

بررسی آماری تحول زایایی وندهای اشتقاقی در نظم فارسی*

وحید مواجی^۱ و محرم اسلامی^۲

^۱ دانشگاه صنعتی شریف

^۲ دانشگاه زنجان

۲۴ بهمن ۱۳۹۱

چکیده

زایایی یکی از ویژگی‌های دستگاه صرف است که مشخص‌ترین خصوصیتش نسبی بودن آن است. یعنی یک فرایند واژه‌سازی در طول زمان می‌تواند زایایی خود را از دست بدهد یا اینکه زایاتر شود. از طرفی بین فرایندهای مختلف واژه‌سازی نیز زایایی نسبی است. زایایی ویژگی طیفی دارد که مقادیر آن می‌تواند در فرایندهای واژه‌سازی بین صفر (کاملاً سترون) و ۱ (بسیار زایا) در نوسان باشد. با دورویکرد می‌توان به مسأله زایایی نگریست: رویکرد هم‌زمانی و رویکرد درزمانی. در این مقاله از هر دو دیدگاه به مسأله نگریسته شده و با ارائه فرمول‌های ریاضی و محاسبات آماری روی پیکره‌ای از شعر فارسی (پیکره گنجور) با بیش از یک میلیون مصراع، زایایی فرایند اشتقاق پسوندی در نظم فارسی بررسی شده است. این پیکره اشعار ۵۲ شاعر فارسی‌زبان از قرن چهارم تا قرن چهاردهم هجری را شامل می‌شود. البته ممکن است تحلیل آماری منابع منظوم و نتایج به دست آمده از آن نماینده صددرصد زبان رایج مردم در دوره‌ای از تاریخ یک زبان نباشد، ولی حداقل بازتاب‌دهنده بخشی از ویژگی‌های ساخت‌واژی در گونه ادبی دوره‌های مختلف زبان فارسی خواهد بود. در این مقاله معیارهای مختلفی که برای سنجش میزان زایایی ارائه شده، مورد بررسی قرار گرفته است و با استفاده از روش تعداد تک‌وقوع‌ها و نیز تجزیه خودکار کلمات به کمک یک تحلیل‌گر صرفی (پارس مورف)، میزان زایایی پسوندهای اشتقاقی از نظر درزمانی و هم‌زمانی با شاخص‌های عددی و آماری نشان داده می‌شود

کلیدواژه‌ها: واژه‌سازی، اشتقاق، زایایی درزمانی، زایایی هم‌زمانی، نظم فارسی

۱ مقدمه

زایایی در صرف یکی از مسائل محوری مطالعات واژه‌سازی است و دیدگاه‌های مختلفی در بحث زایایی مطرح است. به اعتقاد [Bauer \(2001\)](#):

الف) وندها زایا هستند؛ یعنی زایایی، ویژگی یک وند برای ساختن کلمات پیچیده جدید است؛ [Plag \(2003\)](#) و [Fleischer \(1975\)](#).

ب) فرایندهای صرفی زایا هستند یعنی ویژگی یک فرایند صرفی برای تشکیل سازه‌های جدید طبق اصول نظام‌مند، زایایی نام دارد [Anderson \(1982\)](#)؛ [Uhlenbeck \(1978\)](#).

علی‌رغم این که در بین مقالات بر سر این که چه چیزی زایا است، اختلاف وجود دارد، مطالعه کمی زایایی بیشتر حول محور زایایی وند متمرکز شده است. معیارهای کمی مختلفی برای سنجش زایایی ارائه شده است که برخی

* هشتمین همایش زبان‌شناسی ایران، دانشگاه علامه طباطبائی، بهمن ۱۳۹۱

زایایی گذشته و برخی زایایی بالقوه را ارزیابی می‌کنند. باید توجه داشت که زایایی یک پدیده در زمانی است، بدین معنی که یک وند خاص ممکن است در زمان گذشته بسیار زایا بوده باشد مانند پسوند /-ak/ در زبان فارسی، ولی امروزه به ندرت کلمات جدیدی با این پسوند ساخته می‌شوند و لذا این پسوند غیرزایا می‌باشد (طباطبائی، ۱۳۸۲).

