productivity

^۴ می‌توان به طور مختصر به درجه‌ی تشابه بین واژه‌ها، از نظر دور ماندن واژه‌های کم‌بسامد و کارابودن روش در دامنه‌ی ثابتی مانند تصریف‌زمان گذشته در زبان انگلیسی اشاره کرد که فقط گزینه‌های صرفی حامل یک نقش دستوری،

مورد مقایسه‌اند.

⁵ validity

پراکنندگی نوع واژه‌های یک قاعده صرفی ویژه، کارا باشد (Kreyer, 2015). وی فرمول زیر را با تعداد نوع واژه‌هایی که به صورت مکرر رخ می‌دهد، پیشنهاد می‌کند:

$$1. \quad P' = \frac{n_1}{\sqrt{1}} + \frac{n_2}{\sqrt{2}} + \dots + \frac{nx}{\sqrt{x}} = \sum_{i=1}^x \frac{ni}{\sqrt{i}}$$

(X) بسامد نوع واژه‌های ساخته شده توسط قاعده صرفی است. این نوع سنجش وقتی بیشترین مقدار خود را به دست می‌آورد که تک بسامدی‌های بسیاری در پیکره وجود داشته باشند. برای نمونه، اگر یک نوع واژه در پیکره (الف) وجود داشته باشد و این نوع واژه ۱۰۰۰ بار و نوع واژه دیگر در همان پیکره ۵۰۰ بار رخ دهد که هر یک از این دو به طور جداگانه، سنجش خواهند شد. بنابراین، (P') توزیع بسامد کلی نوع واژه‌ها را در نظر می‌گیرد و وزن کمتری به نوع واژه‌های پُربسامد می‌دهد. درباره تعیین روایی این روش، می‌توان به پژوهشی در زبان آلمانی اشاره کرد. یافته‌های پاستیلینیکف و اشنایدر-وایژوفسکی نشان داده که این سنجش برای پسوند آلمانی (-nis)، چندان قابل قبول نیست^۱ (Pustyl'nikov & Schneider-Wiejowski, 2009, p. 3). روشن است با انجام یک یا چند پژوهش نمی‌توان یک روش دارای منطق قوی را رد نمود. این روش تاکنون برای سنجش زیابایی زبان فارسی انجام نشده است.

ج. بسامد موردواژه

روش مفیدتر برای اندازه‌گیری زیابایی، در نظر گرفتن میزان بهره‌گیری از مشتقات، یعنی بسامد موردواژه‌های صورت مشتق آن‌ها است. بسامد موردواژه، شامل همه موارد وقوع نوع واژه با همه صورت‌های آن است. مشکل روش شناختی این سنجش، تعیین پایایی^۲ (اعتماد) تعداد نوع واژه‌های یک دوره معین است. بوی (Booii, 2018) می‌نویسد: «در حالی که شواهد آشکاری در مورد اثرات تنوع ساخت و بسامد موردواژه وجود دارد، باید کار بیشتری برای بررسی شرایطی که در آن بسامد بالای موردواژه، زیابایی یک ساخت را کم و زیاد می‌کند، انجام گردد» (Booii, 2018, p. 566). در پژوهش فرزانه (Farzane, 2016) از کتاب‌های دستور زبان فارسی، وندهای در زیر عنوان دارندگی استخراج گردیده است و با مراجعه به فرهنگ لغت فارسی زانسو، بسامد ۱۸ نند دارای معنی دارندگی سنجیده شده است (Farzane, 2016). یافته‌های به دست آمده از این پژوهش نمایانگر آن است که میزان زیابایی وندهای مورد نظر به این شرح است: «سی» ۶۰۰، «ننده» ۳۹۲، «دار» ۳۸۲، «گر» ۲۳، «ناک» ۱۲۰، «مند» ۹۱، «وار» ۷۶،

^۱ برای مطالعه بیشتر نگاه کنید به پاستیلینیکف و اشنایدر-وایژوفسکی (Pustyl'nikov & Schneider-Wiejowski, 2009).

^۲ reliability

«پذیر» ۶۱، «او» ۵۶، «زار» ۵۲، «سار» ۵۱، «خیز» ۴۶، «ور» ۲۲، «گین» ۱۷، «وند» ۱۴، «آگین» ۱۳، «لو» ۱۱، «کی» ۱۰. همان گونه که مشاهده می‌شود ارقام ارائه‌شده، بسامد «موردواژه» است و نه احتمال زایایی که فقط می‌تواند بین دو عدد صفر و یک، مقدار عددی بپذیرد.

همان گونه که شرح داده خواهد شد، برای اندازه‌گیری دقیق‌تر زایایی به واژه‌های تک‌بسامدی و نوواژه‌ها نیز نیاز داریم. زیرا با وجود شواهد متفاوت فعلی، بهره‌گیری از این روش به عنوان روش بی‌نقص و کامل در مطالعات زبان فارسی، می‌تواند به نتایج گمراه‌کننده‌ای منجر گردد که در زیربخش (۴.۱.۳)، به آن اشاره می‌شود.

۲.۲.۳. تعیین سهم عددی^۱ نوواژه‌ها

با شمارش تعداد نوواژه‌های تأییدشده در زمان معین، زایایی در زمانی سنجیده می‌شود. به باور پلگ (Plag, 2003)، شمارش مشتقات یک‌وند، می‌تواند راهی پرفایده برای تعیین زایایی باشد، البته اگر فقط مشتقات جدید یعنی نوواژه‌های تولیدشده در یک دوره‌ی معین، اندازه‌گیری گردد (Plag, 2003, p. 110). این معیار را می‌توان با یک فرهنگ لغت تاریخی دارای پایایی، تعیین نمود. ولی معمولاً مدخل‌نویسان از واژه‌های جدید ساخته‌شده با الگوهای بسیار زایا، غافلند. روش دیگر، بهره‌گیری از پیکره‌های بزرگ است. برای نمونه، می‌توان مشاهده کرد که چگونه پسوند/مان/ و شبه‌پسوند (-gate)، در واژه‌هایی هم‌چون «گفتمان، سیرمان، چیدمان، یادمان» و (Watergate, Irangate)، دیده می‌شود و معنای آن این است که این وندها در دهه‌های اخیر، زایایی جدیدی به دست آورده‌است (Haspelmath & Sims, 2010, p. 114). در بخش پسین، به نکته‌های دیگری درباره‌ی اهمیت سنجش نوواژه‌ها اشاره خواهد شد. عباسی (Abbassi, 2006)، نیز در این باره بیان نموده‌است که مابین زایایی یک فرایند صرفی در زبان حال و تعداد نوواژه‌های ساخته‌شده به وسیله آن فرایند، رابطه مستقیم وجود دارد.

۳. مبانی نظری

در مبانی نظری این مقاله، سنجش‌های مختلفی به کار گرفته می‌شوند که به آن می‌پردازیم. این

¹ proportion

(proportion) یا سهم عددی، سنجش نسبت کمیت جزء به کل است؛ به گونه‌ای که صورت کسر در مخرج کسر نیز وجود دارد (سهم عددی تک‌بسامدی‌های دارای پسوند -مند) به تمامی تک‌بسامدی‌های همه پسوندها که (-مند) یکی از آنها است).

دسته از سنجش‌های کمی، چهارمین مدل‌های مورد بحث نیز هستند که در تحلیل داده‌ها از آن‌ها استفاده می‌شود.

۳.۱. مدل‌های تعیین سهم عددی تک‌سامدی‌ها (برآورد احتمال)

پلگ (Plag, 2003) از تک‌سامدی‌ها برای اندازه‌گیری زیایی بهره گرفته‌است. با افزایش تعداد واژه‌های تک‌سامدی در پیکره، تعداد نوواژه‌ها (زیایی وند) بیشتر می‌شود.

۳.۱.۱. زیایی به معنای دقیق^۱ (p)

برای برآورد احتمالی نوواژه‌ها در میان واژه‌ها، سهم عددی تک‌سامدی‌های یک وند معین با تعداد کل موردواژه‌های موجود آن وند، اندازه‌گیری می‌شود. این احتمال به وسیله باین و لیبر (Baayen & Lieber, 1991)، با عنوان «زیایی به معنای دقیق»، نامیده شده و پلگ (Plag, 2003)، آن را با فرمول زیر بیان کرده‌است:

$$2. P = n^{1aff} / N^{aff}$$

(n^{1aff})، نشانگر تعداد تک‌سامدی‌های یک وند ویژه (aff) و (N^{aff})، تعداد کل موردواژه‌های آن است. زیایی وند این گونه محاسبه و تفسیر می‌شود: تعداد زیاد تک‌سامدی‌ها، منجر به افزایش (P) یا زیایی زیاد می‌شود. در مقابل، شمار بسیار واحدهای واژگانی پُرسامد منجر به ارزش بالای (N^{aff}) و کاهش (P) می‌شوند که نشانگر زیایی کم است (Plag, 2003, p. 81). باید دانست این روش، تغییر و گسترش فرمول آرونوف (Aronoff, 1983)، باین (Baayen, 1989)، باین و لیبر (Baayen & Lieber, 1991)، و باین (Baayen, 1992)، است که به عنوان نخستین روش فرموله شده به صورت [۳] بوده‌است:

$$3. P = n1/N$$

۳.۱.۲. درجه زیایی مشروط به تک‌سامدی‌ها^۲

سهم عددی واژه‌های تک‌سامدی با الگوی ویژه به تعداد کل موردواژه‌های شکل گرفته با همان الگو در پیکره (به نام معیار p)، یا «درجه زیایی مقوله مشروط» است (Baayen & Lieber, 1993; Baayen, 1991). این معیار از این مشاهده‌ها ناشی می‌شود که قاعده‌های زیای احتمالاً سبب ایجاد «گاه‌واژه^۳» می‌شوند که اهمیت ویژه‌ای در تعیین زیایی دارند. در یک پیکره بسیار

¹ productivity in the narrow sense

² hapax-conditioned degree of productivity

³ to create occasionalism

بزرگ، می‌توان تک‌بسامدی‌ها را، واژه‌هایی دارای این ویژگی فرض کرد:

$$4. P = V_{1,m} / N_m$$

$(V_{1,m})$ ، تعداد واژه‌های تک‌بسامدی در پیکره با الگوی صرفی (m) ، و (N_m) بسامد موردواژه‌های همه واژه‌های دارای الگوی (m) را نشان می‌دهد. سهم عددی، نشان‌دهنده احتمال است که واژه استخراج شده از پیکره، که الگوی ویژه‌ای را به نمایش گذارده، دارای ویژگی آن الگوست.

