

نظام آوایی زبان فارسی در آینه آمار

محرم اسلامی^۱

دانشگاه زنجان

افشین رحیمی^۲

دانشجوی دکتری رشته علوم کامپیوتر، دانشگاه ملبورن

سودابه اسلامی^۳

دانشجوی کارشناسی ارشد زبان‌شناسی، دانشگاه الزهرا

چکیده

در پژوهش حاضر، با تحلیل آماری ویژگی‌های واجی موجود در ۵۴۳۹۱ واژه فارسی، کوشیده‌ایم تا جنبه‌هایی از سازوکار نظام آوایی زبان فارسی را با زبان آمار بیان کنیم. جنبه‌های واجی-آماری موجود در واژگان (ذهنی) فارسی‌زبان‌ها، که در این تحقیق به آن پرداخته‌ایم، مواردی مانند الگوی تکیه واژه در زبان فارسی، هجا و تعداد هجا در واژه فارسی، واج‌آرایی واژه‌های فارسی، بسامد طبقات طبیعی واجی در کل داده‌ها، بسامد طبقات طبیعی واجی در درون هجا، واج‌آرایی خوشه‌های همخوانی، هجای بی‌نشان و بهینه در زبان، فراوانی واج‌ها در زبان فارسی و مسائل مشابه را دربرمی‌گیرد. در انتخاب واژه‌ها، ابتدا یک پیکره متنی ده میلیون کلمه‌ای مورد تجزیه ساختوازی قرار گرفت و در ادامه با حذف صورت‌کلمه‌ها و نیز کلمات خارجی فهرست واژه‌ها به دست آمد و سپس این فهرست با فرهنگ معاصر فارسی / امروز مقابله شد و مدخل‌های جدید با بسامد ۱ به فهرست واژگان اضافه شد و سپس کل واژه‌ها از نظر اطلاعات واجی فوق مورد تجزیه و تحلیل قرار گرفت. نتایج این تحقیق واجی-آماری از یک سو مؤید

1. meslami@znu.ac.ir

2. arahimi@student.unimelb.edu.au

3. sudabeheslami@yahoo.com

اصول و مفاهیم واج‌شناختی مانند طبقه طبیعی واجی، فرایندهای واجی، ساختمان سلسله‌مراتبی هجا و غیره است و از طرف دیگر بسامد واج‌ها و زنجیره‌های واجی را بازگو می‌کند، که در نتیجه آن پیوستار بی‌نشانی و نشان‌داری در نظام آوایی زبان مشخص می‌شود. **کلیدواژه‌ها:** تحلیل واجی - آماری، نظام آوایی، واژگان، واج‌آرایی، زبان فارسی.

۱- مقدمه

بررسی آماری نظام آوایی یا همان واج‌شناسی زبان می‌تواند اطلاعات قابل‌توجهی درباره این بخش از زبان در اختیار ما قرار دهد، که بر مبنای آن می‌توان از سازوکار نظام آوایی زبان به‌عنوان بخشی از توانش ناخودآگاه اهل زبان آگاهی پیدا کرد و به دنبال آن گرایش‌های عام واجی در زبان‌ها و نیز گرایش‌های زبان-ویژه را استخراج کرد. از سوی دیگر از این اطلاعات آماری می‌توان در تبیین فرایندهای واجی، شناسایی زنجیره‌های واجی محتمل و بسامد آنها، و به دنبال آن، زنجیره‌های واجی بی‌نشان و نشان‌دار بهره برد. همچنین از اطلاعات آماری نظام آوایی می‌توان در مطالعات آواشناختی، آواشناسی آزمایشگاهی، مطالعات واج‌شناختی و دیگر بررسی‌های زبانی، از جمله ساختواژه استفاده کرد. در کاربردهای عملی مانند واژه‌گزینی، طراحی آزمون در تجزیه و تحلیل گفتار و به‌خصوص گفتار پاتولوژیک و فناوری‌های مربوط به پردازش گفتار نیز از نتایج بررسی آماری نظام آوایی می‌توان سود جست، چراکه در هرکدام از این زمینه‌های تحقیقاتی دنباله‌های واجی محتمل در زبان و بسامد آنها اهمیت ویژه‌ای دارد.

داده‌های پژوهش حاضر را فهرستی بالغ بر ۵۴۳۹۱ واژه فارسی تشکیل می‌دهد، که در واژگان زبانی فارسی (اسلامی و همکاران، ۱۳۸۳) موجود است. واژگان زایا (همان: ۹-۶) تنها حاوی واژه‌های زبان فارسی است و صورت‌های تصریف‌شده کلمات، یعنی صورت‌کلمه‌ها در آن جایی ندارند. در واژگان زایا برای هر مدخل واژگانی اطلاعات زبانی مانند زنجیره واجی واژه، صورت نوشتاری واژه، مقوله دستوری واژه، ساخت هجایی واژه، بسامد واژه‌ها در پیکره زبانی مورد استفاده در تهیة واژگان زایا و الگوی تکیه واژه ارائه شده است. در پردازش متن فارسی لازم است ساختار تصریفی کلمه در زبان صورت‌بندی شود تا با حذف عناصر تصریفی، به واژه‌ها دست پیدا کرد (اسلامی و علیزاده، ۱۳۸۸)، که این امر در پردازش متن اهمیت بسزایی دارد.

نتایج پژوهش حاضر، بنا به ماهیت آماری، فارغ از هرنوع پیش‌دآوری بوده است و اطلاعات ارائه‌شده به‌نوعی نظام آوایی زبان را به تصویر می‌کشد. پژوهشگران در عرصه‌های مختلف، به

فراخور تخصص و موضوع تحقیق، می‌توانند از آمار و ارقام ارائه‌شده در این پژوهش استفاده کنند.

این مقاله با احتساب مقدمه، به‌عنوان بخش اول، از یازده بخش تشکیل شده است. در بخش دوم به معرفی آثار پیشین می‌پردازیم. در بخش سوم الگوی تکیه در واژه‌های زبان فارسی و در بخش چهارم واژه‌های فارسی از حیث تعداد هجا مورد بررسی قرار می‌گیرند. در بخش پنجم زنجیره‌های واجی در ساختمان هجا و آرایش هجا در درون واژه‌های یک‌هجایی تا واژه‌های شش‌هجایی و بیشتر ارزیابی می‌شوند. در بخش ششم بسامد طبقات طبیعی واجی را، فارغ از ساختمان هجا، و در بخش هفتم طبقات واجی را در ساختمان هجا مورد مطالعه قرار داده‌ایم. در بخش هشتم انواع آرایش‌های واجی در خوشه دوهمخوانی هجای CVCC بررسی شده است. بخش نهم هجای بی‌نشان در زبان فارسی را به بحث می‌گذارد و بخش دهم فراوانی واج‌های زبان فارسی در واژگان زبان را نشان می‌دهد و در پایان نتیجه‌گیری مقاله آمده است.