در این پژوهش با بررسی‌های آماری روی پیکره‌ای نسبتاً غنی از نظم فارسی^۱ با بیش از یک میلیون مصراع شعر، تحول ساخت‌واژی زبان فارسی را در گونه‌ی زبانی نظم و شعر مطالعه می‌کنیم. زبان شعر شاید نماینده‌ی زبان رایج مردم در دوره‌ای از تاریخ نباشد، ولی تحلیل آماری زایایی وندهای اشتقاقی در نظم و نتایج به‌دست‌آمده می‌تواند حداقل نشان‌دهنده‌ی بخشی از ویژگی‌های صرفی زبان فارسی در گونه‌ی ادبی به عنوان تنها آثار باقیمانده از زبان فارسی باشد. شایان ذکر است که آثار منثور فارسی نیز نماینده‌ی تمام‌کمال زبان فارسی رایج در دوره‌ای از تاریخ ایران زمین نمی‌تواند باشد.

یکی از مواردی که می‌توان روی این پیکره بررسی کرد، بحث زایایی فرایندهای واژه‌سازی در زبان فارسی است. این بررسی را می‌توان از دو منظر مورد توجه قرارداد: از دید هم‌زمانی یعنی زایایی نسبی فرایندهای واژه‌سازی در یک بازه زمانی از تاریخ و از دید در زمانی یعنی تحول زایایی یک فرایند واژه‌سازی در طول زمان. فرایندهای واژه‌سازی در فارسی شامل ترکیب و اشتقاق می‌باشد که در این مقاله، فرایند اشتقاق پسوندی در زبان فارسی مورد بررسی قرار می‌گیرد.

بحث اینکه چه معیاری برای سنجش زایایی مناسب می‌باشد بحث گسترده‌ای است. در این مقاله معیارهای مختلفی که برای سنجش میزان زایایی ارائه شده است مورد بررسی قرار گرفته و با استفاده از روش تعداد تک‌وقوع‌ها^۲ معیاری عددی و آماری برای سنجش میزان زایایی وندهای اشتقاقی ارائه می‌شود و نتایج بدست آمده از دید در زمانی و هم‌زمانی ارائه می‌گردد.

۲ معیارهای زایایی صرفی

در ادامه به معرفی مدل‌های ارزیابی زایایی می‌پردازیم.

۱.۲ مدل آرونوف

(Aronoff (1976، به نقل از بائر (Bauer (2001، نسبت کلمات بالفعل ساخته شده در چارچوب یک قاعده واژه‌سازی (WFR^۳) را به کلمات بالقوه قابل تولید از آن قاعده، طبق فرمول زیر محاسبه می‌کند:

$$I = v/s \quad (1)$$

که در آن I شاخص زایایی است؛ v تعداد نوع‌کلمه‌های واقعی است، و s تعداد نوع‌کلمه‌هایی است که WFR می‌تواند تولید کند.

۲.۲ مدل‌های بسامدی

مدل‌های بسامدی روی این فرض استوار هستند که بسامد به طور مستقیم یا غیرمستقیم با زایایی مرتبط است. منظور از «بسامد» تعداد دفعات وقوع یک کلمه در یک پیکره است. در خصوص موضوع بسامد سه نظر مختلف مطرح است که در آنها به ترتیب بسامد نوع^۴، بسامد نمونه^۵ و بسامد نسبی مد نظر می‌باشد. به عنوان مثال اگر

^۱پیکره گنجور

^۲Hapaxes

^۳Word-Formation Rule

^۴Type

^۵Token

۱۰۰ کلمه با پسوند /-ande/ در یک پیکره زبانی فارسی باشد برای مثال در «گوینده، جوینده، کشنده»، می‌گوییم بسامد نوع برابر ۱۰۰ است. بسامد نمونه در همین پیکره زبانی ممکن است بیش از ۱۰۰ باشد، چراکه ممکن است هر کدام از آن ۱۰۰ کلمه بیش از یک بار تکرار شده باشد.

۱.۲.۲ بسامد نوع

یک دیدگاه این است که وندی زایاتر است که تعداد کلمات زیادی با استفاده از آن ساخته شده باشد. بنابراین یکی از پرکاربردترین معیارها در مقالات، شمارش ساده تعداد نوع کلمه‌های بالفعل ساخته شده با آن‌وند در یک دوره زمانی خاص است. البته این معیار به همان اندازه هم مورد مخالفت قرار گرفته است به این دلیل که یک‌وند ممکن است در گذشته کلمات زیادی را می‌ساخته ولی امروزه گویشوران زبان به ندرت کلمه جدیدی را از رهگذر آن می‌سازند. مثال چنین وندی، «-اک» در فارسی یا «-ment» در انگلیسی می‌باشد که در قرون گذشته کلمات جدید زیادی با آن ساخته می‌شده و بسیاری از آن کلمات هنوز به کار می‌روند ولی در حال حاضر کلمه جدیدی با آنها ساخته نمی‌شود و غیرزایا به شمار می‌روند (Plag, 2003).