۳.۱.۳. زایایی جهانی^۱ (P^*)

باین (Baayen, 2009)، روش پیشین را میزان اشباع یک مقوله دانسته‌است؛ یعنی یک قاعده با خطرپذیری کم اشباع، امکان بیشتری برای گسترش دارد و آن را در ابتدا به صورت زیر نشان داده‌است:

$$5. P = V(1,C,N) / N(C)$$

در این روش، در واقع سهم عددی تک‌بسامدی‌های هر قاعده $(V(1,C,N))$ ، نسبت به همه تک‌بسامدی‌های شناسایی‌شده $(N(C))$ ، یعنی معیار (P^*) ، سنجیده می‌شود. Baayen & Lieber (1993; Baayen, 1991). این سنجش، افزون بر اشاره به گرایش قاعده‌های زایا در ایجاد «گاه‌واژه»، به این موضوع می‌پردازد که تا چه اندازه هر قاعده ویژه به میزان رشد تمامی واژه‌ها کمک می‌کند. باین (Baayen, ۱۹۹۳)، صورت‌بندی پیشین [۴] را با اندک تغییری، به صورت [۶] ارائه می‌کند و تأکید دارد این نسبت، میزان رشد تمامی واژه‌های مربوط به مقوله صرفی را تخمین می‌زند:

$$6. P^* = V_{1,m} / V_1$$

در این فرمول همچون روش پیشین، (V_1) ، تعداد کل تک‌بسامدی‌های پیکره است، ولی برتری دیگری دارد که می‌تواند زایایی وند ویژه‌ای را در زمانی معین، اندازه‌گیری کند. باین (Baayen, 2009) با این روش، میزان پیشرفت مقوله صرفی و تولید عناصر جدید را ارزیابی می‌کند. به این ترتیب که مقوله (C) ، که با سرعت بالاتری در حال گسترش است، نسبت به مقوله‌ای که با سرعت کمتری گسترش می‌یابد و یا به هیچ وجه گسترشی ندارد، زایاترست (Baayen, 2009). باین گسترش زایایی^۲ را با کمی تغییر، این‌گونه اندازه‌گیری می‌کند:

$$7. P = V(1,C,N)$$

¹ global productivity

(P^*) این نشانه برای تمایز سنجش زایایی با روش زایایی جهانی از روش‌های پیشین (P) است.

² expanding productivity

این سنجش زیایی با واژه‌های تک‌بسامدی همه مقوله‌ها ($V(1,C)$)، با یک پیکره معین (N) تخمین زده می‌شود. بنابراین فرمول [۸]، برآوردی از سهم عددی مقوله خاص به میزان رشد تمامی واژه‌ها است (Baayen, 1993).

$$8. P^* = V(1,C,N)/V(1,C)$$

فرمول [۸]، معادل فرمولی است که باور (Bauer, 2001)، در فرمول [۹]، با عنوان زیایی جهانی صورت‌بندی می‌کند:

$$9. P^* = (n_{1Et}) / h_1$$

(E) مقوله خاص، (t) تعداد موردواژه‌های موجود در پیکره و (h_1) معرف تعداد تک‌بسامدی‌هاست. سنجش‌های پیکره‌ای مبتنی بر تک‌بسامدی‌ها (P^* , P) به طور چشمگیری استفاده می‌شوند.

در سطرهای پیشین، ارتباط بین فرمول‌های مطرح‌شده در پژوهش‌های گوناگون طی سه دهه، به روشنی بیان گردید. همچنین مشاهده شد که روش‌های سنجش، ناسازگاری‌هایی با یکدیگر دارند ولی در برخی روش‌ها نیز، همپوشانی وجود دارد. باید در نظر داشت که هر معیار و سنجش، بعد متفاوتی از زیایی را برجسته می‌کند. آنچه از خلاصه بحث اخیر به دست می‌آید آن است که اصطلاح «گاه‌واژه» (Baayen, 1993)، همان چیزی است که باین و لیبر، آن‌ها را «تک‌بسامدی» نامیده بودند و برای نمایش آن از نمادهای متفاوتی مانند $V_{1,m}$ ، n^{aff} و $V(1,C,N)$ بهره گرفتند (Baayen & Lieber, 1991). همچنین اصطلاح‌های «میزان اشباع یک مقوله» و «رشد تمامی واژه‌ها» نیز به مجموع «موردواژه‌های» همان مقوله مورد بررسی در پیکره اشاره دارد که باین (Baayen, 2009)، آن‌ها را به کار برده است. نمادهای مورد استفاده برای این مفهوم نیز N^{aff} ، N_m و N_c است. در پژوهش حاضر، برای سنجش زیایی از چهار روش بسامدی و دو روش برآورد احتمالی بهره گرفته شده است.

۴. روش‌شناسی

این پژوهش پیکره‌بنیاد، مطالعه‌ای توصیفی-تحلیلی است که به صورت همزمانی به انجام رسیده است. پیکره مورد استفاده در این پژوهش دادگان «پیکره وابستگی^۱» زبان فارسی است که حدود سی‌هزار جمله و ۴۶۹۶۴۳ واژه دارد. این دادگان پیکره نحوی زبان فارسی است و دارای برجسب‌های صرفی و نحوی است (Rasooli et al., 2013). این پیکره دارای برجسب‌های معنایی و تحلیل‌گفتمانی نیز شده است (Mirzaei & Safari, 2018). این پژوهش بر اساس ۵۰

^۱ این پیکره توسط مرکز تحقیقات کامپیوتری علوم اسلامی نور تهیه شده است.

پسوند معرفی شده به وسیله صادقی (Sadeghi, 1991-1993) انجام پذیرفته است. او در این مجموعه، ۷۶ پسوند غیر فعلی زبان فارسی را مورد بررسی قرار می‌دهد و آن‌ها را به دو گروه عمده فعلی و غیر فعلی بخش می‌کند. این تعداد پسوند با توجه به نوع پسوند و نوع مقوله پایه واژه‌ای که به آن افزوده شده، معرفی شده است. تعداد صورت پسوندهای اشتقاقی غیرفعلی، ۵۰ عدد بوده است که در جدول زیر فهرست می‌شود.

جدول ۱: فهرست پسوندهای اشتقاقی غیر فعلی زبان فارسی (Sadeghi, 1991-1993)، به تفکیک واژه‌ها (نوع واژه‌ها، مورد واژه‌ها، تک‌بسامدی‌ها، نوواژه‌ها) و چند مثال از پیکره

مورد مطالعه

ردیف	نام پسوند	پسوند (IPA)	نوع واژه‌ها	مورد واژه‌ها	تک‌بسامدی‌ها	نوواژه‌ها	مثال‌ها
۱	-ئی	I	۶۷۷	۱۱۰۴	۴۹۷	۱۷	مالی، زبانی
۲	-ه	E	۱۹۲	۳۷۷	۱۲۳	۳	افزونه، سامانه
۳	-ن	Æn	۱۰۹	۱۴۱	۹۲	۵	عمیقاً، قاعدتاً
۴	-ش	ɛʃ	۵۵	۱۱۵	۳۰	۳	گرایش، پردازش
۵	-یت	Jæt	۴۰	۸۰	۲۱	۰	عقلانیت، روحانیت
۶	-دار	Dar	۳۶	۶۳	۲۷	۳	زاویه‌دار، هوادار
۷	-انه	Ane	۳۳	۴۸	۲۷	۲	منفعلان، ناقدانه
۸	-ان	An	۳۲	۵۸	۲۲	۲	پیام‌رسان، لرزان
۹	-یه / -یه	Ijje	۲۹	۵۹	۱۶	۰	سهمیه، فرضیه
۱۰	-گان	Gan	۲۵	۴۶	۱۸	۱	دادگان، خبرگان