۲- پیشینه پژوهش

پیش‌تر جنبه‌هایی از نظام آوایی زبان فارسی در چند بررسی جداگانه مورد مطالعه آماری قرار گرفته است، که در ادامه به معرفی آنها می‌پردازیم. علم‌الهدی (۲۰۰۰)، نظام آوایی زبان فارسی معاصر را از حیث واج‌آرایی و با هدف بررسی میزان رعایت اصل توالی رسایی، به شیوه آماری، مورد مطالعه قرار داده و نشان می‌دهد که (همان: ۵-۱۷۲) در فارسی نو گرایش به نقض محدودیت رسایی در وام‌واژه‌ها، به‌ویژه در واژه‌های برگرفته از عربی، بیش از چنان گرایشی در فارسی سره است (به نقل از احمدی و بی‌جن‌خان، ۱۳۸۹: ۱۴). شایان ذکر است که علم‌الهدی در بررسی خود از دو دادگان هجاهای فارسی نو، که یکی مختص واژه‌های فارسی سره و دیگری برگرفته از تمامی واژه‌های فارسی است، استفاده کرده و تصریح می‌کند: «در نتیجه ورود واژه‌های دخیل الگوی کلی [آماری در واج‌آرایی] خوشه‌های صامت در زبان فارسی [از حیث محدودیت توالی رسایی] و یا حتی نوع ترجیح‌ها^۱ [از این حیث]، تغییری نکرده، بلکه تنها «شدت» ترجیح [گرایش] هاست که تغییر کرده است (همان: ۱۴). اسلامی و بی‌جن‌خان (۱۳۸۴) در پژوهشی با هدف شناسایی زنجیره‌های واجی بی‌نشان جهت استفاده در امر

1. preference

واژه‌گزینی، ویژگی‌های واج‌آرایی و ساختار هجایی فارسی را مورد بررسی آماری قرار داده‌اند. نتیجه پژوهش اسلامی و بی‌جن‌خان نشان می‌دهد که واژه‌های سه و دوهجایی به ترتیب پربسامدترین واژه‌ها در زبان فارسی هستند (همان: ۴۳۵-۴۳۳). نویسندگان مقاله فوق با تکیه بر داده‌های پژوهش خود و شناسایی زنجیره‌های واجی بی‌نشان‌تر بیان می‌کنند که می‌توان، با توجه به نتیجه پژوهش آنها، میزان پذیرش واژه‌های نوساخته را حداقل از نظر آوایی پیش‌بینی کرد. احمدی و بی‌جن‌خان (۱۳۸۹) خوشه‌های صامت پایانی در فارسی میانه و فارسی نو را، از حیث میزان تبعیت از محدودیت توالی رسایی با یکدیگر مقایسه کرده‌اند. آنها در مقایسه خود سطوح مختلف بسامدی، اعم از «گونه‌ها»، «نمونه‌ها»، «تکرار در پایان هجا» و «تکرار در پایان واژه»، را بررسی کرده‌اند و برای استخراج بخشی از آمارها، ناگزیر بر اساس برخی جهانی‌های واج‌شناختی، انگاره‌ای مفروض برای هجابندی کلمات در فارسی میانه نیز ارائه کرده‌اند (همان: ۱۳). تحلیل‌های آمار استنباطی آنها نشان می‌دهد که در هر چهار سطح بسامدی مذکور، الگوی کمی تبعیت خوشه‌ها از محدودیت توالی رسایی برای فارسی میانه و فارسی نو در سطوح بالا با نسبت‌هایی یکسان است. در پژوهش دیگری اسلامی و همکاران (۱۳۹۰)، ساخت آوایی و واج‌آرایی واژه‌های مصوب فرهنگستان زبان و ادب فارسی (مندرج در *دفترهای اول تا پنجم* از سال ۱۳۸۳ تا ۱۳۸۷) را با نتایج تحقیق اسلامی و بی‌جن‌خان (۱۳۸۴) مورد مقایسه قرار داده‌اند تا نشان دهند عملکرد فرهنگستان زبان و ادب فارسی در حوزه واژه‌گزینی به چه میزان با گرایش‌های غالب در نظام آوایی زبان فارسی همسویی دارد و این همسویی از شرایط اصلی اقبال مردم به واژه‌های نوساخته و واژه‌های مصوب است.

۳- الگوی تکیه در واژه‌های زبان فارسی

پیشتر گفتیم که اطلاعات مربوط به الگوی تکیه واژه‌ها در واژگان زبانی فارسی (اسلامی و همکاران ۱۳۸۳) گنجانده شده است. الگوی تکیه واژه‌ها بر اساس نظریه واج‌شناسی لایه‌ای^۱ (لد، ۲۰۰۸) مشخص شده است که در آن بین دو مفهوم برجستگی ذهنی و عینی از تکیه، تمایز قائل می‌شوند. مفهوم ذهنی از تکیه را تکیه واژگانی^۲ و مفهوم عینی از تکیه را تکیه

1. autosegmental-metrical
2. D. R. Ladd
3. stress

زیرومی^۱ می‌گویند (همان: ۵۰-۴۸). در چارچوب نظریه فوق انواع واژه‌های فارسی (مدخل‌های واژگانی) دارای الگوی تکیه واحدی هستند. اگر برای واژه‌های فارسی یک صورت ذهنی در نظر گرفته شود (همان چیزی که در واژگان ذهنی اهل زبان وجود دارد)، معلوم می‌شود که همه واژه‌های فارسی اعم از اسم، فعل، صفت و قید تکیه‌پایانی هستند (اسلامی: ۱۳۸۸) و هجای آخر آنها تکیه‌بر است. مثلاً در /ke'tâb, zi'bâ, jed'dan, še'kan, še'kast/ ملاحظه می‌کنیم که هجای آخر تکیه‌بر است. می‌بینیم که اگر برای فعل هم یک صورت واژگانی، خواه بن ماضی یا بن مضارع، در نظر گرفته شود، در آن صورت فعل هم تکیه‌پایانی خواهد بود و این دستاورد به همه مباحث پیشین (مثلاً سامعی: ۱۳۷۴)، در مورد الگوهای تکیه فعل و دیگر اقسام کلمه پایان می‌دهد. بر اساس آنچه گفته شد، تقریباً همه واژه‌های فارسی تکیه‌پایانی‌اند. تعداد انگشت‌شماری از واژه‌ها، مانند حروف اضافه و حروف ربط مرکب و واژه‌هایی مانند «اگر، ولی، اما، برخی، بعضی»، تکیه غیرپایانی دارند، که البته این امر دلایل تاریخی دارد و در پژوهش دیگری به آن خواهیم پرداخت.

۴- ساخت هجایی واژه‌های فارسی

هجا مفهومی واج‌شناختی است و آرایش هجاها در درون واژه، عناصر واجی تشکیل‌دهنده هجا، هجای تکیه‌بر در واژه، به‌عنوان واحد واجی زیرزنجیری و غیره، در مطالعه نظام آوایی زبان اهمیت بسزایی دارند. به همین دلیل در این بخش واژه‌های زبان فارسی را از حیث تعداد هجا به شش دسته واژه‌های یک‌هجایی و دسته واژه‌های شش و بیش از شش هجا (تعداد واژه‌های بیش از شش هجا بسیار اندک است و به همین دلیل با واژه‌های شش‌هجایی ادغام شده‌اند)، تقسیم و بسامد هر دسته را محاسبه کردیم تا امکان بررسی‌های بیشتر واج‌شناختی در ادامه فراهم شود. جدول ۱ واژه‌های زبان فارسی را از نظر تعداد هجا در دسته‌های ششگانه به تصویر کشیده است. واژه‌های یک‌هجایی، که واژه‌های دستوری مانند «با، در، و، از» و واحدهای واژگانی مانند «مو، رو، پاک، تاک، عشق، مرد، رفت»^۲ را در خود دارد حدود ۴.۴۵٪ از کل

1. pitch accent

۲. در واژگان زایا برای پرهیز از ارائه قواعد متعدد برای بازنمایی ساختن بن ماضی از بن مضارع، هر دو را در واژگان وارد کرده‌اند.

واژه‌ها را به خود اختصاص داده است. یکی از دلایل تعداد کم واژه‌های یک‌هجایی، غیرممکن بودن واژه‌سازی در قالب هجای واحد است، چراکه حاصل واژه‌سازی حتماً بیش از یک هجا خواهد بود. واژه‌های دوهجایی حدود ۲۳٪ از کل را تشکیل می‌دهند که، در این دسته نیز مانند یک‌هجایی‌ها، واژه‌های پایه‌ی زبان مانند «مادر، پدر، پسر، خواهر، بابا» و واژه‌های دیگر مانند «کتاب، لفل، سیاه» و همچنین واژه‌های پیچیده مانند «نادان، بی‌رحم، کارگر» قرار دارند. واژه‌های سه و چهارهجایی به ترتیب ۳۲٪ و ۲۴٪ را شامل می‌شوند و بسامد بالای آنها ناشی از حضور واژه‌های پیچیده در این دو دسته است. ترکیب و اشتقاق دو فرایند اصلی واژه‌سازی در زبان‌ها، از جمله زبان فارسی، هستند که حاصل آن واژه‌های پیچیده سه و چهارهجایی است. دو دسته آخر بسامد کمتری دارند، که نشان می‌دهد اگرچه به لحاظ نظری منعی برای ساختن واژه‌های طولانی وجود ندارد، ولی اصل ارتباط بهینه در نظام زبان مانع آن می‌شود.