۲.۲.۲ بسامد نمونه

از آنجا که بسامد نوع دارای مشکلاتی در ارزیابی زایایی است، یک راهکار دیگر برای بررسی زایایی، محاسبه بسامد نمونه است. (Baayen, 1993) رابطه بین بسامد و زایایی را مورد بررسی قرار داده است. طبق این نظر، یک فرایند صرفی زایا متناظر تعداد زیادی کلمه کم بسامد و تعداد اندکی کلمه پر بسامد است، در حالی که فرایندهای غیرزایا متناظر تعداد زیادی کلمه پر بسامد و تعداد اندکی کلمه کم بسامد می‌باشند. منطق پشت این قضیه این است که کلمات مرکب پر بسامد به صورت یک کلمه واحد در واژگان ذهنی ذخیره می‌شوند و کلمات مرکب کم بسامد به صورت سازه‌های تشکیل دهنده آنها در واژگان ذهنی ذخیره می‌شوند. دلیل این که یک کلمه مرکب تازه ساخته شده (مثلاً dis-represent) برای گویشورانی که قبلاً هیچ وقت آن را نشنیده‌اند قابل فهم است، این است که آنها این کلمه را به سازه‌های تشکیل دهنده‌اش تجزیه می‌کنند، معنی تکواژهای سازنده را استنباط کرده و سپس معنی کل کلمه را ادراک می‌کنند.

در حقیقت باید بین بسامد نوع با بسامد نمونه تمایز قائل می‌شویم (Booij, 2007). بسامد نمونه شامل همه کلمات پیچیده حاصل یک قاعده واژه‌سازی و نیز شامل کل وقوع‌های یک کلمه با تمامی شکل‌هایش می‌باشد. همانطور که پیشتر گفتیم، مثلاً اگر ۱۰۰ صفت مختلف داشته باشیم که به «-وار» ختم شده باشند، بسامد نوع کلمه این پسوند برابر ۱۰۰ است ولی بسامد نمونه آن می‌تواند بسیار بیشتر از این باشد، چون هر یک از این صفت‌ها خود می‌توانند به دفعات در متن مورد مطالعه تکرار شوند.

بسامد نوع کلمه را در یک پیکره نسبتاً بزرگ باید محاسبه کرد نه لغتنامه یا دایره‌المعارف. زایایی صرفی در بهترین حالت، خود را در ساختن کلماتی نشان می‌دهد که در لغتنامه وارد نشده‌اند، چون یک کلمه جدید وقتی ساخته می‌شود مدت زمانی باید صرف شود تا این کلمه تثبیت شود و بعد از مدتی وارد لغتنامه شود یا در مورد برخی کلمات شاید هیچ وقت وارد لغتنامه نشوند. بنابراین یک پیکره معیار مناسبتری است. با این حال، بسامد نوع کلمه حتی در یک پیکره بزرگ هم معیار خوبی برای زایایی بدست نمی‌دهد، زیرا زایایی به معنی قابلیت ساختن کلمات جدید است، کلماتی که هنوز وجود ندارند.

۳.۲.۲ بسامد نسبی

بسامد نسبی، بسامد پایه‌های مشتق شده و واژگانی را با هم در نظر می‌گیرد، با این فرض که یک فرایند زایاتر است اگر بسامد پایه‌های مشتق شده کمتر از بسامد پایه‌های واژگانی باشد و در غیر این صورت زایایی آن کمتر است. در این معیار، بسامد پایه‌های مشتق شده بر بسامد پایه‌های واژگانی تقسیم می‌شود. هر چه عدد حاصله بالاتر باشد، آن‌وند یا فرایند کمتر زایاست.