ردیف	نام پسوند	پسوند (IPA)	نوع واژه‌ها	مورد واژه‌ها	تک بسامدی‌ها	نوع واژه‌ها	مثال‌ها
۱۱	-نده	Ænde	۲۵	۴۷	۱۵	۰	برنده، دمنده
۱۲	-گاه	Gah	۲۲	۲۲	۱۰	۰	گذرگاه، نمایشگاه
۱۳	-کار	Kar	۱۸	۲۲	۱۵	۰	راهکار، سوارکار
۱۴	-ک	Æk	۱۷	۶۳	۱۷	۱	موشک، پیامک
۱۵	-گر	Gær	۱۶	۴۸	۱۰	۱	حسگر، سینماگر
۱۶	-ستان	Estan	۱۶	۳۰	۸	۰	شهرستان، گلستان
۱۷	-بندی	Bændi	۱۲	۱۵	۱۰	۰	رده‌بندی، درزبندی
۱۸	-مند	Mænd	۱۲	۱۸	۷	۲	هدفمندی، توانمندی
۱۹	-ین	In	۱۲	۳۹	۶	۰	زیرین، سنگین
۲۰	-کاری	Kari	۱۷	۱۷	۹	۰	چمن‌کاری، کم‌کاری
۲۱	-ل	A	۱۱	۲۶	۸	۰	سازگارا، جویا
۲۲	-گار	Gar	۱۰	۱۴	۷	۰	رستگار، ماندگار
۲۳	-آر	Ar	۹	۲۳	۳	۰	دیدار، گرفتار
۲۴	-بان	Ban	۸	۱۵	۵	۰	میزبانی، قلعه‌بانی

ردیف	نام پسوند	پسوند (IPA)	نوع واژه‌ها	مورد واژه‌ها	تک‌بسامدی‌ها	نوع واژه‌ها	مثال‌ها
۲۵	-دان / دانی	dan/dani	۸	۱۱	۵	۰	زندان، زندانی
۲۶	-وار	Var	۷	۷	۵	۲	کابوس‌وار، شبح‌وار
۲۷	-ور	Vær	۷	۷	۱	۰	داور، سخنوری
۲۸	-گری	Gæri	۶	۹	۴	۱	مادی‌گری، غریبالگری
۲۹	-بازی	Bazi	۴	۴	۴	۰	آب‌بازی، چتربازی
۳۰	-یار	Jar	۴	۴	۴	۰	دهیار، هوشیار
۳۱	-ناک	Nak	۴	۵	۳	۰	سوزناک، دردناکی
۳۲	-واره	Vare	۴	۷	۳	۱	رنگواره، ماهواره
۳۳	-گون	Gun	۴	۸	۲	۰	واژگون، گوناگون
۳۴	-مان	Man	۴	۹	۲	۱	گفتمان، زایمان
۳۵	-چه	dʒe	۳	۳	۳	۰	دریاچه، باغچه
۳۶	-چی	dʒi	۳	۳	۳	۰	حریرچی، مسلسل‌چی
۳۷	-او	U	۳	۵	۲	۰	نیکو، چاقو
۳۸	-وند	Vænd	۳	۱۷	۱	۰	شهروند، خداوند
۳۹	-ند	Ænd	۳	۱۵	۰	۰	مانند، فرایند

ردیف	نام پسوند	پسوندها (IPA)	نوع واژه‌ها	مورد واژه‌ها	تک‌بسامدی‌ها	نوع واژه‌ها	مثال‌ها
۴۰	-آباد	Abad	۲	۲	۲	۰	ناکجا آباد، خرم آباد
۴۱	-کی	Æki	۲	۲	۲	۰	دزدکی، آبکی
۴۲	-ینه	Ine	۲	۲	۲	۰	کمینِه، گزینه
۴۳	-آسا	asa	۱	۱	۱	۱	مرگ آسا،
۴۴	-گانه	gane	۱	۱	۱	۰	جداگانه
۴۵	-گین	gin	۱	۱	۱	۰	سنگین، خشمگین
۴۶	-زار	zar	۱	۱	۱	۰	چمن زار
۴۷	-فام	fam	۰	۰	۰	۰	-
۴۸	-کده	kæde	۰	۰	۰	۰	-
۴۹	-سان	san	۰	۰	۰	۰	-
۵۰	-وش	væf	۰	۰	۰	۰	-

۴. ۱. استخراج داده‌ها

در این پژوهش به منظور بررسی پسوندهای اشتقاقی غیر فعلی زبان فارسی معاصر، پیکره وابستگی وارد برنامه اکسل شد.^۱ همه پسوندهای شرکت کننده در ساخت واژه‌ها از پیکره‌ی ورد اشاره استخراج گردید. آن گاه تمام موارد به طور جداگانه تفکیک و به صورت دستی تحلیل شد. از آن جا که یکی از امکانات مهم زبان فارسی، افزودن وندهای بومی به وام‌واژه‌ها است؛ تقطیع تکواژی و ریشه‌یابی واژه‌ها از مهم‌ترین عملیات تجزیه و تحلیل واژه‌های پسونددار در این پیکره بود.

^۱ لازم به ذکر است که نخست برای تحلیل و تفکیک داده‌ها از واژه‌پرداز ورداسمیت (Word Smith)، استفاده شد. راستی‌آزمایی در تشخیص پسوندها، حدود ۰٫۲۶ درصد به دست آمد. با توجه به پایین بودن دقت تفکیک در این پیکره زبان فارسی، تحلیل به صورتی که بیان می‌شود، ادامه یافت.

^۲ زبان‌شناسان در مطالعات صرفی برای ریشه از اصطلاح (root) بهره می‌گیرند. به گونه‌ای که با حذف وندها به ریشه‌ای می‌رسیم که مشتمل بر معنای اصلی واژه‌است. ولی آنچه در این مطالعه مورد نظر بوده است وندافزایی است که پسوند ویژه‌ای به پایه افزوده می‌گردد. بنابراین اصطلاح پایه (base) در معنای صورت بی‌نشان و مستقل واژه در مدخل‌های واژگان ذهنی، مورد نظر است.

هدف از انجام این عملیات، یافتن پایه، تعیین نوع مقوله پایه و تعیین انواع پسوندهای موجود در هرواژه بوده است. برای رسیدن به این هدف در تحلیل دستی، از فرهنگ ریشه‌شناسی زبان فارسی محمد حسن دوست، فرهنگ‌های واژگانی دهخدا و معین، بهره گرفته شد. با استفاده از فرهنگ‌های اشاره شده، همه مواردی که ریشه‌ای از زبان‌های دیگر، مانند انگلیسی، فرانسوی، آلمانی و موارد مشابه داشت، حذف گردید. ولی پسوندهای زبان فارسی افزوده شده به وام‌واژه‌های عربی بسیار متداول در زبان فارسی نگهداشته شد. زیرا گویشوران زبان فارسی، چنین واژه‌هایی را همانند واژه‌های بومی می‌انگارند. بر این اساس، برای تعیین زایایی، همه پسوندها (۵۰ پسوند) استخراج و در جدول‌های جداگانه فهرست گردید. پسوندها از جنبه ساختار پایه مورد بررسی قرار گرفتند و پسوندهای پیوسته شده در هر ترکیب مشخص شد و در طبقه‌بندی‌های گوناگون قرار گرفت.

۴.۲. تحلیل خوشه‌ای

برای دسته‌بندی پسوندها بر مبنای بسامد، از تحلیل خوشه‌بندی^۱ استفاده شد. در این ارتباط باید اشاره کرد؛ روش‌های گوناگونی برای خوشه‌بندی داده‌ها در علم آمار وجود دارد که می‌توان به الگوریتم‌های k-میانگین^۲، خوشه‌بندی سلسله‌مراتبی^۳، خوشه‌بندی دومرحله‌ای^۴ و خوشه‌بندی نزدیک‌ترین همسایه^۵ اشاره کرد. در این الگوریتم‌ها، بر مبنای قواعد آماری، اعداد نزدیک به هم در یک دسته یا اصطلاحاً یک خوشه قرار می‌گیرند. باید بیان کرد که همیشه یافته‌های به دست آمده از اجرای (الگوریتم‌های) متفاوت یکسان نیست. ولی این مسئله اهمیت چندانی ندارد زیرا تنها هدف پژوهشگر از انجام تحلیل خوشه‌ای، دسته‌بندی داده‌ها در دسته‌های دلخواه برای تحلیل مورد نظر است. ما به منظور دسته‌بندی زایایی با بسامد نوع‌واژه‌ها، موردواژه‌ها، تک‌بسامدی‌ها و نوواژه‌ها، از (الگوریتم) کای-میانگین^۶ بهره گرفتیم. با استفاده از نرم‌افزار آر، نسخه ۳، ۴، ۰، ۳ (RV4.0.3)، داده‌ها در ۵ خوشه دسته‌بندی شد. اجرای محاسبات (الگوریتم) مشخص کرد که چه پسوندهایی باید در یک خوشه قرار گیرند. پس از اجرای محاسبات (الگوریتم) و مشخص شدن خوشه‌ها، نمودارهای ستونی هر خوشه، به همراه میانگین بسامد پسوندهای موجود در هر خوشه رسم گردید.

^۱ clustering analysis

^۲ K-means clustering

^۳ hierarchical Clustering

^۴ two step clustering

^۵ nearest neighborhood clustering

^۶ در الگوریتم کای-میانگین (k-میانگین)، منظور از درج (k-)، در الگوریتم (k-میانگین)، در نظر گرفتن چند میانگین در این نوع (الگوریتم) است.