جدول ۱- پراکندگی واحدهای واژگان برحسب تعداد هجاهای آنها

تعداد هجا	بسامد واژگانی	درصد بسامد واژگانی
۱	۲۴۲۱	۴.۴۵
۲	۱۲۳۹۵	۲۲.۷۸
۳	۱۷۵۳۰	۳۲.۲۲
۴	۱۳۰۵۱	۲۳.۹۹
۵	۶۳۱۴	۱۱.۶۰
۶ و بیشتر	۲۶۷۳	۴.۹۱
جمع	۵۴۳۹۱	۱۰۰

۵- زنجیره‌های واجی در ساختمان هجا و آرایش هجا در درون واژه‌ها

در این بخش به بررسی زنجیره‌های واجی در ساختمان هجا و آرایش هجاها در درون واژه‌ها می‌پردازیم. هجای فارسی از نظر ساختمان دارای سه ترکیب زنجیری به صورت CVCC, CVC, CV است (ثمره، ۱۳۷۸)، که در آن C نماینده همخوان و V نماینده واکه است. می‌بینیم که ساختمان هجا در زبان فارسی به‌غایت ساده است و به همین دلیل تعیین مرز هجا در

پردازش‌های خودکار زبان فارسی با کمترین هزینه قابل انجام است. در ادامه به بررسی واج‌آرایی در ساختمان واژه‌های فارسی می‌پردازیم.

۵-۱- زنجیره‌های واجی در واژه‌های یک‌هجایی

با توجه به جدول ۲، واژه‌های یک‌هجایی در قالب هجای نشان‌دار CVCC بیشترین بسامد و در قالب هجای بی‌نشان CV کمترین بسامد را دارد. دلیل این امر امکانات ترکیبی عناصر واجی در درون هجای CVCC است. در قالب هجای CV بالقوه می‌توان از ترکیب ۲۳ همخوان و ۶ واکه، ۱۳۸ واژه داشت که البته با بسامد ۱۶۱ واژه در جدول ۲ مغایرت دارد. اگرچه خلأهای واژگانی هم در این ساخت هجایی هست که نباید نادیده گرفته شود. مثلاً ما واژه‌های /xe, xi, .../ در زبان فارسی نداریم. برای بررسی این تناقض، واژگان زایا را به صورت دستی بررسی کردیم و متوجه شدیم که از یک طرف حروف الفبا و تلفظ اختصارات به عنوان مدخل آمده است، چراکه هدف از تهیه واژگان زایا استفاده از آن در تبدیل خودکار متن به گفتار در زبان فارسی بوده است و به همین دلیل لازم بوده تلفظ حروف و اختصارات در واژگان گنجانده شود. از سوی دیگر برخی از مدخل‌ها، مقوله‌های دستوری چندگانه داشتند. مثلاً /do/ هم با مقوله عدد و هم با مقوله فعل (بن مضارع) وارد واژگان شده است و همچنین مواردی مانند /pâ/ که به طور همزمان مقوله اسم و فعل (بن مضارع) دارد. در شمارش دستی معلوم شد که ۶۷ واژه در داده‌های ما ساخت CV دارند.

جدول ۲- پراکندگی ساخت هجایی واژه‌های تک‌هجایی

ساخت هجایی	بسامد واژگانی	درصد بسامد واژگانی
CVCC	۱۲۳۵	۵۱.۰۱
CVC	۱۰۲۵	۴۲.۳۳
CV	(۶۷)۱۶۱	۶.۶۵
جمع	۲۴۲۱	۱۰۰

در نگاه اول تصور می‌شود که در واژه‌های یک‌هجایی، هجای CVCC بیشتر از دو ساخت دیگر بسامد دارد، ولی اگر نسبت عناصر تشکیل‌دهنده هجا را در نظر بگیریم، متوجه می‌شویم

که باز نسبت هجای CV، با توجه به تعداد واج‌های موجود در آن، بیشتر از بقیه است. جدول ۳ این موضوع را نشان می‌دهد:

جدول ۳- واژه‌های یک‌هجایی و بسامد آنها با توجه به نسبت عناصر واجی در درون هجا

هجای	تعداد ممکن	تعداد موجود	درصد موجود	درصد خلأ
CV	۱۳۸	۶۷	٪۴۹	٪۵۱
CVC	۳۱۷۴	۱۰۲۵	٪۳۲	٪۶۸
CVCC	۷۳۰۰۲	۱۲۳۵	٪۱.۷	٪۹۸.۳

با توجه به جدول ۳، اگرچه عملاً تعداد واژه‌های یک‌هجایی CV اندک است ولی، اگر نسبت عناصر تشکیل‌دهنده هجا را در نظر بگیریم، می‌بینیم که هجای CV، ٪۴۹ ترکیب‌های ممکن را دارد و بسامد آن بیشتر از دو ساخت دیگر است. این به آن معناست که مثلاً در هجای CVCC تنها ٪۱.۷ امکان ترکیبی واج‌ها فعلیت پیدا می‌کند و ٪۹۸.۳ موارد دارای خلأ واج‌آرایی هستند. یعنی زبان فارسی قابلیت ساخت ۷۳۰۰۲ واژه در قالب هجای CVCC را دارد، ولی از اندک ظرفیت خود استفاده می‌کند. حال در واژه‌های چندهجایی این توان ترکیبی به اعداد نجومی می‌رسد. به‌عنوان مثال در قالب واژه دوهجایی CVCC.CVCC می‌توان ۵۳۲۹۲۹۲۰۰۴ (۷۳۰۰۲*۷۳۰۰۲) واژه داشت، درحالی‌که در واژگان زایا تنها ۱۳۰ واژه دوهجایی با این ساخت وجود دارد. این در حالیست که در مقام تلفظ، غالباً همخوان پایانی در خوشه دوهمخوانی حذف می‌شود. مثلاً تلفظ /dastband/ به‌صورت [dasban] است. تعداد نسبی واژه‌های یک‌هجایی با ساخت CV و نیز حذف همخوان از خوشه دوهمخوانی و تبدیل CVCC به CVC و هجاسازی مجدد با درج واکه و به دنبال آن تبدیل CVC به CV مثلاً در /kârgar/ به‌صورت [kâregar]، همه حکایت از این موضوع دارد که هجای CV هجای مطلوب در زبان فارسی است و در فرایندهای واجی، هجاهای دیگر به آن تبدیل و یا نزدیک می‌شوند. البته با در نظر گرفتن محدودیت‌های حاکم بر زبان، مطالب فوق، و شواهدی که در ادامه خواهیم دید، حکایت از نظام بهینه‌ای به نام زبان دارد که همهٔ ارکانش از جمله نظام آوایی آن در بهینه‌ترین صورت قرار

دارد. نظامی چنان قوی که با بخش ناچیزی از قابلیت‌هایش کار ارتباط صورت می‌گیرد. نظامی که عناصر اولیه آن را مثلاً ۲۹ واج تشکیل می‌دهد.

۵-۲- زنجیره‌های واجی در واژه‌های دوهجایی

جدول ۳ بسامد واژگانی واژه‌های دوهجایی و واج‌آرایی‌های ممکن در این واژه‌ها را نشان می‌دهد. بخش اعظم این واژه‌ها، در فارسی امروز، بسیط و در زمره کلمات پایه‌ای زبان به حساب می‌آیند. مانند: «دختر، مادر، پسر، پدر، خواهر، ...». بسامد بالا نشانه بی‌نشانی ساخت‌های هجایی است، اگرچه، همان‌طور که در بالا گفتیم، ساخت‌های نشان‌دار هم در اثر فرایندهای واجی به ساخت‌های بی‌نشان تبدیل و یا نزدیک می‌شوند. در جدول ۴ می‌بینیم که ساخت‌های هجایی CVC-CVC و CV-CVC بیشترین بسامد را دارند. این موضوع نشان می‌دهد که به‌عنوان مثال واژه «بالگرد» با ساخت هجایی CVC-CVCC دارای ساختار بی‌نشان‌تری نسبت به واژه «چرخ‌بال» با ساخت هجایی CVCC-CVC است و احتمال پذیرش آن از سوی اهل زبان بیشتر است.