۳.۲ مدل احتمالاتی

مدل‌های احتمالاتی عمدتاً توسط باین (Baayen (1989); Baayen and Lieber (1991); Baayen (1992) ارائه شده‌اند. مجموعه این مدل‌ها به صورت آماری، احتمال وقوع یک کلمه جدید در قالب یک فرایند ساخت‌واژی را محاسبه می‌کنند. در مدل‌های احتمالاتی، محاسبه زیایی ناگزیر مستلزم در نظر گرفتن تک‌وقوع‌ها است. تک‌وقوع‌ها کلماتی هستند که تنها یک بار در پیکره رخ می‌دهند. طبق نظر باین، اگر بخواهیم زیایی را مطالعه کنیم باید تک‌وقوع‌ها را مورد بررسی قرار دهیم. بنابراین می‌توان میزان زیایی یک فرایند صرفی یا P را بدین صورت تعریف نمود (Baayen (1992):

$$P = n1/N \quad (2)$$

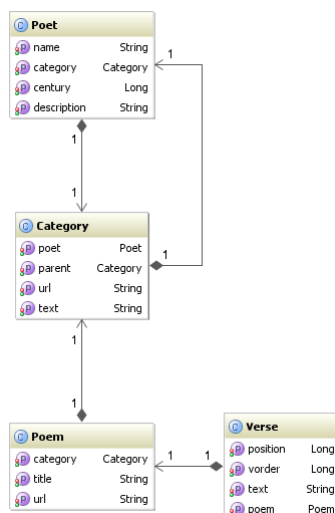
که $n1$ تعداد تک‌وقوع‌های نوع کلمه و N تعداد کل وقوع‌های آن نوع کلمه (نمونه) در پیکره می‌باشد. در این مقاله ما از همین روش برای محاسبه زیایی وندهای مختلف در نظم فارسی استفاده خواهیم کرد.

۳ منابع نظم فارسی

منابع مورد نظر از پروژه گنجور^۶ استخراج شده است. این پروژه به صورت متن‌آزاد و داوطلبانه انجام می‌شود و در زمان نگارش این مقاله شامل اشعار ۵۲ شاعر فارسی‌زبان از قرن چهارم تا قرن چهاردهم می‌باشد. این پروژه برای ذخیره دادگان خود از فرمت پایگاه داده Sqlite بهره می‌برد. جداول دادگان گنجور بدین شرح می‌باشند:

- جدول Poet: این جدول اطلاعات شاعران را نگهداری کرده و شامل ویژگی‌های نام شاعر، قرن شاعر مربوطه و شرح مختصری از زندگی شاعر است.
 - جدول Category: این جدول نقش مقوله بندی شعرا و شعرها را دارد.
 - جدول Poem: این جدول اطلاعات شعرها را نگهداری می‌کند (مثلاً یک غزل از حافظ یا یک حکایت از سعدی).
 - جدول Verse: این جدول که اصلی‌ترین جدول می‌باشد، اطلاعات تک مصراع‌ها را در خود دارد.
- ارتباط این جدول‌ها با هم و نمودار UML آنها در شکل ۱ آمده است.

^۶<http://ganjoor.net>



شکل ۱: ارتباط جداول موجود در پایگاه داده گنجور

جدول Poet شامل ۵۲ سطر، جدول Category شامل ۶۱۲ سطر، جدول Poem شامل ۳۸۷۴۲ سطر و جدول Verse شامل ۱۰۶۷۲۹۲ سطر می‌باشند که منبعی غنی برای تحلیل آماری نظم فارسی را در اختیار ما می‌گذارند.

۴ الگوریتم به‌کاررفته

برای بررسی وندها، از دادگان زایا یا FLexicon (اسلامی و دیگران، ۱۳۸۳) استفاده گردید. این دادگان دارای جدولی به نام Affix می‌باشد که شامل ۱۷۰ وند است و از این تعداد ۸۳ وند اشتقاقی داریم که در حقیقت پسوند می‌باشند. الگوریتم به‌کاررفته، تک‌تک وندهای اشتقاقی را با استفاده از تحلیلگر صرفی پارس‌مورف (مواجی و دیگران، ۱۳۹۰) روی اشعار شاعران مختلف پیدا کرده و با استفاده از فرمول ۲ زایایی آنها را محاسبه می‌کند. به علت تعدد وندها و شاعران، نتایج بسیار زیادی به‌دست آمد که برای داشتن یک دید در زمانی، این نتایج بر حسب قرن طبقه‌بندی گردید. مثلاً زایایی وند «-کده» در قرن چهارم الی چهاردهم به چه نحوی بوده است. بنابراین ۸۳ نمودار در زمانی بدست می‌آید که تحول زایایی این وندها را طی ۱۰ قرن به‌دست می‌دهند. از طرفی ۱۰ نمودار هم‌زمانی مربوط به هر قرن به‌دست آمد که زایایی ۸۳ وند در هر کدام از قرن‌ها را نشان می‌دهد. مسأله‌ای که وجود داشت این بود که توزیع کمی شعرها برای قرن‌های مختلف برابر و یکسان نبود. لذا با استفاده از تقریب مهندسی و مقیاس‌گذاری^۷ دادگان، سعی گردید تا حد امکان نمودارهایی هموار^۸ بدست آید.