۵. یافته‌ها

۵.۱. تحلیل بسامدی

۵.۱.۱. تحلیل خوشه‌بندی نوع واژه‌ها

با توجه به بسامد نوع واژه‌ها، از روش خوشه‌بندی پایگانی استفاده شد و پسوندها در ۵ خوشه (دسته) دسته‌بندی شدند. نمودار ستونی به دست آمده از خوشه‌بندی در زیر نمایش داده شده است. در نمودار ستونی، میانگین بسامد پسوندهای موجود در هر خوشه در نظر گرفته شده است. میزان زایایی هر پسوند با عنوان‌های غیر زایا، با زایایی پایین، زایایی متوسط، با زایایی بالا و کاملاً زایا، از بالا به پایین با الگوهای مختلف مشخص شده است. در همه نمودارها، الگوی مربع چین تیره و روشن، نشان‌دهنده خوشه کاملاً زایاست. الگوی نقطه چین کم تراکم، مربوط به خوشه غیر زایاست. سه خوشه با زایایی بالا، خوشه زایایی متوسط و خوشه با زایایی پایین در بین دو خوشه کاملاً زایا و خوشه غیر زایا نشان داده شده‌اند. برای نشان دادن این سه نوع زایایی به ترتیب از الگوهای خطوط عمودی، خطوط افقی و نقطه چین پُر تراکم بهره گرفته شده است.^۱

جدول ۲: جدول راهنمای الگوهای استفاده شده در نمودارها

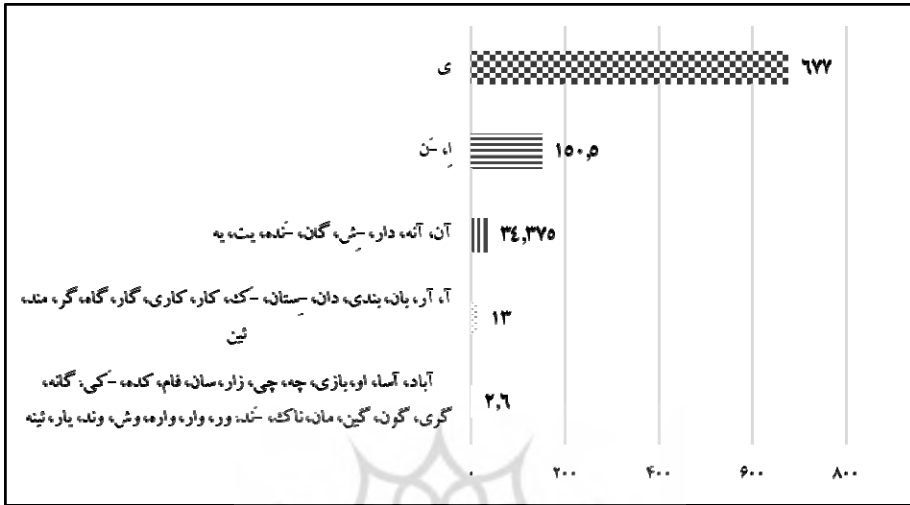
				
مربع چین	نقطه چین پُر تراکم	خطوط افقی	خطوط عمودی	نقطه چین کم تراکم
خوشه کاملاً زایا	خوشه با زایایی پایین	خوشه زایایی متوسط	خوشه با زایایی بالا	خوشه غیرزایا

همان گونه که در جدول بالا مشاهده می‌شود پسوند «-ی»، بیش‌ترین بسامد (۶۷۷)، را در بین پسوندها دارد و خوشه کاملاً زایا را به تنهایی تشکیل داده است. پسوندهای «-ن» و «-ا»، نیز با میانگین بسامد (۱۵۰/۵)، در خوشه با زایایی بالا قرار گرفته‌اند. بیشتر پسوندها در خوشه‌هایی با میانگین پایین قرار گرفته‌اند. همان گونه که دیده می‌شود در این روش از بسامد و میانگین آن بهره گرفته شده است.

^۱ لازم به اشاره است که به دلیل محدودیت فضا، نام پسوندها در هر نمودار، بدون نشانه خط تیره (-) فهرست شده است.

^۲ -ن مصدری و -ا-ه.

این تحلیل اعدادی را در اختیار گذارده است که ممکن است فقط در این پیکره، دارای روایی باشد. همچنین این نتیجه مربوط به زایایی گذشته است و در باره زایایی کنونی، دارای پایایی نیست.



شکل ۲: نمودار ستونی میانگین بسامد نوع واژه‌ها

۵. ۱. ۲. تحلیل خوشه‌بندی موردواژه‌ها

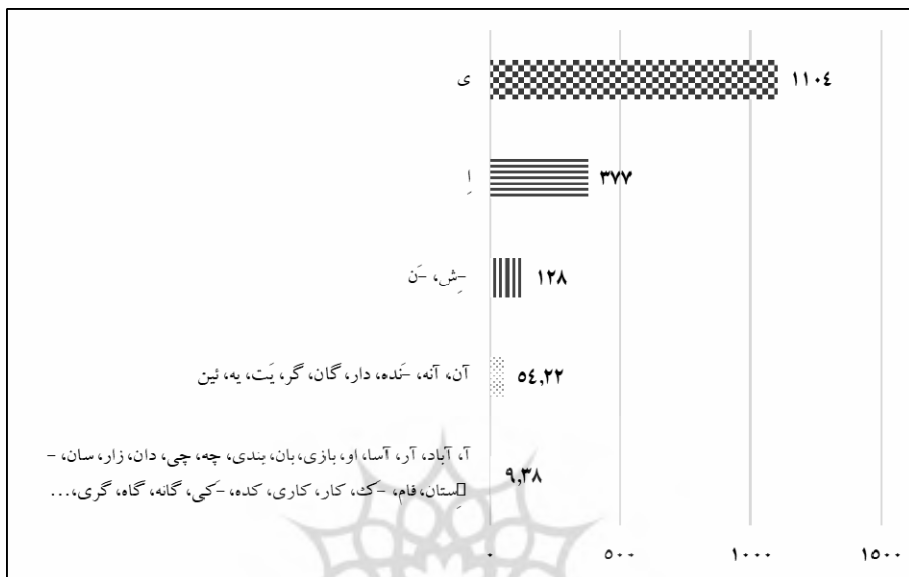
بر مبنای خوشه‌بندی پایگانی، پسوندها بر اساس بسامد موردواژه‌ها، در ۵ خوشه قرار گرفتند. در این نمودار پسوند «ی» با بیشترین بسامد (۱۱۰۴)، در خوشه کاملاً زایا قرار دارد. پسوند «ا» نیز به تنهایی با بسامد (۳۷۷)، خوشه با زایایی بالا را تشکیل داده است. از این نمودار می‌توان نتیجه گرفت که بیشتر پسوندها دارای بسامد موردواژه‌ای پایین هستند. همان‌گونه که مشاهده می‌شود اگر در سنجش زایایی صرفی به دو روش بسامد نوع واژه و موردواژه بسنده شود، با چند مشکل روبه‌رویم:

نخست آنکه در خوشه با زایایی بالا، یک‌بار پسوندهای «ن» و «ا» و یک‌بار پسوند «ا»، قرار می‌گیرند. یعنی پسوند «ن»، مطابق نوع واژه، در خوشه با زایایی بالا و مطابق موردواژه در خوشه زایایی متوسط قرار می‌گیرند. با نگاه گذرای به سه خوشه دیگر نیز می‌توان به تفاوت‌های موجود پی برد.

ایراد دوم این سنجش آن است که میزان کم و زیاد بودن زایایی پسوندها، تا اندازه‌ای مشخص است ولی نه رقم ثابتی است و نه زایایی پسوندها، قابل مقایسه است.

مشکل سوم به این نکته بر می‌گردد که حتی در صورت چشم‌پوشی از موارد گفته‌شده، این سنجش فقط نتیجه‌ای از زایایی گذشته و میزان استفاده کاربران از واژه‌های موجود زبان است. به

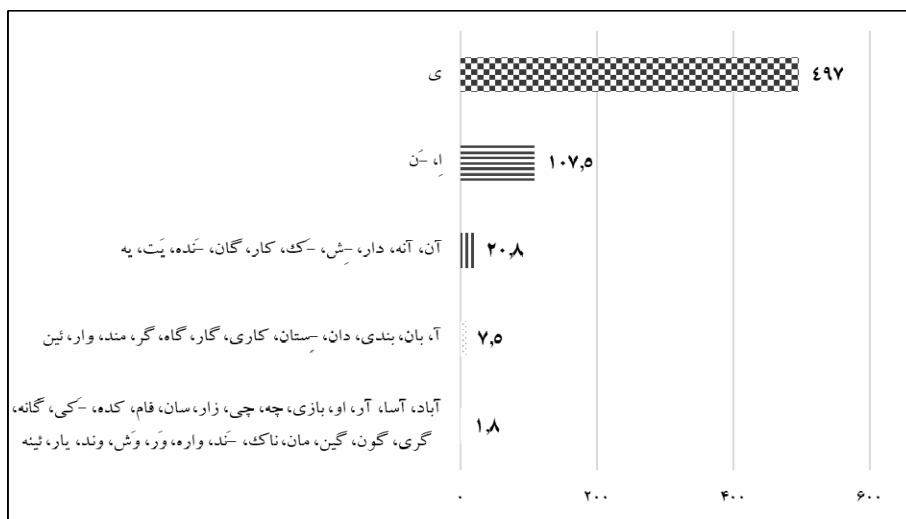
دیگر سخن، سنجش‌های مربوط به تعیین بسامد نوع واژه‌ها و موردواژه‌ها در دقیق‌ترین حالت نیز، نمایشی از زایایی صرفی کنونی به دست نمی‌دهد.



