

بررسی میزان تأثیر بسامد بر طول واژه

شهلا شریفی*

الهام ضیایی قوچان عتیق**

چکیده

بسامد در زبان‌شناسی هم‌چون دیگر حوزه‌های علوم مختلف کاربردهایی فراوان دارد و برهمین اساس تعاریفی متنوع برای آن ارائه شده است. آنچه در این پژوهش از مفهوم بسامد برمی‌آید، تعداد دفعات ظهور یک عنصر زبانی در یک متن است و از همین روی مستلزم انجام کارهای آماری است.

در این مقاله نگارندگان بر آنند تا به بررسی «قانون زیف» یعنی رابطه میان بسامد و طول واژه در متون فارسی بپردازند. به همین جهت دو متن قدیمی و امروزی شامل ۱۵۰۰ واژه، انتخاب شده است و تأثیر بسامد بر طول واژگان قاموسی و نقشی مورد بررسی قرار گرفته است که نتیجه بدست آمده، قانون زیف را در زبان فارسی نیز تأیید می‌کند.

کلید واژه

بسامد - قانون زیف.

*استادیار دانش‌گاه فردوسی مشهد.

**کارشناسی ارشد زبان‌شناسی همگانی دانش‌گاه فردوسی مشهد.

مقدمه

بسامد به عنوان معیار سنجش، در شاخه‌ها و روی کردهای مختلف زبان‌شناسی هم‌چون دیگر علوم بکار گرفته شده و بسته به حوزه بکار رفته، دسته‌بندی‌ها و تعاریفی متفاوت از آن ارائه شده است. به عنوان مثال بسامد معیار سنجش زبانی در ساخت واژه، معیار سنجش نشان‌داری عناصر زبانی در معنی‌شناسی و رده‌شناسی است. هم‌چنین در مطالعات تاریخی نیز نقشی قابل توجه دارد. به هر جهت از میان انواع بسامد، در این جستار نوعی از بسامد مورد نظر نگارندگان است که به آن بسامد مورد یا بسامد متنی گفته می‌شود که در واقع به تعداد دفعات ظهور یک عنصر زبانی در یک متن اشاره دارد و بررسی آن مستلزم کارهای آماری است. بدین ترتیب در این مقاله نگارندگان برآنند تا به بررسی رابطه متغیر طول واژه‌ها (قاموسی و نقشی) با بسامدشان بپردازند.

پیشینه پژوهش

انجام کارهای آماری در حوزه‌های مختلف زبانی، منجر به یک رویکرد آماری در زبان شده است که از جمله این موارد - با توجه به ویژگی‌های آوایی و دستوری زبان‌های مختلف - می‌توان به سبک‌شناسی نویسندگان مختلف از جهت بکاربردن طول واژه‌ها و جملات در متن، استفاده از واژگان خاص و... و هم‌چنین پی‌بردن به نسل زبان‌هایی که از یک خانواده منشعب می‌شود، اشاره کرد.^۱ اول کسی که به صورت منسجم در این زمینه به بررسی و پژوهش پرداخته، جرج زیف است که در طول دهه‌های ۱۹۲۰ تا ۱۹۳۰ با هم‌راهی دانش‌جویانش در دانش‌گاه هاروارد، آزمایش‌هایی روی شمارش واژگان مختلف در زبان‌های انگلیسی، لاتین، یونانی، داکوتا و نوتکا (زبان‌های اسکیمویی) و حتی گفتار کودکان در سنین مختلف انجام داد و دریافت که میان بسامد و مرتبه واژه در هر بافت رابطه‌ای مستقیم وجود دارد. در واقع این، اولین قانون زیف بود که پایه‌ای برای انجام پژوهش‌های دیگر در زمینه ارتباط بسامد در حوزه‌های مختلف زبانی شد که از آن جمله می‌توان به بررسی «واگوپلسکی» (۱۹۸۶) در حوزه زبان‌شناسی تاریخی تطبیقی اشاره کرد که تطابق آوایی میان پانزده مفهوم ثابت جهانی را در تغییرات زبان‌شناسی تاریخی مورد بررسی قرار داد.