جدول ۴- پراکندگی ساخت هجایی واژه‌های دوهجایی

ساخت هجایی		بسامد واژگانی	درصد بسامد واژگانی
CVC	CVC	۴۲۰۲	۳۳.۹۰
CV	CVC	۳۵۴۲	۲۸.۵۷
CVC	CV	۱۵۱۳	۱۲.۲۰
CV	CV	۱۱۸۲	۹.۵۳
CVC	CVCC	۷۸۲	۶.۳۰
CV	CVCC	۵۸۶	۴.۷۲
CVCC	CVC	۳۹۱	۳.۱۵
CVCC	CVCC	۱۳۰	۱.۰۴
CVCC	CV	۶۷	۰.۴۵
جمع		۱۲۳۹۵	۱۰۰

۵-۳- زنجیره‌های واجی در واژه‌های سه‌هجایی

واژه‌های سه‌هجایی بالقوه ۲۷ ساخت هجایی می‌توانند داشته باشند، که در جدول ۵ آمده است. این جدول بسامد ساخت‌های هجایی و واج‌آرایی واژه‌های سه‌هجایی را از بیشترین به کمترین نشان می‌دهد، که از این بین تنها ۱۲ ساخت هجایی بسامد نسبی بالا دارند و ۱۵ ساختار سه‌هجایی بسامدی کمتر از ۳٪ دارند. که حکایت از نشان‌داری این ساخت‌ها است و به همین دلیل، جهت پرهیز از طولانی‌شدن گزارش نتایج پژوهش حاضر، از ذکر آنها و موارد مشابه در ادامه خودداری کرده‌ایم.

جدول ۵- زنجیره‌های واجی در واژه‌های سه‌هجایی

ساخت هجایی	بسامد واژگانی	درصد بسامد واژگانی
cvc-cv-cv	۳۰۲۹	۱۷.۲
cv-cv-cv	۲۷۵۹	۱۵.۷
cv-cv-cvc	۲۷۴۰	۱۵.۶
cvc-cv-cvc	۲۲۱۸	۱۲.۶
cv-cvc-cvc	۱۹۰۳	۱۰.۸
cvc-cvc-cvc	۱۰۵۷	۶
cv-cvc-cv	۸۰۰	۴.۵
cvc-cvc-cv	۷۱۶	۴
cvcc-cv-cv	۴۶۷	۲.۶
cv-cv-cvcc	۴۳۹	۲.۵
cvc-cv-cvcc	۳۸۵	۲.۱
cv-cvc-cvcc	۳۳۹	۱.۹
cvcc-cv-cvc	۱۷۰	۰.۹
سایر	۵۰۸	۲.۸
جمع	۱۷۵۳۰	۱۰۰

۵-۴- زنجیره‌های واجی در واژه‌های چهارهجایی

واژه‌های چهارهجایی فارسی دارای توان ترکیبی در قالب ۸۱ ساخت هجایی هستند و در داده‌های ما ۱۷ ساختار، بسامد نسبی بالا و ۵۸ ساختار دیگر، مجموعاً ۶.۸٪ بسامد دارند، که

نظام آوایی زبان فارسی در آینه آمار / ۷۵

زبان فارسی برای بسیاری از این ساخت‌های چهارهجایی واژه‌ای ندارد. جدول ۶ ریز اطلاعات را نشان می‌دهد.

جدول ۶- زنجیره‌های واجی در واژه‌های چهارهجایی

ساخت هجایی	بسامد واژگانی	درصد بسامد واژگانی
CV-CV-CV-CV	۱۶۹۸	۱۳
CVC-CV-CV-CV	۱۶۰۸	۱۲.۳
CV-CVC-CV-CV	۱۴۹۸	۱۱.۴
CV-CV-CV-CVC	۱۲۳۳	۹.۴
CVC-CVC-CV-CV	۱۱۰۷	۸.۴
CVC-CV-CV-CVC	۸۸۷	۶.۷
CV-CV-CVC-CVC	۶۶۸	۵.۱
CV-CVC-CV-CVC	۶۴۹	۴.۹
CVC-CVC-CV-CVC	۴۴۰	۳.۳
CV-CV-CVC-CV	۴۰۸	۳.۱
CVC-CV-CVC-CVC	۴۰۰	۳
CVC-CV-CVC-CV	۳۷۸	۲.۸
CV-CVC-CVC-CV	۳۴۸	۲.۶
CVCC-CV-CV-CV	۲۴۴	۱.۸
CV-CV-CV-CVCC	۲۲۴	۱.۷
CV-CVC-CVC-CVC	۲۱۴	۱.۶
CVC-CVC-CVC-CV	۱۴۲	۱
سایر	۹۰۵	۶.۸
جمع	۱۳۰۵۱	۱۰۰

۵-۵- زنجیره‌های واجی در واژه‌های پنج‌هجایی

در واج‌آرایی واژه‌های پنج‌هجایی، بالقوه می‌توان ۲۴۳ آرایش هجایی را دید، که از بین آنها ۲۳ مورد بسامد نسبی بالا، و سایر ۲۲۰ ساخت هجایی مجموعاً بسامدی کمتر از ۱.۱٪ دارند، که در بیشتر این ساخت‌ها خلأ واژگانی داریم. جدول ۶ این اطلاعات را بازنمایی می‌کند.

جدول ۷- زنجیره‌های واجی در واژه‌های پنج‌هجایی

ساخت هجایی	بسامد واژگانی	درصد بسامد واژگانی
CV-CV-CV-CV-CV	۷۸۷	۱۲.۴
CVC-CV-CV-CV-CV	۵۹۵	۹.۴
CV-CVC-CV-CV-CV	۵۸۶	۸.۹
CV-CV-CVC-CV-CV	۴۲۸	۶.۷
CV-CV-CV-CV-CVC	۴۰۹	۶.۴
CVC-CVC-CV-CV-CV	۳۸۸	۶.۱
CVC-CV-CVC-CV-CV	۳۳۶	۵.۳
CVC-CV-CV-CV-CVC	۲۷۱	۴.۲
CV-CVC-CVC-CV-CV	۲۲۸	۳.۶
CVC-CVC-CV-CVC-CV	۲۲۰	۳.۴
CV-CVC-CV-CVC-CV	۱۹۰	۳
CV-CV-CVC-CV-CVC	۱۵۰	۲.۳
CV-CV-CV-CVC-CV	۱۴۸	۲.۳
CV-CVC-CV-CV-CVC	۱۴۶	۲.۳
CVC-CV-CV-CVC-CV	۱۲۳	۱.۹
CV-CV-CV-CVC-CVC	۱۱۵	۱.۸
CVC-CV-CVC-CV-CVC	۱۱۴	۱.۸
CVC-CVC-CVC-CV-CV	۱۰۶	۱.۶
CVC-CV-CV-CVC-CVC	۱۰۱	۱.۵
CVC-CVC-CV-CV-CVC	۷۸	۱.۲
CV-CV-CVC-CVC-CVC	۷۵	۱.۱
CV-CV-CVC-CVC-CV	۶۸	۱
سایر	۶۷۰	۱۰.۶
جمع	۶۳۱۴	۱۰۰

۵-۶- زنجیره‌های واجی در واژه‌های شش‌هجایی

برای واژه‌های شش‌هجایی، ۷۲۹ آرایش همجایی ممکن متصور است که ۲۶ مورد بسامد نسبی بالا، و بقیه ۷۰۳ ساخت همجایی، بسامدی کمتر از ۱۷٪ دارند. اگرچه بیشتر واج‌آرایی‌های ممکن برای واژه‌های شش‌هجایی فاقد نمونه‌ای در زبان فارسی است. جدول واج‌آرایی واژه‌های شش‌هجایی را به دلیل محدودیت حجم مقاله و نیز شباهت آن به اطلاعات گروه‌های دیگر در اینجا نیاورده‌ایم. برای واژه‌های بیش از شش هجا نیز، به دلیل اندک‌بودن تعداد آنها، توضیح جداگانه‌ای در نظر نگرفته‌ایم، ولی تعداد آنها را در ارائه آمار نظام آوایی زبان محاسبه کرده‌ایم.