شکل ۲: نمودار ستونی میانگین بسامد موردواژه‌ها

۵. ۱. ۳. تحلیل خوشه‌بندی تک‌بسامدی‌ها

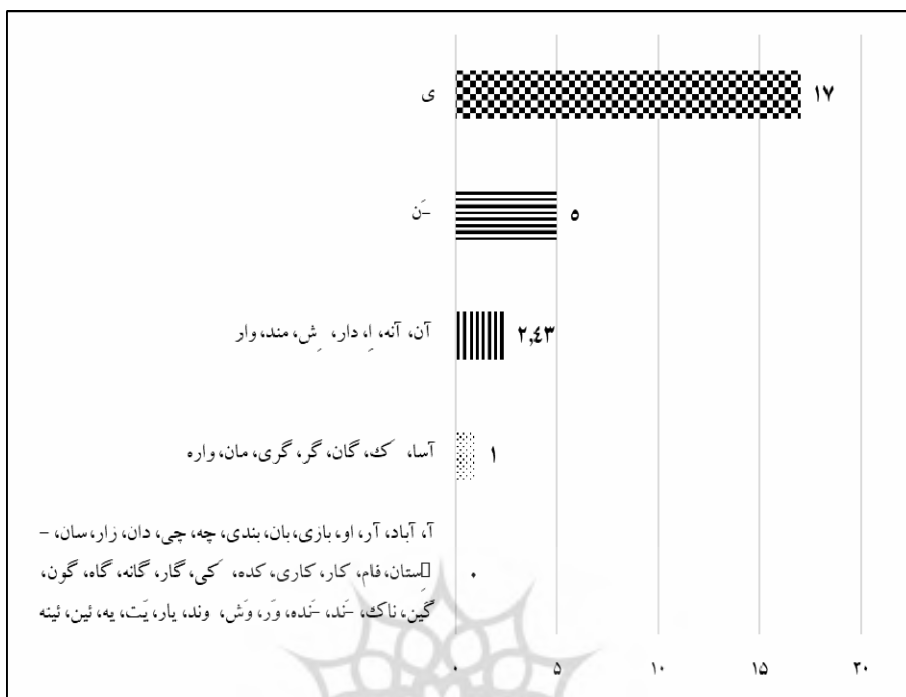
جدول زیر، نمودار ستونی میانگین بسامد تک‌بسامدهای پیکره را در هر خوشه نشان می‌دهد. این بار نیز پسوند «ی» با بسامد (۴۹۷)، به تنهایی در خوشه کاملاً زایا در این پیکره شناخته می‌شود. همچنین دو پسوند «-ن» و «-ا» نیز با میانگین بسامد (۱۰۷/۵)، در خوشه با زایایی بالا جای گرفتند. پسوندهای «-آن، -آنه، -دار، -ش، -ک، -کار، -گاه، -نده، -یت، -یه»، با میانگین بسامد (۲۰/۸)، در خوشه با زایایی متوسط قرار گرفتند. بیش تر پسوندها در دو خوشه‌ی آخر با میانگین بسامد (۷/۵) و (۱/۸)، جای دارند که به ترتیب مربوط به خوشه با زایایی پایین و خوشه غیر زایا شناخته می‌شود.



شکل ۳: نمودار ستونی میانگین بسامد تک‌بسامدی‌ها

۵. ۱. ۴. تحلیل خوشه‌بندی نوواژه‌ها

نتیجه‌ی به‌دست‌آمده از خوشه‌بندی نوواژه‌ها در نمودار ستونی زیر نمایش داده شده‌است. در این نمودار، میانگین بسامد نوواژه‌های هر پسوند در خوشه‌های آماری در نظر گرفته شده‌است. همان‌گونه که مشاهده می‌شود پسوند «-ی»، با بیش‌ترین بسامد (۱۷)، در بین پسوندها، خوشه کاملاً زایا را تشکیل می‌دهد. همچنین پسوند «-ن» با بسامد (۵)، در خوشه زایایی بالا قرار گرفته‌است. بیشتر پسوندها در خوشه‌ای با میانگین صفر در خوشه‌ی غیر زایا قرار گرفته‌اند و این دال بر کم‌بودن بسامد نوواژه‌ها در این پیکره‌هاست. بنابراین، این تحلیل نیز همانند تحلیل بسامد تک‌بسامدی‌ها نمی‌تواند زایایی صرفی را در هر پیکره‌ای با هر حجم نشان دهد. این موضوع در پیکره‌های کوچکی چون «پیکره وابستگی»، ما را از واقعیت زایایی صرفی پسوندها دور می‌کند.



شکل ۴: نمودار ستونی میانگین بسامد نوواژه‌ها

۲.۵. تحلیل سهم عددی یا برآورد احتمالی^۱

۲.۵.۱. خوشه‌بندی پسوندها بر اساس زایایی دقیق

همان‌گونه که اشاره شد، زایایی صرفی مطابق واژه‌های تک‌بسامدی، به دو روش عمده سنجیده می‌شود. یکی از این روش‌ها به «زایایی دقیق»، موسوم است. در این روش تعداد تک‌بسامدی‌های هر مقوله معین بر بسامد کل موردواژه‌های خود در پیکره دسته‌بندی می‌شود. برای نمونه، می‌توان گفت در «پیکره وابستگی»، پسوند «-ی» دارای (۶۷۷) نوع واژه، (۱۱۰۴) موردواژه، (۴۹۷) تک‌بسامدی و مجموع تک‌بسامدی‌های همه پسوندها نیز (۱۰۵۵)، بوده است. بر مبنای فرمول شماره [۵]، در صفحه‌های پیشین، زایایی دقیق این پسوند، چنین سنجیده می‌شود:

$$10. \text{ زایایی دقیق} = \frac{497}{1104} = 0/4501$$

البته در این بخش نیز، پسوندها بر اساس میزان زایایی دقیق‌شان در پیکره مورد نظر، خوشه‌بندی

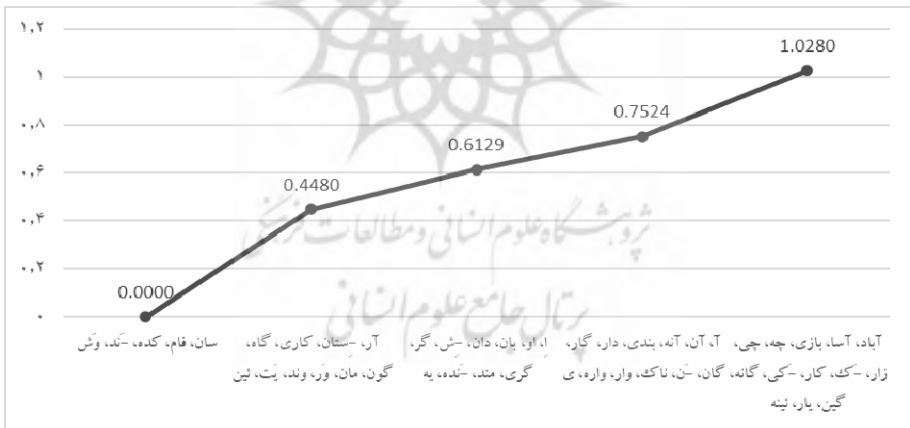
^۱ همان‌طوری که در بخش‌های گذشته از نظر گذشت، در سنجش سهم عددی، هر صورت، جزیی از مخرج کسر است. در سنجش احتمال ساخت یک کلمه جدید توسط کاربران، در هر دو نوع سنجش زایایی دقیق و زایایی جهانی، مقدار صورت در مخرج نیز وجود دارد. لذا در این تحلیل می‌توانیم سنجش سهم عددی و برآورد احتمالی را در معنای هم‌راستا با هم، به کار ببریم.

شده‌اند. پس از محاسبه‌ی زایایی دقیق هر پسوند مطابق فرمول و اجرای خوشه‌بندی سلسله‌مراتبی، پسوندها در ۵ خوشه دسته‌بندی شدند. مقدار میانگین زایایی دقیق هر خوشه بالای هر نقطه نشان داده شده‌است. در این نمودار خطی، سه نکته مهم وجود دارد:

یکم- در این روش، موردواژه‌های هر پسوندی که در مخرج کسر قرار می‌گیرد، متغیر است. بنابراین رقم مربوط در سنجش زایایی صرفی برای هر پسوند در پیکره‌ها با حجم‌های متفاوت، پیوسته تغییر می‌کند.

دوم- پسوند «-زار» در این سنجش، دارای زایایی کامل (یک) بوده‌است. در حالی که این به سبب برابر بودن بسامد نوع واژه، موردواژه و تک‌بسامدی (یعنی یک مورد) بوده‌است. پس این نوع سنجش در پیکره‌های کوچک که معمولاً بسامد نوع واژه، موردواژه و تک‌بسامدی‌ها، پایین است، یافته‌های کاملاً گمراه‌کننده‌ای را به دست می‌دهد.