هم‌چنین زیف در سال ۱۹۳۵ در کتاب روان‌شناسی زبان اظهار داشت که طول واژه با بسامد نسبی آن رابطه معکوس دارد.^۲ این امر توجه بسیاری از پژوهش‌گران از

جمله «ویمر» (۱۹۹۴)، «بست» (۱۹۹۶) و «ویلسون» (۱۹۹۸) را بخود جلب کرد و با انجام تحقیقات گسترده‌تر دریافتند که توزیع بسامد واژگان با تعداد حروف و آواهای آن‌ها مطابقت دارد. «هسیلمت» (۲۰۰۲) نیز از جمله پژوهش‌گرانی است که به بررسی رابطه تأثیر بسامد در واژه‌سازی پرداخته و به رابطه میان طول واژه و بسامد آن اشاره می‌کند و معتقد است که هر چه واژه کوتاه‌تر باشد، بسامد آن بیش‌تر است^۳ «بایی» (۲۰۰۲) نیز این قانون را به گونه‌ای دیگر بیان می‌کند و معتقد است که تولید واژگان پربسامد به صورت خودکار روی می‌دهد. هم‌چنین به اعتقاد وی، تولید واژگان پربسامد از تولید واژگان کم بسامد برای از هم گسستن واژه مشکل‌تر است. به عبارت دیگر، گسستن واژگانی پربسامد مثل حرف تعریف the سخت‌تر از واژگان کم بسامد در یک بافت است.^۴

با توجه به مطالبی که ارائه شد، در این مقاله نگارندگان برآنند تا قوانین و ادعاهای زیف را در زبان فارسی مورد بررسی قرار دهند.

بررسی رابطه میان بسامد و طول واژه

در بخش‌های پیشین اشاره شد که زیف معتقد بود میان بسامد هر واژه و مرتبه آن در بافت، رابطه‌ای مستقیم وجود دارد. با توجه به این ادعاها و اثبات آن‌ها در برخی زبان‌ها، نگارندگان برآنند تا صحت و سقم آن‌ها را در زبان فارسی مورد بررسی قرار دهند. به همین منظور دو متن متفاوت (یک متن تاریخی از کتاب تاریخ بیهقی و یک متن امروزی از کتاب اترک‌نامه) مشتمل بر ۱۵۰۰ واژه، مورد بررسی قرار گرفت که در این متون واژگان قاموسی و نقشی براساس تعداد واج‌های تشکیل دهنده‌شان مورد ارزیابی قرار گرفته است.

بررسی بسامد در متن قدیمی

بررسی رابطه بسامد با واژگان نقشی در متن قدیمی

با توجه به این که شیوه نگارش متون قدیمی با متون امروزی متفاوت است، متنی از کتاب تاریخی «تاریخ بیهقی» مشتمل بر ۱۵۰۰ واژه (شامل واژگان قاموسی و نقشی ساده و مرکب) انتخاب شد و براساس تعداد واج‌ها مورد بررسی قرار گرفت و در

نهایت، بسامد آن‌ها براساس فرمول ریاضی $F = \frac{n}{N} \times 100$ محاسبه شد که در آن: بسامد = F، تعداد واج‌ها = n، تعداد کل واژه‌ها = N است.

نتایج بدست آمده در جدول‌های زیر ارائه شده است:

بسامد به درصد	تعداد	واژگان نقشی بسیط (حروف اضافه)
۲/۴۶۶	۳۷	از
۱/۴۶۶	۲۲	در
۱/۰۶۶	۱۶	به
۰/۹۳۳	۱۴	تا
۰/۸۶۶	۱۳	بر
۰/۱۳۳	۲	با

جدول ۱ - بسامد حروف اضافه بسیط در متن قدیمی

بسامد به درصد	تعداد	واژگان نقشی مرکب (حروف اضافه)
۰/۱۳۳	۱	با آن‌که
۰/۰۶۶	۱	بدان
۰/۰۶۶	۱	بدین

جدول ۲ - بسامد حروف اضافه مرکب در متن قدیمی

بسامد به درصد	تعداد	واژگان نقشی بسیط (حروف ربط)
۱۰/۴	۱۵۶	و
۳/۶۶۶	۵۵	که
۱/۷۳۳	۲۶	را
۰/۶	۹	چون
۰/۵۳۳	۸	چه
۰/۴	۶	یا
۰/۱۳۳	۲	هم
۰/۱۳۳	۲	اگر
۰/۰۶۶	۱	اما

جدول ۳ - بسامد حروف ربط بسیط در متن قدیمی

بسامد به درصد	تعداد	واژگان نقشی مرکب (حروف ربط)
۰/۴۶۶	۷	چنان‌که
۰/۴۶۶	۷	آن‌چه
۰/۲۶۶	۴	هرچند
۰/۲	۳	همچنین
۰/۱۳۳	۲	هرچه
۰/۰۶۶	۱	هم‌چون
۰/۰۶۶	۱	آن‌که
۰/۰۶۶	۱	اگرچه

جدول ۴ - بسامد حروف ربط مرکب در متن قدیمی

همان‌طور که از مشاهده نتایج بدست آمده در جدول‌های بالا برمی‌آید، بسامد واژه‌های نقشی (حروف ربط و حروف اضافه) ساده و مرکب تقریباً با طول آن‌ها رابطه عکس دارد، بطوری که اگر واژگان نقشی ساده و مرکب را با هم مقایسه کنیم، این نتیجه محسوس‌تر است. بنابراین می‌توان گفت هرچه واژه‌ای، بسامدی بیش‌تر داشته باشد کوتاه‌تر است، اما بنظر می‌رسد این قانون در مورد واژه‌هایی که کمتر از سه واج دارد، مصداق پیدا نمی‌کند، بطوری که برخی واژه‌ها از لحاظ تعداد واج‌ها یک‌سان، اما بسامد وقوعشان با هم متفاوت است.