۶- بسامد طبقات واجی در زبان فارسی

در این بررسی، همخوان‌های زبان فارسی را از نظر رسایی به دو طبقه رسا و غیررسا تقسیم کرده‌ایم. همخوان‌های رسا را با c2 و همخوان‌های غیررسا را با c1 نمایش می‌دهیم. همخوان‌های رسا شامل /r,y,l,m,n/ و سایر همخوان‌ها غیررسا هستند. واکه‌ها بر اساس پسین و پیشین بودن جایگاه تولید به دو طبقه واکه‌های پیشین (a, e, i/v1) و واکه‌های پسین (â, o, u/v2) تقسیم و همچنین بر اساس ارتفاع زبان به سه طبقه واکه‌های باز (a, â/v3)، واکه‌های متوسط (e, o/v4) و واکه‌های بسته (i,u/v5) تقسیم شده‌اند. جدول ۸ آمار جالبی از توزیع و بسامد نسبی واج‌های زبان فارسی و نیز طبقات واجی در اختیار ما می‌گذارد. با نگاهی به آمار ارائه‌شده در این جدول می‌بینیم که:

الف. طول متوسط هر واژه در زبان فارسی حدود ۸ واج است، که از تقسیم تعداد واج‌ها (۴۴۴۶۳۸) به تعداد واژه‌ها (۵۴۳۹۱) به‌دست می‌آید.

ب. بیش از ۴۰٪ واج‌ها در ساختمان واژه فارسی را واکه‌ها تشکیل می‌دهند، البته مبنای تحلیل ما واژه‌های فارسی است و اگر کسره اضافه و تکواژهای غیراشتقاقی، مانند یای نکره و یای بند موصولی، که تنها یک واکه هستند و از طرف دیگر بسامد بالایی هم دارند، را به این تعداد اضافه کنیم، متوجه خواهیم شد که در گفتار روزمره، این رقم بالاتر نیز می‌رود. این درحالی است که اگر بسامد نسبی واکه به همخوان محاسبه شود، مطابق جدول ۸ همین رقم به حدود ۷۲٪ می‌رسد.

ج. بسامد همخوان‌های رسا بیش از دو برابر همخوان‌های غیررسا است.

د. پسین و پیشین‌بودنِ واژه‌ها تفاوت معناداری از حیث بسامد دارد و شاهدهی بر طبیعی‌بودنِ این دو طبقهٔ واجی می‌تواند باشد.

ه. ارتفاع زبان با بسامد رابطهٔ عکس دارد. یعنی هرچه ارتفاع زبان بالاتر باشد، بسامد طبقهٔ واجی کمتر است و برعکس.

و. ارتفاع زبان با رسایی رابطهٔ عکس دارد. یعنی هرچه ارتفاع زبان بالاتر باشد، رسایی طبقهٔ واجی کمتر است و برعکس.

ز. بسامد طبقهٔ واجی و رسایی رابطهٔ مستقیم دارند. یعنی هرچه بسامد بالاتر باشد، واج و طبقهٔ واجی رساتر است.

بسامد طبقات واجی در جدول ۸ بدون در نظر گرفتن ساخت هجایی محاسبه شده است. در این محاسبه بسامد با توجه به تعداد اعضای هر گروه نرمال شده است تا معیار دقیق‌تری از بسامد هر طبقه در واژگان به دست آید. نرمال‌سازی با تقسیم بسامد هر طبقه بر تعداد اعضای طبقه و نسبت آن با مکمل آن طبقه محاسبه شده است. به‌عنوان مثال بسامد واژگانی گروه C1 تقسیم بر ۱۸ و بسامد واژگانی گروه C2 تقسیم بر ۵ شده است. با این تقسیم، نسبت نرمال‌شدهٔ فراوانی طبقهٔ C2 به C1 برابر با ۲۰۹ شده است. محاسبهٔ نسبت C به ۷ نیز دقیقاً به همین روش انجام گرفته است. بسامد نسبی طبقات از روی نسبت نرمال‌شده محاسبه می‌شود.

در جدول ۸، ما طبقات واجی، تعداد اعضا و بسامد هر طبقه، درصدی که هر کدام از طبقات واجی به خود اختصاص می‌دهند، نسبت طبقات به همدیگر و بسامد نسبی طبقات واجی را می‌بینیم. جدول ۸ کاملاً مقیاس رسایی پیش‌بینی‌شده در نظریه‌های زبانی را تأیید می‌کند.

۷- طبقات واجی در ساختمان هجا

در این بخش به حضور طبقات واجی در درون هجاهای زبان فارسی می‌پردازیم. بسامد و رفتار این طبقات می‌تواند شاهد طبیعی‌بودنِ آنها و نیز درستی ساختمان سلسله‌مراتبی برای هجا باشد. برای این منظور هجاهای سه‌گانهٔ فارسی از تمامی واژگان استخراج شد و در هر هجا بسامد طبقه‌های واجی مورد بررسی قرار گرفت و سپس بسامد به‌دست‌آمده، با توجه به تعداد اعضای طبقه‌های واجی، نرمال شد تا معیار نسبی دقیق‌تری از حضور طبقه‌های واجی در هجای فارسی به دست آید. همان‌گونه که در جدول ۹ مشاهده می‌شود، در هجای CV بسامد رخداد

نظام آوایی زبان فارسی در آینه آمار / ۷۹

واکه‌های باز (v3) بیشتر از سایر واکه‌ها و همچنین میزان رخداد گروه همخوان‌های رسا بیشتر از غیررسا است، که تأییدی بر اعتبار اصل رسایی در واژگان فارسی است.

جدول ۸- پراکندگی طبقات واجی (فارغ از ساخت هجایی) و اصل رسایی

طبقه واجی	تعداد اعضا	بسامد	درصد طبقه واجی	نسبت	بسامد نسبی	
c1	۱۸	۱۶۶۹۱۱	۳۷.۷۷	۱	۳۲.۳۶	c23=%100
c2	۵/m,n,l,r,y/	۹۷۲۸۱	۲۱.۷۵	۲.۰۹	۶۷.۶۴	
v1	۳/a,e,i/	۱۰۰۷۵	۲۵.۰۴	۱	۶۱	v6=%100
v2	۳/ â,o,u/	۷۰۳۷۱	۱۵.۴۳	۱	۳۹	
v3	۲/a, â /	۸۴۱۶۵	۱۹.۸۹	۱	۴۶.۶۴	v6=%100
v4	۲/e,o/	۵۵۱۸۴	۱۱.۷۳	۱	۳۰.۵۸	
v5	۲/i,u/	۴۱۰۹۷	۸.۸۵	۱	۲۲.۷۷	
c	۲۳	۲۶۴۱۹۲	۵۹.۵۳	۱	۲۷.۷	c=%100
v	۶	۱۸۰۴۴۶	۴۰.۴۷	۲.۶۱	۷۲.۳	v=%100
جمع	۲۹	۴۴۴۶۳۸	۱۰۰		۱۰۰	

۷-۱- طبقات واجی در هجای CV

در جدول ۹ پراکندگی طبقات واجی در هجای CV فارسی را شاهد هستیم، که در آن بسامد طبقات طبیعی واکه‌ها با ارتفاع زبان ارتباط عکس دارد و واکه‌های باز: الف. در هجای CV صرف‌نظر از نوع همخوان دارای بیشترین بسامد هستند، ب. نوع همخوان در هجای CV دارای اهمیت چندانی نیست و این رفتار واجی وجود ساخت سلسله‌مراتبی برای هجا را تأیید می‌کند، که به موجب آن واکه، به‌عنوان هسته هجا، همخوان پیش از خود را تحت فرمانش سازه‌ای ندارد و بر آن مسلط نیست. در این مورد در ادامه بیشتر بحث خواهیم کرد.