سوم- حاصل جمع مقادیر زایایی بیشتر از عدد یک است (یعنی حاصل جمع پنج خوشه). به بیان دیگر، مفهومی غیر از احتمال زایایی، سنجش شده‌است. پیش از این اشاره شد که احتمالی زایایی باید مقادیری بین عددهای صفر و یک باشد.



شکل ۵: نمودار خطی یا پیوستار احتمالی زایایی پسوندها بر مبنای زایایی دقیق

۵.۲.۲. خوشه‌بندی پسوندها بر اساس زایایی جهانی

روش دیگر سنجش زایایی صرفی بر مبنای واژه‌های تک‌بسامدی، روش «زایایی جهانی» است. در این روش، تعداد تک‌بسامدی‌های هر مقوله معین بر مجموع تک‌بسامدی‌های پیکره دسته‌بندی می‌شود. بر مبنای مثال ارائه‌شده در بخش (۱،۲،۵)، در «پیکره وابستگی»، پسوند «-ی» دارای ۴۹۷ تک‌بسامدی و مجموع تک‌بسامدی‌های همه پسوندهای این پیکره، ۱۰۵۵ بوده‌است. بر مبنای

فرمول‌های شماره [۸] و [۹]، در صفحه‌های پیشین، زایایی جهانی این پسوند، این گونه سنجیده می‌شود:

$$11. \text{ زایایی جهانی} = \frac{497}{1155} = 0/471$$

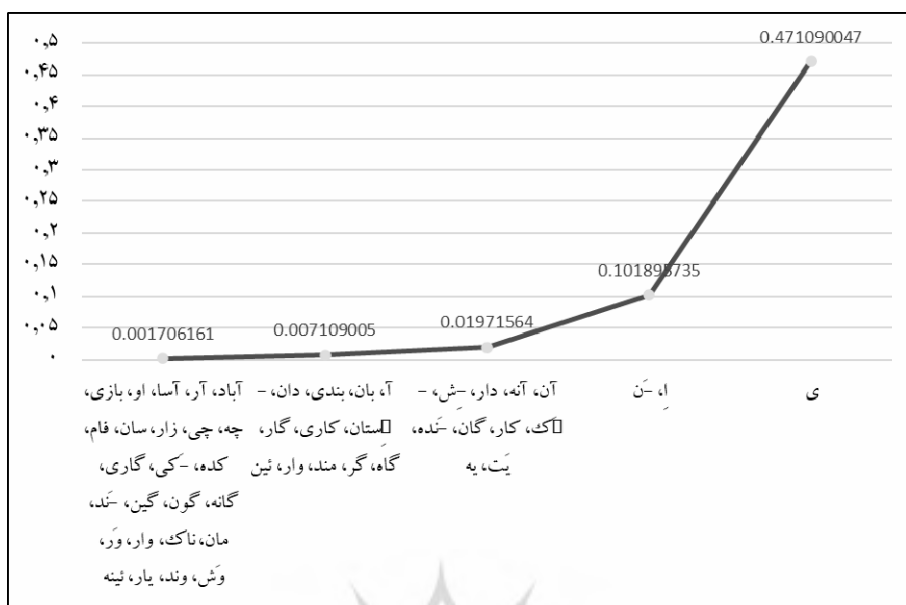
پس از سنجش زایایی جهانی پسوندها مطابق فرمول و اجرای خوشه‌بندی سلسله‌مراتبی، پسوندها بر اساس میانگین زایایی جهانی در ۵ خوشه قرار گرفتند. نمودار خطی شماره (۶-۵)، نشان می‌دهد که پسوند «ی» دارای بالاترین زایایی جهانی با احتمال زایایی (۰/۴۷۱)، است که به‌تنهایی و با اختلاف بسیار زیادی نسبت به دیگر پسوندها در سمت راست نمودار با بیشترین زایایی جهانی قرار گرفته‌است. پس از آن دو پسوند «ن» و «ا»، با احتمال زایایی (۰/۱۰)، در خوشه‌ی با زایایی بالا قرار گرفته‌اند. ده پسوند «آن» - «آنه» - «دار» - «ش» - «ک» - «کار» - «گان» - «نده» - «یت» و «یه»، در خوشه‌ی زایایی متوسط با احتمال زایایی (۰،۰۱۹)، قرار گرفته‌اند. همان گونه که مشاهده می‌شود، ۱۲ پسوند «آ» - «بان» - «بندی» - «دان» - «ستان» - «کاری» - «گار» - «گاه» - «گر» - «مند» - «وار» و «ئین»، با احتمال زایایی (۰،۰۰۷) در خوشه‌ی با زایایی پایین قرار دارند. ۲۵ پسوند دیگر با احتمال زایایی ناچیز (۰،۰۰۱۷)، در خوشه‌ی غیر زایا واقع شده‌اند. یافته‌های به‌دست آمده در باره‌ی پنج پسوند همسو با دستاوردهای عباسی (Abbassi, 2006)، است. این پنج پسوند شامل «ی» در خوشه‌ی کاملاً زایا، پسوند «مند» در خوشه‌ی با زایایی پایین و پسوندهای «ار» - «مان» - «ئینه» در خوشه‌ی غیر زایا می‌شود.^۱ در این سنجش نیز نکته‌های مهمی وجود دارد:

یکم- در این روش، برای سنجش زایایی هر پسوند، همه‌ی موارد تک‌بسامدی‌ها در مخرج کسر قرار می‌گیرد. بنابراین رقم به‌دست آمده در سنجش زایایی صرفی برای همه‌ی پسوندهای یک پیکره با هر حجمی، کاملاً ثابت است.

دوم- در فرمول زایایی جهانی برخلاف فرمول زایایی دقیق، عدد به‌دست آمده تعیین می‌کند که هر پسوند خاص، تا چه اندازه در افزایش واژه‌های پیکره‌ی مورد مطالعه، نقش داشته‌است.

سوم- حاصل جمع مقادیر زایایی برابر یک است (یعنی حاصل جمع پنج مقدار در پنج خوشه در نمودار). به دیگر سخن، مفهوم احتمال زایایی، سنجیده شده‌است. پس مقادیر زایایی هر پسوند در پیوستار زایایی صرفی، احتمالی بین صفر و یک را به خود اختصاص می‌دهد. پس آنچه در این روش برای سنجش زایایی صرفی، استفاده شده‌است، احتمال زایایی صرفی کنونی را اندازه‌گیری کرده‌است. به بیان دیگر، این روش دارای پایایی (اعتماد) است.

^۱ همان گونه که در پیشینه این مقاله گفته شد در مطالعه زایایی صرفی عباسی (۲۰۰۶)، تعداد ۲۰ پسوند مورد بررسی قرار گرفته‌اند. خوشه‌بندی این ۲۰ پسوند برای مقایسه با نتایج این مطالعه، توسط نگارندگان این مقاله انجام شده است.



شکل ۶: نمودار خطی یا پیوستار احتمالی زایایی پسوندها بر مبنای زایایی جهانی

۳.۵. بحث

همان گونه که شرح داده شد، برای سنجش دقیق تر زایایی به واژه‌های تک‌بسامدی و نوواژه‌ها نیز نیاز داریم. بر مبنای پژوهش پلگ (Plag, 2003) واژه‌های دارای یک وند معین که تک‌بسامدی نیستند نسبت به تک‌بسامدی‌های همان وند، احتمال بیشتری دارد که در یک فرهنگ‌واژگان بسیار بزرگ فهرست شده باشند (Plag, 2003, p. 70). سنجش این موضوع به سادگی با جستجوی نوواژه‌های اشاره‌شده در «پیکره وابستگی» و یافت نشدنشان در فرهنگ لغت‌های زبان فارسی، این موضوع را تأیید می‌کند. پس هرچه تعداد تک‌بسامدی‌های یک وند ویژه بیشتر شود، تعداد نوواژه‌ها بیشتر می‌شود، از این رو، احتمال روبه‌رو شدن با یک واژه تازه ایجادشده، یعنی زایایی وند بیشتر است.

نکته مهم در این جا آن است که تعداد تک‌بسامدی‌های یک پسوند معین باید با تعداد نوواژه‌های آن همبسته باشد، به گونه‌ای که بتوان تعداد تک‌بسامدی‌ها را به عنوان شاخص زایایی مشاهده کرد. به دیگر سخن، ادعا نمی‌شود که واژه‌های تک‌بسامدی، نوواژه‌اند. در این پیکره، با توجه به اینکه بر مبنای آزمون شاپیروویلیک^۱، دو متغیر تک‌بسامدی‌ها و نوواژه‌ها، از پراکندگی نرمال پیروی نمی‌کنند، در اجرای تحلیل همبستگی^۲ از ضریب همبستگی کندال^۳ بهره گرفته شد. مقدار ضریب همبستگی کندال

¹ Shapiro-Wilk Test

² correlation analysis

³ Kendall correlation coefficient

برای رابطه‌ی میان بسامد تک‌بسامدی‌ها و بسامد نوآژه‌ها، معادل (۰/۴۷۱) به دست آمد. این مقدار مثبت، نشان می‌دهد که رابطه‌ی مستقیمی بین این دو متغیر وجود دارد. برای معناداری این ضریب مقدار (p-value)، برابر (3.058×10^{-5}) ، سنجش شد که کمتر از (۰/۰۵) است. این مقدار، نشان می‌دهد که میان بسامد تک‌بسامدی‌ها و بسامد نوآژه‌ها، همبستگی معناداری وجود دارد.

