



رابطه میان فراوانی واژگان تک کلمه‌ای قرآن و طول آنها بر حسب حرف (بررسی قانون اول زیف در قرآن کریم)

ولی جهانگیری^{۱*}

محمدجواد آزادی^۲

محمدنبی احمدی^۳

چکیده

بررسی رابطه بین فراوانی واژگان تک کلمه‌ای قرآن، و طول آنها بر حسب تعداد حروفشان؛ نشان از وجود رابطه‌ای عکس بین این دو دارد. به عبارت دیگر کلمات کوتاه‌تر قرآن، نسبت به کلمات بلندتر دارای فراوانی بیشتری هستند. این رابطه که به قانون زیف (۱۹۳۶) معروف است، علاوه بر اینکه در کل قرآن صدق می‌کند در ۲۶ سوره قرآن نیز صادق است. نتیجه بررسی این قانون در زبان‌های مختلف نشان داده است این قانون در هر زبانی در محدوده‌ی خاصی کاربرد دارد. در قرآن کریم نیز این قانون برای کلمات کوتاه‌تر از ۴ حرف صادق نیست. با استفاده از فرمول $F\% = \exp(a + b * L + c) * L^2$ می‌توان قانون زیف را در دو سطح: کل قرآن و در هریک از ۲۶ سوره مزبور، به زبان ریاضی قالب‌بندی کرد.

کلیدواژه‌ها:

قانون زیف، نمودار زیف، واژگان قرآنی، زبان‌شناسی آماری، توزیع فراوانی کلمات

✉ | valijahangiry@yahoo.com

۱- دانشجوی دوره دکتری علوم قرآن و حدیث دانشگاه فردوسی

مشهد* (نویسنده مسئول)

✉ | azadi_ravagh@yahoo.com

۲- دانشجوی دوره دکتری علوم قرآن و حدیث دانشگاه کاشان

✉ | mn.ahmadi217@yahoo.com

۳- استادیار گروه زبان و ادبیات عرب دانشگاه رازی کرمانشاه

مقدمه

قانون پراکندگی زیف، قانون تجربی است که به نام استاد زبان شناسی دانشگاه هاروارد جورج کینگسلی زیف^۱ (۱۹۵۰-۱۹۰۲) نامگذاری شده است؛ زیف قانونش در رابطه با اختصارات^۲ را این گونه معرفی می‌کند: «با توجه به شواهد، طول یک کلمه با فراوانی نسبی آن رابطه‌ی عکس دارد» (زیف، ۱۹۳۶: ۳۶). یعنی کلمات کوتاه‌تر بسامدشان بیشتر از کلمات طولانی‌تر است؛ زیف «صرفه جویی در زمان و تلاش» (همو) را علت این پدیده می‌داند، البته علت دیگری که مطرح شده این است که هر چقدر یک کلمه بیشتر به کار رود، آن کلمه بیشتر کوتاه‌تر می‌شود (روسو^۳، ۲۰۰۲: ۱۴).

هر چند زیف اولین کسی نبوده که این قانون را کشف کرده است؛ اما مطمئناً وی تنها شخصی است که بیشترین فعالیت را در توضیح این پدیده و توسعه فرضیاتش به دیگر رشته‌های علمی، مخصوصاً جامعه‌شناسی داشته است (پرن^۴ و زیف^۵، ۲۰۰۲: ۶). به طوری که گفته می‌شود اهمیت زیف در علم زبان‌شناسی همانند اهمیت نیوتن در علم فیزیک است (آلمن^۶، ۲۰۰۲: ۲۵). با توجه به اهمیت این قانون، تاکنون تحقیقات متعددی در این باره انجام شده است و میزان کارایی آن در متون متعددی از میان زبان‌های مختلف مورد بررسی قرار گرفته است.^۷

قرآن کریم به‌عنوان کتابی مقدّس، نزد مسلمانان از جایگاه ویژه‌ای برخوردار است، از آنجایی که منبع این کتاب، مافوق بشری و الهی است می‌تواند گزینه مهمّی جهت انجام پژوهش‌های زبان‌شناختی نظیر قانون زیف باشد؛ و با توجه به اینکه تاکنون چنین پژوهشی در مورد متن قرآن کریم صورت نگرفته است، اهمّیت این پژوهش بیشتر مشخص می‌شود.

در مرحله اول این پژوهش، رابطه بین طول واحدهای زبانی تک کلمه‌ای و فراوانی آنها را در کلّ قرآن، به‌عنوان متنی واحد و یکپارچه؛ در مرحله بعد در سوره‌های مختلف قرآن، به صورت مجزا و مستقل، مورد بررسی قرار خواهد گرفت.