جدول ۹- پراکندگی طبقات واجی (در درون هجای CV) و اصل رسایی

طبقة واجی	بسامد طبقه واجی	درصد بسامد	بسامد طبقه واجی در نسبت اعضا	درصد بسامد طبقه واجی در نسبت اعضا
c1v3	۲۹۰۷۴	۲۷.۸۹	۸۲۱۶	۲۱.۷۲
c1v4	۲۲۲۵۱	۲۱.۳۵	۵۳۵۵	۱۴.۱۵
c1v5	۱۸۳۲۴	۱۷.۵۸	۴۷۲۰	۱۲.۴۸
c2v3	۱۲۱۱۲	۱۱.۶۲	۷۰۳۰	۱۸.۵۸
c2v4	۱۱۵۵۸	۱۱.۰۹	۶۶۰۳	۱۷.۴۵
c2v5	۱۰۸۹۲	۱۰.۴۵	۵۹۱۰	۱۵.۶۲
جمع	۱۰۴۲۱۱	۱۰۰		۱۰۰

۷-۲- طبقات واجی در هجای CVC

در جدول ۱۰ حضور طبقات واجی در هجای CVC را می‌بینیم که: الف. توزیع واکه‌ها در هجای CVC با ارتفاع زبان ارتباط عکس دارد، ب. همخوان آغازی در هجای CVC تحت فرمانش سازه‌ای هسته هجا نیست، لذا حضور آنها قبل از هسته تصادفی است. همخوان‌ها در هجای CVC ، بعد از هسته، مقید به ساختمان هجا هستند و به همین دلیل بسامد همخوان‌های رسا بیشتر از غیررسا است. به‌ویژه اگر نسبت اعضای این دو طبقه در نظر گرفته شود، موضوع مقیدبودن همخوان‌ها در هجای CVC بعد از هسته به ساختمان هجا بیشتر خودنمایی می‌کند. مثلاً هجای c2v3c2 اگرچه دارای بسامد ۴۸۷۸ است، ولی با احتساب نسبت ۵ به ۱۸ دو طبقه رسا و غیررسا، این هجا دارای بیشترین بسامد می‌شود. این امر در بقیه موارد نیز صادق است.

جدول ۱۰- پراکندگی طبقات واجی (در درون هجای CVC) و اصل رسایی

طبقة واجی	بسامد طبقه واجی	درصد بسامد	بسامد طبقه واجی در نسبت اعضا	درصد بسامد طبقه واجی در نسبت اعضا
c1v3c2	۱۵۳۳۷	۲۲.۳۲	۱۳۳.۳۶	۱۸.۷۷
c1v3c1	۱۱۷۵۵	۱۷.۱	۲۲.۲۲	۳.۱۲
c1v4c2	۷۱۹۴	۱۰.۴۷	۶۲۵۵	۸.۸
c1v4c1	۶۶۱۹	۹.۶۳	۱۲.۵۱	۱.۷۶
c2v3c1	۶۳۷۶	۹.۲۸	۵۵.۴۴	۷.۸
c2v3c2	۴۸۷۸	۷.۱	۱۹۵.۱۲	۲۷.۴۸
c1v5c2	۳۷۲۲	۵.۴۲	۳۲.۳۶	۴.۵۵
c2v4c1	۳۶۳۴	۵.۲۹	۳۱.۶	۴.۴۴
c1v5c1	۳۴۵۰	۵.۰۲	۶.۵۲	۰.۹
c2v5c1	۲۲۹۰	۳.۳۳	۱۹.۹۱	۲.۸
c2v4c2	۱۷۳۸	۲.۵۳	۶۹.۵۲	۹.۷۸
c2v5c2	۱۷۳۱	۲.۵۲	۶۹.۲۴	۹.۷۸
جمع	۶۸۷۲۴	۱۰۰		۱۰۰

۷-۳- طبقات واجی در هجای CVCC

جدول ۱۱ پراکندگی طبقات واجی در هجای CVCC را نشان می‌دهد و در اینجا نیز شاهد: الف. بسامد بالای طبقات واکه‌ای رساتر هستیم، ب. توزیع همخوان‌ها در خوشه همخوانی تابع ساخت سلسله‌مراتبی هجا و اصل رسایی است، ج. مطابق جدول ۱۲ در خوشه همخوانی در ۷۹٪ مواقع همخوان اول رسا و در ۲۱٪ غیررسا است و این امر نشان می‌دهد که در اکثر مواقع از هسته به پایان هجا شاهد افت رسایی هستیم و جدول ۱۲ این موضوع را نشان می‌دهد.

جدول ۱۱- پراکندگی طبقات واجی (در درون هجای CVCC) و اصل رسایی

طبقه واجی	بسامد طبقه واجی	درصد بسامد	بسامد طبقه واجی در نسبت اعضا	درصد بسامد طبقه واجی در نسبت اعضا
c1v3c2c1	۱۸۰۲	۲۳.۹۹	۶۸۱	۱۴
c1v3c1c1	۱۰۴۴	۱۳.۹	۸۵	۲
c1v4c2c1	۶۴۱	۸.۵۳	۲۴۲	۵
c2v3c2c1	۶۴۱	۸.۵۳	۱۱۱۴	۲۳
c1v4c1c1	۴۷۸	۶.۳۶	۳۹	۱
c1v3c1c2	۴۵۶	۶.۰۷	۱۷۲	۳
c2v3c1c1	۳۸۳	۵.۱	۱۴۴	۳
c1v4c1c2	۳۰۹	۴.۱۱	۱۱۶	۲
c1v4c2c2	۲۶۲	۳.۴۹	۴۵۵	۹
c1v5c1c1	۲۳۹	۳.۲۸	۱۹	۰
c2v4c1c1	۲۲۰	۲.۹۳	۸۳	۲
c1v3c2c2	۱۵۲	۲.۰۲	۲۶۴	۵
c2v5c1c2	۱۳۵	۱.۸	۲۳۴	۵
c2v3c1c2	۱۳۱	۱.۷۴	۲۲۷	۵
c2v5c1c1	۱۲۸	۱.۷	۴۸	۱
c2v4c2c1	۱۲۷	۱.۶۹	۲۲۰	۴
c2v4c1c2	۱۲۱	۱.۶۱	۲۱۰	۴
c1v5c1c2	۸۴	۱.۱۲	۳۱	۱
c1v5c2c1	۶۱	۰.۸۱	۲۳	۰
c2v4c2c2	۳۲	۰.۴۳	۲۵۶	۵
c2v3c2c2	۲۴	۰.۳۲	۱۹۲	۴
c1v5c2c2	۲۱	۰.۲۸	۳۶	۱
c2v5c2c1	۲۰	۰.۲۷	۳۴	۱
جمع	۲۰۲۴	۱۰۰		۱۰۰

جدول ۱۲- ارزیابی اصل رسایی در جایگاه عضو اول خوشه دوهمخوانی در هجای CVCC

درصد طبقه طبیعی	طبقه طبیعی
۷۸.۶	رساها
۲۱.۴	غیر رساها
۱۰۰	جمع

۸- آرایش واجی در خوشه دوهمخوانی هجای CVCC

در ادامه به بررسی حضور، آرایش و بسامد طبقات واجی در خوشه دوهمخوانی CVCC می‌پردازیم.