موضوع مهم دیگر آنکه، واژه‌های تک‌بسامدی با توجه به یک پیکره معین تعریف می‌شود، بنابراین امکان دارد به جای یک مشتق تازه تولیدشده، یک واژه کیمیا ب زبان یا برخی ابداعات کاربران باشد. در این جا اندازه پیکره نیز نقش مهمی در تعیین ماهیت تک‌بسامدی‌ها بر عهده دارد. وقتی پیکره‌ای مانند «پیکره وابستگی» کوچک باشد، بیشتر تک‌بسامدی‌ها، واژه‌های شناخته‌شده زبان هستند (مانند سخنور، زندانبان، پرستار در این پیکره). با این وجود، با افزایش اندازه پیکره، تعداد نوآژه‌ها در بین واژه‌های تک‌بسامدی افزایش می‌یابد و دقیقاً در بین واژه‌های تک‌بسامدی بیشترین تعداد نوآژه‌ها نمایان می‌شوند (پیامک، طالبان، سامانه).

با توجه به هر چهار سنجش کمی زایایی صرفی (یعنی سنجش با بسامدهای نوع‌واژه، موردواژه، تک‌بسامدی‌ها و نوآژه‌ها)، می‌توان تفاوت‌های بسیاری بین پسوندها مشاهده کرد. با نگاهی به ستون عمودی نمودارها، می‌توان گفت که برخی از پسوندها دارای بسامد موردواژه بالایی هستند (مانند -ی، -ا، -ن)، یعنی برخی از واژه‌ها دارای این پسوندها، بسیار مورد استفاده قرار می‌گیرند. انواع دیگر مشتقات اغلب مورد استفاده قرار نمی‌گیرند و بسامد موردواژه نسبتاً کمی دارند به ویژه پسوندهای «-آسا، زار، قام، کده، گانه، گون، گین، مان، ئینه، ور، وش و موارد مشابه». برای پرهیز از سنجش‌هایی که چیزی غیر از زایایی کنونی را اندازه‌گیری می‌کند؛ برای سنجش پایایی (دارای اعتماد) زایایی صرفی پسوندهای اشتقاقی غیر فعلی زبان فارسی، از روش‌های برآورد احتمال زایایی دقیق و زایایی جهانی، بهره گرفتیم.

بحث پیرامون این سنجش را با یک مثال دنبال می‌کنیم. باین و لیبر (Baayen & Lieber, 1991) بر مبنای پیکره ۱۸ میلیون واژه‌ای، چنین بیان کرده‌اند که بسامد نوع‌واژه مقوله (-ness) با بسامد ۴۹۷ نمی‌تواند نشان دهد که از مقوله (-ity) با بسامد ۴۰۵، زیاتر است. ولی با مقایسه زایایی جهانی یعنی عدد ۰,۰۴۴ برای مقوله (-ness) و عدد ۰,۰۰۷ برای مقوله (-ity)، تفاوت زایایی این دو را می‌توان نشان داد که مطابقت بیشتری با برآوردهای شمی دارد (Baayen & Lieber, 1991). با این قیاس می‌توان گفت که پسوند «-گاه» با بسامد نوع‌واژه (۲۲)، زیاتر از پسوندهای «-بان و -دان/دانی»، با بسامد (۸)، است. ولی بر مبنای زایایی جهانی هر سه پسوند، در خوشه با زایایی پایین با احتمال زایایی ۰,۰۰۷، قرار می‌گیرند. همین موضوع درباره توجه به اعداد خام بسامد موردواژه نیز صدق می‌کند، به گونه‌ای که پسوند «-نده» با بسامد

موردواژه (۴۷)، زیایتر از پسوند «-کار»، با بسامد موردواژه (۲۲)، به نظر می‌رسد ولی هر دو پسوند، پس از اجرای محاسبات (الگوریتم) k -میانگین، در خوشه زایایی متوسط با احتمال زایایی (۰,۰۱)، واقع می‌شوند. دلیل مهم آن است که در روش بسامدی، فقط به شمارش خام هر پسوند به کاررفته در ساخت واژه، پرداخته می‌شود. در حالی که روش زایایی جهانی، بر آورد احتمال درست زایایی هر پسوند برای تولید واژه جدید، به دست داده می‌شود.

افزون بر این، در سنجش با روش زایایی دقیق، بسامد تک‌بسامدی‌های هر پسوند به بسامد موردواژه‌های خودش سنجیده می‌شود. اگر اعداد به دست آمده از این روش را با هم جمع ببندیم، احتمال زایایی به دست نمی‌آید زیرا حاصل جمع بیشتر از عدد یک است. ولی همان‌گونه که مشاهده شد در زایایی جهانی، سهم عددی تک‌بسامدی‌های هر پسوند به مجموع سهم عددی تک‌بسامدی‌های همه پسوندها، سنجیده می‌شود. عدد حاصل هر پسوند نیز بین اعداد صفر و یک است، که در مجموع عدد یک می‌شود. این نشان‌دهنده احتمال زایایی هر پسوند در تولید واژه‌های جدید در زبان است. بنابراین با تحلیل روش‌های سنجش کمی و دقت در آنچه که در عمل در زایایی صرفی، سنجش می‌شود؛ به پایایی (اعتماد)، قابل قبول در پژوهش‌های کمی زایایی صرفی نزدیک می‌شویم.

بهره‌گیری از روش‌های تعیین بسامد نوع‌واژه و موردواژه، به دلیل متفاوت بودن مفهوم‌شان با زایایی صرفی، پیشنهاد نمی‌شوند؛ هر چند برای سنجش زایایی گذشته، یافته‌هایی به دست می‌دهند. به موجب گسترش مطلوب رشته زبان‌شناسی پیکره‌ای، استفاده از روش‌های سنجش کمی زایایی (با برآورد احتمال)، با بهره‌گیری از پیکره‌های تولیدشده به وسیله مراکز مطالعات پژوهشی، روش‌های پایاتر و رواتری هستند. ولی برای افزایش روایی در اندازه‌گیری‌ها، مقایسه یافته‌های به دست آمده از سنجش زایایی در پیکره‌های مختلف، ضروری است.

۶. نتیجه‌گیری

هرگاه از زایایی صرفی سخن می‌گوییم به آن معناست که یک فرایند صرفی (در این مطالعه پسوندافزایی)، تا چه اندازه مورد استفاده کاربران زبان قرار می‌گیرد. واژه‌های ممکن مطابق قاعده‌های صرفی زبان، امکان ساخته شدن دارند و واژه‌های موجود در فرهنگ‌های لغت و پیکره‌های زبانی دیده می‌شوند. زایایی پیوستاری احتمالی است که میزان آن بین صفر و یک، متغیر است. احتمال زایایی هر پسوند با میزان کاربرد واژه‌های ساخته شده با هر پسوند، ارتباط دارد ولی با آن برابر نیست. زایایی گذشته باید از زایایی کنونی هر پسوند متمایز گردد. در این مقاله، گروه‌بندی نوینی از همه سنجش‌های کمی زایایی صرفی بر مبنای مفاهیم آمار و احتمالات ارائه

گردید که در چهار مدل کلی تعیین نسبت، تعیین بسامد، تعیین نوواژه‌ها و تعیین واژه‌های تک‌بسامدی (برآورد احتمال)، گنجانده شد. زایایی کنونی از طریق زایایی دقیق و زایایی جهانی باور (Bauer, 2004) و باین (Baayen, 2009)، مورد بررسی و تحلیل قرار گرفت. این سنجش کمی، نقطه نظرات باین و لیبر (Baayen & Lieber, 1991)، باین (Baayen, 1993) و باور (Bauer, 2001)، را در خود دارد. در این پژوهش، پیوستار احتمالی زایایی پسوندهای اشتقاقی غیر فعلی زبان فارسی در «پیکره وابستگی» به دست داده شد. در این پیکره، احتمال زایایی پسوندهای «ی» و «ن و ـ» و «ـا»، به ترتیب (۰،۴۷)، (۰،۱) و (۰،۱) است که دارای بیشترین مقدارند. ۲۵ پسوند دارای کمترین احتمال زایایی یعنی (۰،۰۰۱) هستند. ۲۲ پسوند دیگر با احتمالی بین (۰،۰۱) و (۰،۰۰۷)، در میانه پیوستار زایایی واقع شدند.