بررسی رابطه بسامد با واژگان قاموسی در متن قدیمی

با توجه به این که قانون زیف در زبان فارسی در مورد واژه‌هایی که کمتر از سه واج دارد صدق نمی‌کند، نگارنده بر آن است تا به بررسی رابطه بسامد با واژگان قاموسی بپردازد. بر همین اساس واژگان قاموسی براساس واج‌های تشکیل دهنده از ۲ تا ۹ واج تقسیم‌بندی شده است که نتایج در جدول زیر قابل دسترسی است:

تعداد واج‌ها	میزان تکرار (بسامد)	بسامد به درصد
۲	۲۰	۰/۱۳۳
۳	۵۸۰	۳۸/۶۶۶
۴	۵۳۵	۳۵/۶۶۶
۵	۲۱۳	۱۴/۲
۶	۸۶	۵/۷۳۳
۷	۳۰	۲
۸	۲۴	۱/۶
۹	۱۲	۰/۸

جدول ۵ - بسامد واژگان قاموسی در متن قدیمی

همان‌طور که مشاهده می‌شود واژگانی که مشتمل بر ۳ واج است، بالاترین میزان بسامد را دارد. به همین ترتیب هرچه بر تعداد واج‌های واژه افزوده می‌شود و یا به عبارت دیگر هر چه طول واژه‌ای بیش‌تر می‌گردد، از میزان بسامدش کاسته می‌شود که البته در مورد واژه‌هایی که ۲ واج دارد استثنایی وجود دارد که همان‌طور که گفته شد، قانون زیف در مورد واژه‌هایی که کمتر از ۳ واج دارد مصداق پیدا نمی‌کند؛ یعنی هر چند که واژه‌های مشتمل بر ۲ واج از نظر طول کوتاه‌تر است، اما تعداد دفعات ظهورشان در متن کمتر است که خود دلیلی محکم برای اثبات قانون زیف است.

بررسی بسامد در متن امروزی

بررسی رابطه بسامد با واژگان نقشی در متن امروزی

به منظور دستیابی به اطمینان بیش‌تر از نتایج بدست آمده، علاوه بر متن قدیمی، یک متن امروزی نیز از کتاب «اترک‌نامه» انتخاب شده است که شامل ۱۵۰۰ واژه قاموسی و نقشی است و نتایج بدست آمده از آن در جدول‌های زیر قابل دسترسی است:

بسامد به درصد	تعداد	واژگان نقشی بسط (حروف اضافه)
۳/۶	۵۴	در
۳/۴	۵۲	به
۳/۲	۴۸	از
۰/۸	۱۲	با
۰/۵۳۳	۸	برای
۰/۵۳۳	۸	تا
۰/۲۶۶	۴	بر
۰/۰۶۶	۱	زیرا

جدول ۶ - بسامد حروف اضافه بسیط در متن امروزی

بسامد به درصد	تعداد	واژگان نقشی مرکب (حروف اضافه)
۰/۲۶۶	۴	به محض
۰/۲	۳	درباره
۰/۱۳۳	۲	پیش از
۰/۱۳۳	۲	در اطراف

جدول ۷ - بسامد حروف اضافه مرکب در متن امروزی

بسامد به درصد	تعداد	واژگان نقشی بسط (حروف ربط)
۹/۲	۱۳۸	و
۲/۹۳۳	۴۴	که
۲/۴۶۶	۳۷	را
۰/۶۶۶	۱۰	یا
۰/۲	۳	ولی

جدول ۸ - بسامد حروف ربط ساده (بسیط) در متن امروزی

بسامد به درصد	تعداد	واژگان نقشی بسط (حروف ربط)
۰/۴۶۴	۷	همچنین
۰/۱۳۳	۲	آن‌که
۰/۰۶۶	۱	گرچه

جدول ۹ - بسامد حروف ربط مرکب در متن امروزی

همان‌طور که مشاهده می‌شود بسامد حروف ربط و حروف اضافه بسیط (ساده) نسبت به حروف ربط و حروف اضافه مرکب بیش‌تر است که در واقع با همان قانون زیف در رابطه با طول واژه مطابقت می‌کند؛ یعنی هرچه بر طول واژه افزوده می‌شود، از تعداد دفعات ظهورش در متن کاسته می‌گردد. به عبارتی در متن امروزی نیز مانند متن قدیمی طول واژه با بسامدش رابطه مستقیم دارد و این در حالی است که همان‌طور که زیف نیز اشاره کرده بود، این قانون در مورد واژه‌های کمتر از سه واج صدق نمی‌کند و شاید علت وجود بی‌نظمی در بسامد واژگان نقشی بسیط نیز به همین خاطر باشد.