۸-۱- بسامد همخوانی در خوشه دوهمخوانی هجای CVCC

در اینجا تمامی خوشه‌های دوهمخوانی در هجای CVCC از کل واژگان استخراج گردید و بسامد رخداد ترکیب‌های مختلف واجی محاسبه شد. همان گونه که در جدول ۱۳ مشاهده می‌کنیم بسامد واج /n/ در جایگاه همخوان اول خوشه و همچنین میزان رخداد واج /t/ در جایگاه همخوان دوم خوشه، خودنمایی می‌کند. از آنجا که جمعاً ۵۲۹ (۲۳*۲۳) آرایش واجی ممکن در این خوشه وجود دارد و از این میان بسیاری از ساخت‌ها دارای خلأ و بسیاری نیز دارای بسامد بسیار پایین (در کل ۰.۴۲٪) بودند، برای پرهیز از طولانی‌شدن بحث موارد فوق را در جدول ۱۳ نیاورده‌ایم.

۸-۲- بسامد واجی عضو اول خوشه در هجای CVCC

جدول ۱۴ بسامد واجی عضو اول خوشه در هجای CVCC را به نمایش می‌گذارد، که در آن سه واج /n,t,s/ بیش از ۵۴٪ از کلیه واج‌های رخ داده در جایگاه همخوان اول خوشه دوهمخوانی را تشکیل داده‌اند. لازم به توضیح است که پدیدآورندگان واژگان زایا (اسلامی و همکاران، ۱۳۸۳) بنا به اهداف کاربردی (تبدیل خودکار متن به گفتار)، در صورت واجی واژه‌هایی که در ساختمان آنها، به لحاظ آوایی، اصطلاحاً واکه مرکب [ou] وجود داشته، آن را به صورت /ow/

نشان داده‌اند تا تولید ماشینی کلماتی مانند «موز، قوم، نوروز» طبیعی‌تر باشد و از طرف دیگر حجم کار برنامه‌نویسی به حداقل برسد. لذا W را به فهرست همخوان‌ها اضافه کرده‌اند.

جدول ۱۳- پراکندگی خوشه CC در هجای CVCC

عضو اول خوشه	عضو دوم خوشه	بسامد واژگانی	درصد بسامد
n	d	۸۶۵	۱۱.۵۲
s	t	۶۹۰	۹.۱۹
n	g	۴۵۵	۶.۰۶
r	d	۴۰۲	۵.۳۵
š	T	۲۹۶	۳.۹۴
n	j	۲۰۵	۲.۷۳
s	m	۱۹۴	۲.۵۸
x	t	۱۸۹	۲.۵۲
h	r	۱۵۷	۲.۰۹
f	t	۱۳۵	۱.۸
x	s	۱۲۰	۱.۶
r	g	۱۰۸	۱.۴۴
r	m	۱۰۵	۱.۴
y	n	۱۰۳	۱.۳۷
r	z	۹۵	۱.۲۶
k	s	۸۹	۱.۱۸
t	r	۸۴	۱.۱۲
r	t	۷۷	۱.۰۳
سایر		۳۱۴۲	۴۲

جدول ۱۴- بسامد واجی عضو اول خوشه در هجای CVCC

عضو اول خوشه	بسامد واجی	درصد واجی	عضو اول خوشه	بسامد واجی	درصد واجی
n	۱۷۲۹	۲۳.۰۲	q	۱۷۱	۲.۲۸
r	۱۲۹۵	۱۷.۲۴	z	۱۶۹	۲.۲۵
s	۱۰۹۶	۱۴.۵۶	w	۱۵۳	۲.۰۴
š	۵۰۱	۶.۶۷	b	۱۴۱	۱.۸۸
x	۳۶۰	۴.۷۹	ɟ	۹۴	۱.۲۵
y	۳۴۹	۴.۶۵	d	۹۳	۱.۲۴
h	۲۹۰	۳.۸۶	ʃ	۳۷	۰.۴۹
l	۲۳۶	۳.۱۴	p	۹	۰.۱۲
f	۲۲۸	۳.۰۴	v	۷	۰.۰۹
k	۲۰۳	۲.۷	g	۳	۰.۰۴
M	۱۷۴	۲.۳۲	ʒ	۱	۰.۰۱
t	۱۷۲	۲.۲۹			

۸-۳- بسامد واجی عضو دوم خوشه در هجای CVCC

جدول ۱۵ بسامد واجی عضو دوم خوشه در هجای CVCC را نشان می‌دهد، که در آن سه واج /t,d,m/ بیش از ۴۷٪ از ظرفیت جایگاه دوم در خوشهٔ دوهمخوانی را پر کرده‌اند.

۹- هجای بی‌نشان در زبان فارسی

کلیهٔ هجاهای سه‌گانهٔ فارسی از ۵۴۳۹۱ واژهٔ زبان فارسی استخراج و میزان رخداد هریک در ساختار هجایی واژه‌ها بررسی و بسامد آنها محاسبه شد. مشخص شد که ۵۷٪ از کل هجاهای زبان فارسی، ساخت هجایی CV دارند و بسامد هجاهای CVC و CVCC به ترتیب ۳۸٪ و ۴٪ است. بنابراین هجای CV، به دلیل بسامد بالا، هجای بی‌نشان زبان فارسی محسوب می‌شود و همان‌طور که پیشتر گفتیم این در حالیست که، در مقام تلفظ، غالباً با حذف همخوان پایانی در خوشهٔ دوهمخوانی و هجاسازی مجدد با درج واکه و به دنبال آن تبدیل CVC به CV مثلاً در /kârgar/ به صورت [kâregar]، می‌بینیم که هجای CV هجای مطلوب در زبان فارسی است و

در فرایندهای واجی، هجاهای دیگر به آن تبدیل و یا نزدیک می‌شوند. جدول ۱۶ بسامد هرکدام از ساخت‌های هجایی فارسی را نشان می‌دهد.

جدول ۱۵- بسامد واجی عضو دوم خوشه CC در هجای CVCC

عضو دوم خوشه	بسامد واجی	درصد واجی	عضو دوم خوشه	بسامد واجی	درصد واجی
t	۱۵۵۵	۲۰.۷	š	۲۰۴	۲.۷۲
d	۱۴۶۵	۱۹.۵	b	۱۹۲	۲.۵۶
m	۶۵	۸.۰۵	q	۱۶۸	۲.۲۴
g	۵۷۴	۷.۶۴	f	۱۲۱	۱.۶۱
r	۵۵۰	۷.۳۲	ž	۱۱۱	۱.۴۸
s	۳۵۰	۴.۶۶	h	۹۰	۱.۲
l	۳۱۷	۴.۲۲	x	۸۱	۱.۰۸
j	۲۶۹	۳.۵۸	v	۴۹	۰.۶۵
k	۲۶۰	۳.۴۶	y	۳۴	۰.۴۵
z	۲۴۲	۳.۲۲	p	۲۹	۰.۳۹
n	۲۲۱	۲.۹۴	č	۱۴	۰.۱۹

جدول ۱۶- بسامد هجاهای فارسی

هجا	بسامد	درصد
cv	۱۰۴۲۱۱	۵۷.۷۵
cvc	۶۸۷۲۳	۳۸.۰۹
cvcc	۷۵۱۲	۴.۱۶
جمع	۱۸۰۴۴۶	۱۰۰

۱۰- فراوانی واج‌های زبان فارسی در واژگان زبان

جدول ۱۷ فراوانی واج‌های زبان فارسی در واژگان زبان را نشان می‌دهد. همان‌طور که دیده می‌شود، همخوان‌های /r,n,d,t,m,s/ پس از واکه‌ها بیشترین بسامد را دارند. به طوری‌که بسامد همخوان‌های /r,n/ حتی بیشتر از واکه‌های /o,u/ است.