از آنجایی که داده‌های مورد نیاز این تحقیق قبلاً توسط کسی گردآوری نشده است، لذا کلیه عملیات محاسبه طول واژگان قرآن و فراوانی هریک از آنها، برای اولین بار در این تحقیق صورت می‌گیرد. برای این کار با استفاده از نرم افزار اکسل و با رعایت نکات زیر متن قرآن مورد تجزیه و تحلیل قرار گرفت:

1. George Kingsley Zipf
2. Abbreviation
3. Ronald Rousseau
4. Claudia Prün
5. Robert Zipf
6. Gabriel Altmann

۷. جهت اطلاع از کارهای علمی که در خصوص قانون زیف انجام گرفته، مراجعه شود به مقاله:

در پژوهش حاضر، طول واحدهای زبانی یک کلمه‌ای^۱ از طریق شمارش تعداد حروف آنها به دست می‌آید و ملاک تمایز واحدهای زبانی، شکل ظاهری آنها (بر اساس رسم الخط عثمان طه و قرائت حفص از عاصم) می‌باشد؛ بر این اساس عبارت «بَنِي إِسْرَائِيلَ» (بقره/۴۰) دو کلمه، کلماتی از قبیل «رَحْمَن» شش حرف (بدون احتساب الف مدی روی حرف میم)، حروف مشدّد یک حرف و الف پایه تنوین نصب در شمارش کلمات نیز به عنوان یک حرف مستقل به حساب آمده است. همچنین، کلماتی که به واو جمع ختم می‌شوند از قبیل «كَفَرُوا» (بقره/۶) الف آخرشان نیز به عنوان یک حرف در نظر گرفته شده است؛ آیه بسمله در ابتدای تمام سوره‌های قرآن (غیر از سوره توبه) به عنوان یک آیه مستقل محسوب شده است؛ و همزه استفهام (أ) فقط در حالتی که پیش از واو قرار گرفته باشد به عنوان یک حرف به حساب می‌آید.

فراوانی و طول واژگان در کل قرآن

در جدول شماره یک، فراوانی واژگان قرآن براساس طولشان ارائه شده است. این داده‌ها برحسب درصد فراوانی نسبی تنظیم شده‌اند. درصد فراوانی نسبی که کار مقایسه مستقیم فراوانی‌ها را آسان می‌کند از طریق رابطه زیر محاسبه می‌شود:

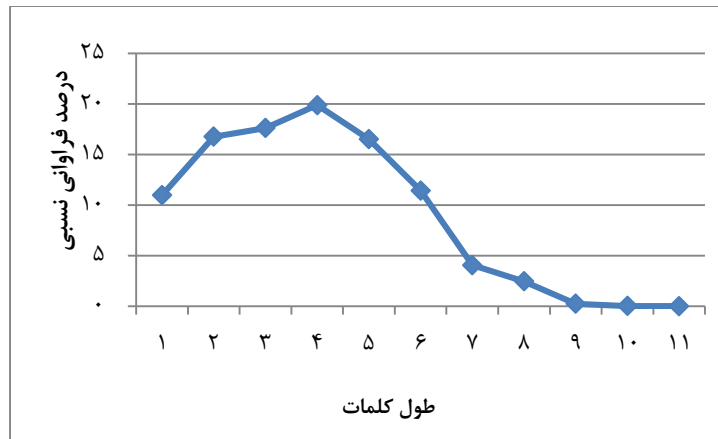
$$F\% = (F_i / N) \times 100$$

(F_i فراوانی هر واژه؛ N تعداد کل فراوانی واژه‌ها)

جدول ۱: طول واژگان و درصد فراوانی نسبی آنها در کل قرآن

طول کلمات	یک	دو	سه	چهار	پنج	شش	هفت	هشت	نه	ده	یازده
درصد فراوانی نسبی	۱۰,۹۷۸	۱۶,۷۶۹	۱۷,۶۲۵	۱۹,۸۸۶	۱۶,۵۲	۱۱,۴۲۸	۴,۰۶۱	۲,۴۶۶	۰,۲۴۳	۰,۰۲۲	۰,۰۰۱

همان طور که از جدول شماره یک بر می‌آید، در کلمات چهار حرفی و بیشتر، هر چه طول کلمات بیشتر می‌شود درصد فراوانی نسبی آنها نیز کم می‌شود. این بدان معناست که قانون زیف در قرآن برای کلمات چهار حرفی و بیشتر صادق است. درک این مطلب به کمک نمودار راحت‌تر صورت می‌گیرد، لذا در نمودار شماره ۱ درصد فراوانی نسبی کلمات بر روی محور عمودی و طول کلمات بر حسب تعداد حروفشان، بر روی محور افقی قرار گرفته است. همان طور که نمودار نشان می‌دهد، بیشترین فراوانی متعلق به کلمات ۴ حرفی است و کلمات کوتاه‌تر یا بلندتر از ۴ حرف، فراوانی کم‌تری دارند.