فهرست منابع

- دهخدا، علی اکبر (بی تا). لغت‌نامه دهخدا. تهران: سیروس.
- شقایق، ویدا (۱۳۸۶). مبانی صرف. تهران: سمت.
- شقایق، ویدا (۱۳۹۴). فرهنگ توصیفی صرف. تهران: انتشارات علمی.
- شکری، سارا و آزیتا عباسی (۱۳۹۷). «بررسی محدودیت‌های ناظر بر ترتیب پسوندهای اشتقاقی در زبان فارسی بر مبنای رویکرد محدودیت‌های پردازشی». پژوهش‌های زبان‌شناسی. سال ۱۰. شماره ۱. صص ۹۹-۱۱۴.
- صحرايي، رضامراد، امیرحسین مجیری فروشانی و مروارید طالبی (۱۳۹۸). «واژه‌های پایه زبان فارسی مبتنی بر متون مطبوعاتی». زبان پژوهی. سال ۱۱. شماره ۳۳. صص ۳۵۳-۳۷۸.
- صفری، علی و لیلیا نیک‌نسب (۱۳۹۹). «تحلیل واژه‌سازی زبان فارسی براساس نظریه صرف ساخت‌بنیاد: شواهدی از کلمات مرکب». زبان پژوهی. سال ۱۲. شماره ۳۶. صص ۱۶۵-۱۸۶.
- عباسی، آزیتا (۱۳۸۴). زایایی در فرایند اشتقاق زبان فارسی. رساله دکتری. دانشگاه تهران.
- عباسی، آزیتا (۱۳۸۴). «محدودیت‌های صرفی و نحوی در زایایی فرایند اشتقاق در زبان فارسی». زبان و زبان‌شناسی. دوره ۲. شماره ۴. صص ۳۹-۵۶.

References

- Abbassi, A. (2006). Morphological and syntactic constraints of productivity in Persian derivation. *Language and Linguistics*, 2 (4), (pp. 39-56). (in Persian)
- Abbassi, A. (2006). Productivity in Persian derivation (Doctoral dissertation). Tehran University). (in Persian)
- Anshen, F., & Aronoff, M. (1998). *Morphology in real time*. Dordrecht: Kluwer Academic Publishers.
- Aronoff, Mark (1976). *Word formation in generative grammar*. Cambridge,

- Massachusetts: MIT Press.
- Aronoff, M. (1980). The relevance of productivity in a synchronic description of word formation. In J. Fisiak (Ed.), *Historical morphology* (pp. 71-82). Berlin, New York: De Gruyter Mouton.
<https://doi.org/10.1515/9783110823127.71>.
- Aronoff, M. (1983). Potential words, actual words, productivity and frequency. *Proc. XIII In journal of Poznan studies in contemporary linguistics* (pp. 163-171). <https://doi.org/10.2478/v10010-010-0010-x>.
- Baayen, H. (1991). *Quantitative aspects of morphological productivity*. In G. Booij (Eds.), *Yearbook of morphology* (pp.109-149). Dordrecht: Kluwer.
- Baayen, H. (1992). On frequency, transparency, and productivity. In G. Booij & J. Van Marle (Eds.), *Yearbook of morphology* (pp.181-208). Dordrecht: Kluwer.
- Baayen, H. (1993). On frequency, transparency, and productivity. In G. Booij & J. Van Marle (Eds.), *Yearbook of morphology* (pp.181-208). Dordrecht: Kluwer.
- Baayen, H. (1989). A corpus-study of *morphological productivity*. Statistical analysis and psycholinguistic interpretation. Vrije Universiteit dissertation.
- Baayen, R. H. (2009). *Corpus linguistics in morphology: Morphological productivity*. In A. Lüdeling & M. Kytö (Eds.), *Corpus linguistics: An international handbook* (pp. 900-919). Berlin: Mouton De Gruyter.
- Baayen, R. H. & Lieber, R. (1991). Productivity and English word-formation: A corpus-based study. *Journal of Linguistics*, 29 (5), (pp. 801-843). <https://doi.org/10.1515/ling.1991.29.5.801>.
- Bauer, L. (2001), *Morphological Productivity*, Cambridge: Cambridge University Press.
- Bauer, L. (2004). *Productivity: Theories*. In P. Štekauer. & R. Lieber (Eds.), *Handbook of word-formation* (315-334). Dordrecht: Springer.
- Bolozky, Sh. (1999). *Measuring productivity in word formation*. Leiden: Brill.
- Booij, G. (2005). *The grammar of words: Oxford textbooks in linguistics*. Oxford & New York: Oxford university press.
- Booij, G. (2018). *The Construction of Words. Introduction and overview*. In B. Dancygier (Ed.), *The Cambridge handbook of cognitive linguistics* (pp. 229-245). Cambridge: Cambridge University Press.
- Dehkhoda, A. (no date). *Dehkhoda dictionary*. Tehran: Sirus. (in Persian)
- Evert, S., & Lüdeling, A. (2001). *Measuring morphological productivity: Is automatic preprocessing sufficient?* In P. Rayson, A. Wilson, T. McEnery, A. Hardie, & S. Khoja (Eds.), *Proceedings of the Corpus Linguistics 2001 Conference: 167-175*. Lancaster: Peter Lang.
- Haspelmath, M. & Sims, A. D. (2010). *Understanding morphology*. London: Hodder Education.
- Hay, J. (2003). *Causes and consequences of word structure*. New York: Routledge.
<https://doi.org/10.4324/9780203495131>.
- Hay, J. B. (2001). Lexical frequency in morphology: Is everything relative? *Journal of Linguistics* 39 (6), (pp. 1041-1070).
<https://doi.org/10.1515/Ling.2001.041>.
- Johnson, R. A., & Dean W. W. (2002). *Applied multivariate statistical analysis*. Englewood Cliffs, New Jersey: Prentice Hall.
- Knowles, E., & Elliott, J. (1997). *The Oxford dictionary of new words*. New York:

- Oxford University Press.
- Kreyer, R. (2015). *The Marburg corpus of intermediate learner English (MILE)*. Amsterdam: John Benjamins.
- Kruisinga, E. (1932). *A handbook of present-day English*, 5. Groningen: Noordhoff.
- McEnery, T. & Hardie, A. (2012). *Corpus linguistics: Method, theory and practice*. Cambridge & New York: Cambridge University Press.
- McEnery, T. & Wilson, A. (2001). *Corpus linguistics: An introduction*. Edinburgh: Edinburgh University Press.
- Mirzaei, A. & Safari P. (2018). Persian discourse treebank and coreference corpus. In N. Calzolari, K. Choukri, C. Cieri, T. Declerck, S. Goggi, K. Hasida, H. Isahara, B. Maegaard, J. Mariani, H. Mazo, A. Moreno, J. Odijk, S. Piperidis, & T. Tokunaga (Eds.), *Proceedings of Eleventh International Conference on Language Resources and Evaluation* (pp. 4049-4055). Paris: European Language Resources Association
- Plag, I. (1999). *Morphological productivity: Structural constraints in English derivation*. Berlin: Mouton de Gruyter.
- Plag, I. (2003). *Word-formation in English*. Cambridge: Cambridge University Press.
- Plag, I. (2006). *Productivity*. In K. Brown (Eds.). *Encyclopedia of language and linguistics*. Amsterdam: Elsevier.
- Pustyl'nikov, O. & Schneider-Wiejowski, K. (2009). Measuring morphological productivity. In R. Köhler (Ed.). *Issues in quantitative linguistics* (pp. 106-125). Lüdenscheid: RAM-Verlag. <https://doi.org/urn:nbn:de:0070-pub-24868658>.
- Pustyl'nikov, O. (2009). Modeling learning of derivation morphology in a multi-agent simulation. *Conference of Proceedings of IEEE Africon* (pp. 1-6). <https://doi.org/10.1109/AFRCON.2009.5308103>.
- Rasooli, M. S., Kouhestani, M., & Moloodi, A. (2013). Development of a Persian syntactic dependency treebank. In Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL HLT), Atlanta, USA.
- Safari, A., & Niknasab, L. (2020). A constructional account of word formation in Persian: Evidence from Compounding. *Journal of Language Research*, 12 (36), (pp.165-186) (in Persian) <https://doi.org/10.22051/jlr.2019.24205.1645>.
- Sahraei, R. M., Mojiri-Foroushani, A. H. and Talebi, M. (۲۰۱۹). Persian Basic Words Based on Texts by Press. Tehran: *Journal of Language Research*, 11 (33), (pp. 353-378). (in Persian) <https://doi.org/10.22051/jlr.2019.19564.1520>.
- Schultink, H. (1961). *Produktiviteit als morfologisch fenomeen*. Forum der Letteren 2, 110-125.
- Shaghghi, Vida (2015). *Descriptive dictionary of morphology*. Tehran: Entesharat Elmi (in Persian)
- Sharifi, Sh. & Erfanian-ghonsouli, L. (2011). The study of productive and non-productive derivational Persian affixes. In F. Ghatre & S. Modarress Khayabani (Eds.), *The Journal of Third Conference of Morphology*. 123-128. Ferdowsi University Press. (in Persian)
- Van Marle, J. (1992). The relationship between morphological productivity and

- frequency: A comment on Baayen's performance-oriented conception of morphological productivity. In G. E. Booij, and J.V. Marle (Eds.), *Yearbook of morphology 199*, (pp. 109-149). Dordrecht: Kluwer.
https://doi.org/10.1007/978-94-011-2516-1_9.
- Willmanns, W. (1899). *Deutsche grammatik. Abteilung. wortbildung*. Strassburg: Verlag von Karl J. Trübner.



© 2020 Alzahra University, Tehran, Iran. This article is an open-access article distributed under the terms and conditions of the Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC-ND 4.0 license) (<https://creativecommons.org/licenses/by-nc-nd/4.0/>).