جدول ۱۷- فراوانی واج‌های زبان فارسی

واج	بسامد واج	درصد بسامد هر واج از کل	واج	بسامد واج	درصد بسامد هر واج از کل
a	۴۲۵۹۷	۹.۵	k	۱۰۲۷۷	۲.۳۳
â	۴۱۵۶۸	۹.۳	h	۱۰۱۲۱	۲.۲
e	۳۵۲۱۶	۷.۹	y	۹۹۵۱	۲.۲
i	۳۲۲۶۲	۷.۲	š	۹۴۸۰	۲.۱
r	۳۰۰۲۵	۶.۷	u	۸۸۳۵	۱.۹
n	۲۵۶۱۳	۵.۷	q	۷۰۱۸	۱.۵
o	۱۹۹۶۸	۴.۴	f	۶۷۲۱	۱.۵
d	۱۹۰۱۶	۴.۲	g	۶۶۳۷	۱.۴
t	۱۸۲۰۲	۴	p	۵۴۰۸	۱.۲
m	۱۷۹۸۸	۴	x	۵۳۲۴	۱.۱
s	۱۵۹۴۲	۳.۵	v	۵۲۴۰	۱.۱
?	۱۴۵۹۲	۳.۲	ǰ	۴۴۶۹	۱
b	۱۳۷۹۶	۳.۱	č	۱۹۴۷	۰.۴
l	۱۳۷۰۴	۳	w	۱۱۴۷	۰.۲
z	۱۰۷۷۱	۲.۴	ž	۸۰۲	۰.۱
جمع کل بسامد واج‌ها = ۴۴۴۶۳۸			جمع کل درصد = ۱۰۰		

۱۱- نتیجه‌گیری

در پژوهش حاضر تلاش کردیم که نظام آوایی زبان فارسی را در آینه آمار بازتاب دهیم، تا برخی ویژگی‌های این نظام طبیعی را روشن کرده باشیم. بررسی آماری اطلاعات واجی موجود در ۵۴۳۹۱ واژه فارسی، به‌عنوان نماینده واژگان ذهنی فارسی‌زبانان، این امکان را به ما داد تا (به دلیل ماهیت آماری پژوهش و بدون هر نوع پیش‌داوری) بتوانیم بسامد، توزیع و رفتارهای نظام‌مند عناصر و واحدهای فعال در نظام آوایی زبان را از طریق آمار توضیح دهیم. پژوهش حاضر، از حیث حجم داده‌ها و جنبه‌های متنوع مورد بررسی و نتایج به‌دست‌آمده، با پژوهش‌های پیشین تفاوت اساسی دارد. نتایج تحلیل‌های آماری نشان می‌دهد که تقریباً تمامی

واحد‌های واژگان اعم از اسم، فعل، صفت، قید و غیره در زبان فارسی تکیه‌پایانی است و هجای آخر آنها برجسته است و اهل زبان نسبت به این امر دانش ناخودآگاه دارند و به همین دلیل نوواژه‌ها و واژه‌های قرضی را با برجسته‌کردن هجای پایانی تلفظ می‌کنند. تحلیل آماری اطلاعات مختلف واجی، مانند تحلیل واژه‌های فارسی از حیث تعداد هجا، زنجیره‌های واجی در ساختمان هجا و آرایش هجا در درون واژه‌های فارسی، بسامد طبقات طبیعی واجی در زبان فارسی فارغ از ساختمان هجا، طبقات واجی در ساختمان هجا، آرایش‌های واجی در خوشه دوهمخوانی، هجای بی‌نشان و بهینه در زبان و فراوانی واج‌های زبان فارسی در واژگان زبان، اطلاعات بسیار مهمی را درخصوص سازوکار نظام آوایی زبان در اختیار پژوهشگران در عرصه‌های مختلف قرار می‌دهد، تا به‌فراخور تخصص و موضوع تحقیق بتوانند از آمار و ارقام ارائه‌شده در این پژوهش استفاده کنند.

نتایج واجی-آماری این تحقیق، از یک سو گواهی بر روایی اصول و مفاهیم پذیرفته‌شده واج‌شناختی مانند طبقه طبیعی واجی، فرایندهای واجی، ساختمان سلسله‌مراتبی هجا و غیره است و از طرف دیگر بسامد واج‌ها و زنجیره‌های واجی را مشخص می‌کند، که در نتیجه آن پیوستار بی‌نشانی و نشان‌داری در نظام آوایی زبان مشخص می‌شود. نیز بررسی نتایج آماری این پژوهش شاهدی بر گرایش‌های عام واجی در زبان‌ها و نیز گرایش‌های زبان-ویژه است که از آنها می‌توان در تبیین فرایندهای واجی، شناسایی زنجیره‌های واجی محتمل و اطلاع از بسامد آنها بهره برد. همچنین از اطلاعات آماری نظام آوایی می‌توان در مطالعات آواشناختی، آواشناسی آزمایشگاهی، مطالعات واج‌شناختی و ساختواژه و نیز در کاربردهای عملی مانند واژه‌گزینی، گفتاردرمانی، فناوری‌های مربوط به پردازش گفتار استفاده کرد.

منابع

- احمدی، مهدی و محمود بی‌جن‌خان. (۱۳۸۹). «خوشه‌های صامت پایانی در فارسی میانه و فارسی نو: قیاسی با سنجۀ محدودیت توالی رسایی». *مجله پژوهش‌های زبان‌شناسی*. س ۲، ش ۲، ۳۵-۱۳.
- اسلامی، محرم، مسعود شریفی آتشگاه، صدیقه علیزاده لمجیری، و طاهره زندی. (۱۳۸۳). «واژگان زایای زبان فارسی». *مجموعه مقالات اولین کارگاه پژوهشی زبان فارسی و رایانه*. تهران: دانشگاه تهران، ۶-۱۲.

نظام آوایی زبان فارسی در آینه آمار / ۸۹

- اسلامی، محرم و محمود بی‌جن‌خان. (۱۳۸۴). «واژه‌گزینی و واج‌آرایی». علی کافی (گردآورنده)، مجموعه مقالات دومین هم‌اندیشی واژه‌گزینی و اصطلاح‌شناسی. تهران: فرهنگستان زبان و ادب فارسی. ۴۴۸-۴۳۰.
- اسلامی، محرم. (۱۳۸۸). «تکیه کلمه در زبان فارسی»، *دوفصل‌نامه پردازش علائم و داده‌ها*. س ۶، ش ۱، پیاپی ۱۱، ۱۲-۳.
- اسلامی، محرم و صدیقه علیزاده لمجیری. (۱۳۸۸). «ساختار تصریفی کلمه در زبان فارسی». *دوفصل‌نامه زبان و ادب فارسی*. نشریه دانشکده ادبیات و علوم انسانی دانشگاه تبریز، سال ۵۲، مسلسل ۲۱۱، ۱۸-۱.
- اسلامی، محرم و سمانه ابهری و صفورا فدایی‌زاده. (۱۳۹۰). «واژه‌گزینی و دانش زبانی». *نامه فرهنگستان*، س ۱۲، ش ۳، پیاپی ۴۷، ۱۸۰-۱۵۹.
- ثمره، بداله. (۱۳۷۸) *آواشناسی زبان فارسی*. ویرایش دوم. تهران: مرکز نشر دانشگاهی.
- سامعی، حسین. (۱۳۷۴). «تکیه فعل در زبان فارسی: بررسی مجدد». *نامه فرهنگستان*. س ۱، ش ۴، پیاپی ۴، ۲۱-۶.
- صدری افشار، غلامحسین، نسرین حکمی و نسترن حکمی. (۱۳۸۱). *فرهنگ معاصر فارسی امروز*. تهران: فرهنگ معاصر.
- فرهنگستان زبان و ادب فارسی. (۱۳۸۳، ۱۳۸۴، ۱۳۸۵، ۱۳۸۶، ۱۳۸۷). *واژه‌های مصوب دفترهای اول تا پنجم*. تهران: فرهنگستان زبان و ادب فارسی.

Alamolhoda, Seyyed Morteza (2000). *Phonostatistics and Phonotactics of the Syllable in Modern Persian*. Helsinki: Finnish Oriental Society.

Ladd, D. Robert (2008). *Intonational Phonology*. second edition. Cambridge: Cambridge University Press.



پښتونستان ښار علمي او مطالعاتي مرکز
پښتونستان ښار علمي او مطالعاتي مرکز