نمودار ۱: درصد فراوانی نسبی کلمات بر اساس طولشان در کل قرآن

باید توجه داشت که زیف قانون خویش را بدون هیچ‌گونه استثنایی ارائه داده است، یعنی مطابق نظر وی نمودار باید به صورت یک خط مستقیم باشد که در آن کلمات تک حرفی دارای بیشترین فراوانی باشند؛ اما نمودار شماره ۱ نشان می‌دهد که کلمات قرآنی که کوتاه‌تر از چهار حرف هستند از این قانون تبعیت نمی‌کنند. همچنین سیگارد و همکارانش^۱ (۲۰۰۴: ۱-۵) اثبات نموده‌اند که در زبان‌های انگلیسی و سوئدی قانون زیف در کلمات کوتاه‌تر از سه حرف صدق نمی‌کند؛ بنابراین، می‌توان گفت که قانون زیف در هر زبانی فقط در یک بازه‌ی خاص از طول کلمات صدق می‌کند.

تعیین فراوانی کلمات بر اساس طولشان در کل قرآن

نمایش داده‌ها از طریق نمودار کار مقایسه را آسان می‌کند اما برای پیش‌بینی پذیر کردن داده‌ها، فرمول‌های ریاضی بهترین وسیله است؛ لذا در این قسمت به بررسی فرمولی می‌پردازیم که بتواند فراوانی کلمات را بر اساس طولشان در کل قرآن با دقت قابل توجهی پیش‌بینی نماید.

زیف برای به‌دست آوردن فراوانی واژگان از طریق طولشان فرمول خاصی ارائه نکرده است، اما این مسئله مورد توجه تعداد زیادی از محققان قرار گرفته است؛ که هر یک فرمولی را ارائه داده‌اند؛ از این میان، فرمول ارائه شده توسط سیگارد و همکارانش از دقت بیشتری برخوردار است.

فرمول پیشنهادی سیگارد و همکارانش عبارت است از:

$$F\% = a * L^b * c^L \quad (0 \leq C < 1) \quad (1)$$

در این فرمول $F\%$ درصد فراوانی نسبی و L معرف طول واژه و متغیرهای a , b , c اعداد ثابت هستند.

1. Sigurd Bengt, Eeg-Olofsson Mats & Weijer, Joost van de

سیگارد و همکارانش فرمول مذکور را پس از محاسبه مقادیر متغیرهای سه گانه در زبان‌های انگلیسی و سوئدی، این چنین بیان می‌کنند: $F\% = 11.74 * L^3 * 0.4^L$

از آنجایی که فرمول فوق بر اساس یک تحلیل تجربی و با توجه به داده‌های مربوط به زبان‌های انگلیسی و سوئدی به دست آمده است (سیگارد، ۲۰۰۴: ۶)، پس احتمالاً دقت لازم را در متن قرآن نخواهد داشت؛ برای یافتن فرمول مورد نظر، داده‌های جدول شماره یک را با استفاده از نرم افزار Data Fit مورد تحلیل قرار داده و در نهایت به فرمول زیر دست یافتیم:

$$F\% = \exp(a+b*L + c*L^2) \quad (۲)$$

پس از محاسبه متغیرهای سه گانه‌ی a, b, c با توجه به داده‌های مذکور، فرمول فوق بدین صورت در می‌آید:

$$F\% = \exp(1.75 + 0.71*L - 0.1*L^2) \quad (۳)$$

حال، به مقایسه کارایی فرمول پیشنهادی در این پژوهش و فرمول ارائه شده توسط سیگارد و همکارانش خواهیم پرداخت تا معلوم شود کدام یک از این دو فرمول، داده‌های جدول شماره یک را بهتر پیش‌بینی می‌کند؛ برای این کار ابتدا باید متغیرهای موجود در فرمول سیگارد و همکارانش را با توجه به داده‌های جدول شماره یک بدست آورد، زیرا سیگارد و همکارانش تصریح دارند که مقدار این متغیرها به متن مورد بررسی بستگی دارد (همو). محاسبه این متغیرها با استفاده از نرم افزار مذکور به فرمول زیر منجر شد:

$$F\% = 18.14 * L^{2.14} * 0.47^L$$

حال، باید دید کدام فرمول، داده‌های جدول شماره یک را بهتر پیش‌بینی می‌کند. برای این منظور از ضریب تشخیص^۱ که با علامت R^2 نمایش داده می‌شود، استفاده می‌کنیم.