بررسی رابطه بسامد با واژگان قاموسی در متن امروزی

به منظور دستیابی به نتایج دقیق‌تر، همان‌روندی که در مورد متن قدیمی اعمال شد، در مورد متن امروزی نیز صورت گرفت، به عبارت دیگر واژگان قاموسی براساس واج‌های تشکیل دهنده به واژگان ۲ تا ۹ واج دسته‌بندی شد که نتایج بدست آمده در جدول زیر قابل دسترسی است:

تعداد واج‌ها	میزان تکرار (بسامد)	بسامد به درصد
۲	۴۵	۳
۳	۶۳۵	۴۲/۳۳۳
۴	۴۲۰	۲۸
۵	۱۶۲	۱۰/۸
۶	۱۰۴	۶/۹۳۳
۷	۷۵	۵
۸	۵۴	۳/۶
۹	۵	۰/۳۳۳

جدول ۱۰ - بسامد واژگان قاموسی در متن امروزی

از مشاهده نتایج بدست آمده در این جدول نیز چنین استنباط می‌شود که بسامد با طول واژه رابطه معکوس دارد، بطوری که هر چه بر میزان واج‌های هر واژه افزوده می‌شود، از میزان بسامد آن کاسته می‌گردد که در مورد واژه‌های کمتر از ۳ واج نیز استثنا وجود دارد که هر چند تعداد واج‌هایشان کمتر است، اما بسامدشان نیز در متن کمتر است.

نتیجه‌گیری

با توجه به داده‌های آماری بدست آمده از متن قدیمی (تاریخ بیهقی) و متن امروزی (اترک‌نامه) در پژوهش حاضر چنین استنباط می‌شود که قانون زیف که رابطه معکوس بین طول واژه و بسامد آن را نشان می‌دهد، در زبان فارسی نیز مصداق دارد. همان‌طور که مشاهده شد طول هر واژه با تعداد واج‌های تشکیل دهنده‌اش سنجیده می‌شود که هر چه بر تعداد واج‌های تشکیل دهنده یک واژه افزوده می‌شود، یعنی واژه طولانی‌تر می‌شود، از میزان بسامد یا همان تعداد دفعات تکرار واژه در بافت کاسته می‌شود که طبق استثنایی که در قانون زیف وجود دارد (واژه‌های کوتاه‌تر از سه واج را در بر می‌گیرد) داده‌های زبان فارسی نیز همین استثنا را تأیید می‌کند. بدین ترتیب می‌توان دریافت که زبان فارسی نیز مانند زبان انگلیسی و زبان سوئدی که توسط زیف مورد بررسی قرار گرفته است، تأیید کننده قانون زیف و استثنای آن است.

پی‌نوشت‌ها

۱. مالمکیر، ۲۰۰۴: ۷۴.
۲. سیگارد و سایرین، ۲۰۰۷.
۳. هسپلمت، ۲۰۰۲: ۲۴۳.
۴. بایبی، ۲۰۰۲: ۸۶-۹۵.

کتاب‌نامه

- بیهقی، ابوالفضل محمدبن حسین (۱۳۷۵)، تاریخ بیهقی، ج ۳، به کوشش خلیل خطیب رهبر، تهران: مهتاب.
- شاکری، رمضان‌علی (۱۳۸۱)، ترک‌نامه تاریخ جامع قوچان، تهران: امیرکبیر.
- Bybee, j. (2002) , Morphology: A Study of the Relation between Meaning and Form, Amsterdam: John Bergamins.
- Haspelmath, Martin, (2002) , Understanding Morphology, Londen: Arnold,
- Malmkyaer, Kristen (2004) , The linguistic Encyclopedia, NewYork: New York Publisher
- Sigurd, Bength and the others (2007) , Word Length, Sentence length and frequency _ Zipf Revisited.
available at: [http://person.sol.lu.se/joost van De weijer/Texts/studing. pdf](http://person.sol.lu.se/joost%20van%20De%20weijer/Texts/studing.pdf).