۵ نتایج به‌دست‌آمده

همانطور که در بخش قبل گفته شد، دو نوع نمودار بدست آمد:

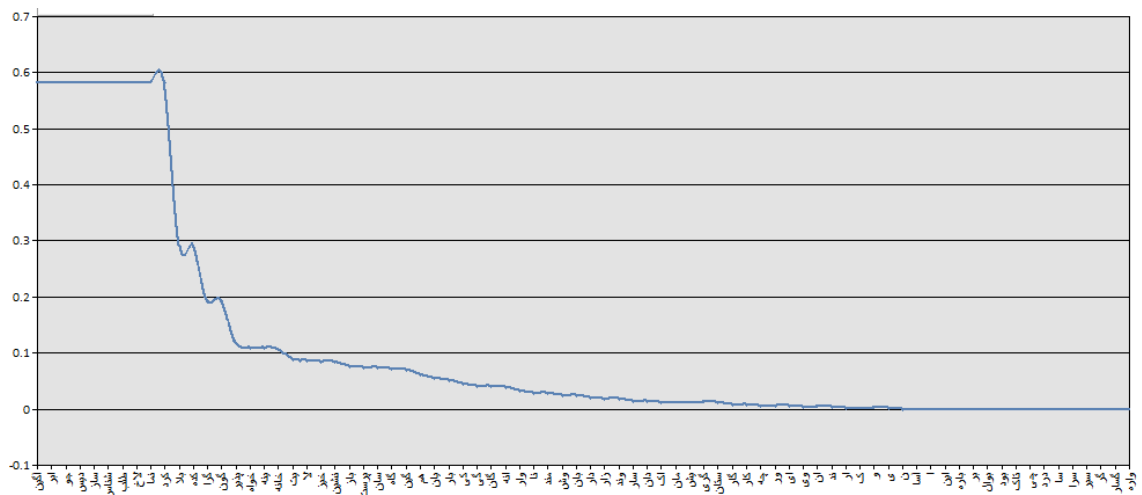
- ۱۰ نمودار هم‌زمانی برای هر یک از ۱۰ قرن که در آنها زایایی هر کدام از ۸۳ پسوند اشتقاقی آمده است.

⁷Scaling

⁸Smooth

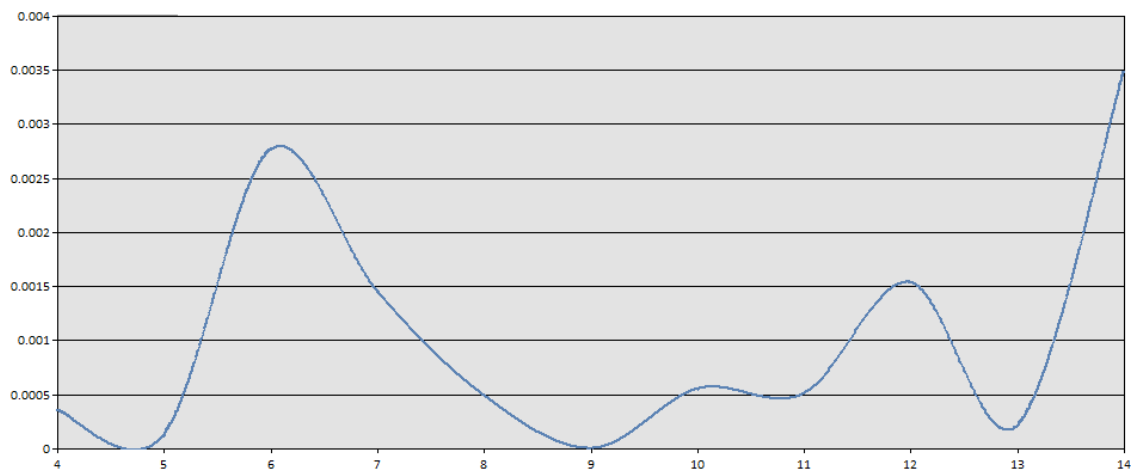
• ۸۳ نمودار در زمانی که در هر کدام از آنها تحول زیایی یک پسوند اشتقاقی خاص در طی ۱۰ قرن نشان داده شده است.

برای مثال نمودار هم‌زمانی زیایی وندها برای قرن هفتم در شکل ۲ آمده است. همانطور که از نمودار مشخص است وندهایی مانند: «-آگین»، «-ابر»، «-جو»، «-دیس» و «-ساز» دارای بیشترین زیایی و وندهایی مثل «-سیر»، «-گر»، «-گسار» و «-واره» دارای کمترین زیایی در قرن هفتم بوده‌اند. وندهایی هم مثل «-گرا»، «-گون»، «-پذیر» و «-خواه» دارای زیایی بینابینی بوده‌اند.

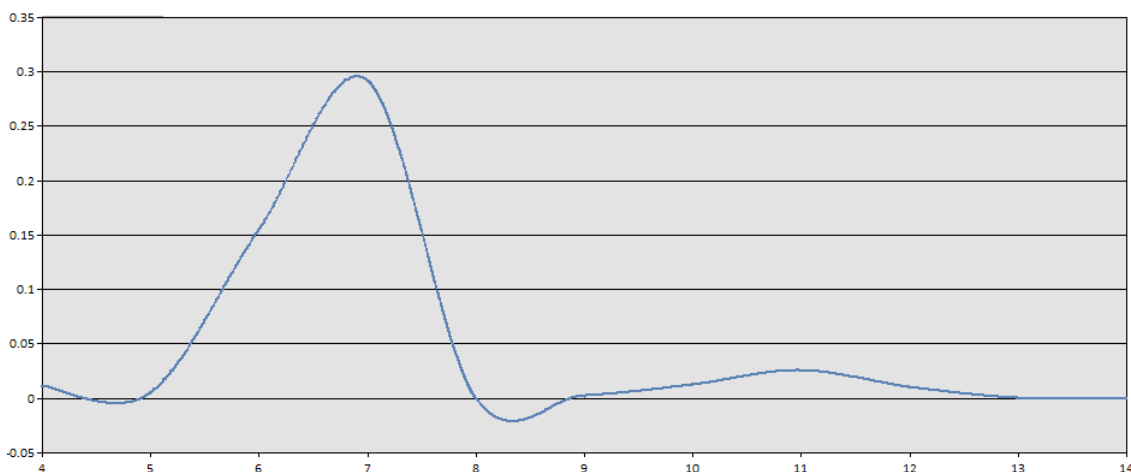


شکل ۲: نمودار هم‌زمانی زیایی وندها در قرن هفتم (محور افقی، وندها و محور عمودی میزان زیایی را نشان می‌دهد)

در شکل ۳، مثالی از نمودار در زمانی تحول زیایی وندها برای وندها «-ار» آورده شده است (محور افقی نشان‌دهنده قرن و محور عمودی میزان زیایی را نشان می‌دهد). همانطور که دیده می‌شود در بخش‌های قبلی این مقاله نیز ذکر گردید، زیایی وندها مقدار ثابتی نیست و با گذشت زمان ممکن است کمتر یا بیشتر گردد. مثلاً در قرن ششم، این وندها نسبتاً زیای بوده تا اینکه در قرن نهم به کمترین میزان زیایی خود رسیده و باز روند افزایش تا قرن دوازدهم داشته و در قرن سیزدهم زیایی آن کمتر شده و در قرن چهاردهم به بیشترین میزان زیایی خود رسیده است. البته باید توجه داشت که تمام این نتایج روی متون نظم فارسی گرفته شده است و اگر روی پیکره‌های دیگر مثل روزنامه‌ها، مجلات متون منشور و کتب بررسی شود، ممکن است نتایج متفاوتی بدست آید. نمودار مشابهی هم برای وندها «-کده» به منظور مقایسه در شکل ۴ آورده شده است.



شکل ۳: نمودار در زمانی زایایی وند «-ار» در طی قرن ۴ الی ۱۴



شکل ۴: نمودار در زمانی زایایی وند «-کده» در طی قرن ۴ الی ۱۴

۶ نتیجه گیری

در این تحقیق کوشیده‌ایم موضوع زایایی را در حوزه اشتقاق پسوندی به عنوان بخشی از ویژگی‌های ساخت‌واژی زبان فارسی مورد مطالعه قرار دهیم. در تحقیق حاضر، مسأله زایایی متناسب با طبیعت موضوع به دو صورت در زمانی و هم‌زمانی مورد بررسی قرار گرفته که در آن حیات ۸۳ پسوند فارسی در ۱۰ قرن جداگانه ترسیم شده است که به دلیل محدودیت در حجم مقاله تنها به ارائه چند نمونه اکتفا کرده‌ایم. بررسی پیکره عظیم گنجور که برخوردار از بیش از یک میلیون مصراع داده منظم است می‌تواند بستر مناسبی را برای مطالعه زایایی در اختیار ما بگذارد و به دلیل گستردگی داده‌ها نتایج نیز از اعتبار و روایی لازم برخوردار باشد.

از نتایج این تحقیق می‌توان در شناخت مسیر تحول زبان فارسی به طور عام و تحول در حوزه ساخت‌واژه به طور خاص در مطالعات تاریخی و نیز رده‌شناختی استفاده کرد. مقایسه دو فرایند عمده واژه‌سازی ترکیب و اشتقاق در ۱۰ قرن می‌تواند گرایش‌های زبان فارسی را از نظر ساخت‌واژی روشن کند که به لحاظ نظری و کاربردی در امر واژه‌سازی و واژه‌گزینی حائز اهمیت است. همچنین نتایج این پژوهش شاید بتواند گوشه‌ای از تاریخ اجتماعی این

مرزبوم را در معنای واژگانی که در این ۱۰ قرن ساخته شده‌اند روشن کند که واژه‌های ساخته‌شده در هر دوره اشاره به باورها، مناسبات زیست‌اجتماعی، ابزارهای تولید، عناوین در سلسله‌مراتب طبقات اجتماعی و غیره دارند.

مراجع

- Anderson, S. R. (1982), "Where's Morphology?" *Linguistic Inquiry*, 13, 571–612. 1
- Aronoff, M. (1976), *Word Formation in Generative Grammar*, *Linguistic inquiry monographs*, MIT Press. 2
- Baayen, H. (1992), *Quantitative aspects of morphological productivity*, Dordrecht: Springer Netherlands, pp. 109–149. 4
- (1993), *On frequency, transparency and productivity*, Dordrecht: Springer Netherlands, pp. 181–208. 3
- Baayen, H. and Lieber, R. (1991), "Productivity and English Derivation: a Corpus-Based Study," *Linguistics*, 29, 801–843. 4
- Baayen, R. (1989), *A Corpus-based Approach to Morphological Productivity: Statistical Analysis and Psycholinguistic Interpretation*, *Centrum voor Wiskunde en Informatica*. 4
- Bauer, L. (2001), *Morphological Productivity*, *Cambridge Studies in Linguistics*, Cambridge University Press. 1, 2
- Booij, G. (2007), *The Grammar of Words: An Introduction to Linguistic Morphology: An Introduction to Linguistic Morphology*, *Oxford Textbooks in Linguistics*, OUP Oxford. 3
- Fleischer, W. (1975), *Wortbildung der deutschen Gegenwartssprache*, VEB Bibliographisches Institut. 1
- Plag, I. (2003), *Word-Formation in English*, *Cambridge Textbooks in Linguistics*, Cambridge University Press. 1, 3
- Uhlenbeck, E. (1978), *Studies in Javanese Morphology*, *Translation series*, Martinus Nijhoff. 1
- اسلامی، محرم، شریفی آتشگاه، مسعود، علیزاده لمجیری، صدیقه، وزندی، طاهره (۱۳۸۳)، "واژگان زایای زبان فارسی"، در اولین کارگاه پژوهشی زبان فارسی و رایانه، دانشگاه تهران. ۵
- طباطبائی، علاءالدین (۱۳۸۲)، اسم و صفت مرکب در زبان فارسی. مرکز نشر دانشگاهی. ۲
- مواجی، وحید، اسلامی، محرم، و وزیرنژاد، بهرام (۱۳۹۰)، "پارس مورف: تحلیل‌گر صرفی زبان فارسی"، *دوفصل‌نامه علمی-پژوهشی پردازش علائم و داده‌ها*. ۸. ۵