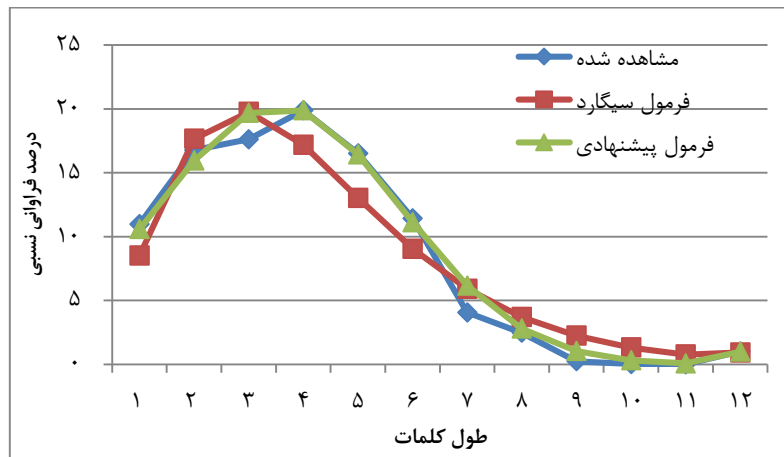
ضریب تشخیص، میزان احتمال هم بستگی میان دو دسته داده در آینده را نشان می‌دهد. این ضریب در واقع نتایج تقریبی پارامتر مورد نظر را در آینده، بر اساس مدل ریاضی تعریف شده، که منطبق بر داده‌های موجود است، بیان می‌دارد. (در این تحقیق ضریب تشخیص، با استفاده از تابع RSQ در نرم افزار Excel محاسبه می‌شود.) مقدار این ضریب در محدوده ۰ تا ۱ قرار دارد و در حالتی که R^2 مساوی یک باشد نشان دهنده هم بستگی کامل بین دو دسته داده است.

همان‌طور که جدول شماره دو نشان می‌دهد، مقدار R^2 در فرمول پیشنهادی بیشتر از میزان آن در فرمول سیگارد و همکارانش می‌باشد. پس می‌توان گفت، فرمول پیشنهاد شده در پژوهش حاضر، بهتر از فرمول سیگارد و همکارانش درصد فراوانی نسبی کلمات را بر اساس طولشان، در کل قرآن پیش‌بینی می‌کند. این مطلب در نمودار شماره دو نیز قابل مشاهده است.

1. Coefficient of Multiple Determination

جدول ۲: درصد فراوانی نسبی کلمات هم طول در کل قرآن بر اساس فرمول پیشنهادی و فرمول سیگارد و مقایسه آن با داده‌های مشاهده شده

طول کلمات	درصد فراوانی نسبی		
	مشاهده شده	بر اساس فرمول سیگارد و همکارانش	بر اساس فرمول پیشنهادی
۱	۱۰,۹۷۸	۸,۵۲۶	۱۰,۵۹۱
۲	۱۶,۷۶۹	۱۷,۶۶۲	۱۵,۹۵۹
۳	۱۷,۶۲۵	۱۹,۷۶۸	۱۹,۶۸۸
۴	۱۹,۸۸۶	۱۷,۱۹۶	۱۹,۸۸۶
۵	۱۶,۵۲	۱۳,۰۲۹	۱۶,۴۴۵
۶	۱۱,۴۲۸	۹,۰۴۶	۱۱,۱۳۴
۷	۴,۰۶۱	۵,۹۱۳	۶,۱۷۲
۸	۲,۴۶۶	۳,۶۹۹	۲,۸۰۱
۹	۰,۲۴۳	۲,۲۳۷	۱,۰۴۱
۱۰	۰,۰۲۲	۱,۳۱۷	۰,۳۱۷
۱۱	۰,۰۰۱	۰,۷۵۹	۰,۰۷۹
R ²	۱	۰,۹۳	۰,۹۹



نمودار ۲: درصد فراوانی نسبی کلمات هم طول در کل قرآن بر اساس فرمول پیشنهادی و فرمول سیگارد و مقایسه آنها با داده‌های مشاهده شده در کل قرآن

فراوانی و طول واژگان در هر یک از سوره قرآن کریم

پس از اینکه مشخص شد قانون زیف در کل قرآن برای واژه‌های بزرگتر از ۴ حرف صادق است، در این مرحله قانون زیف را در تک تک سوره قرآن مورد بررسی قرار می‌دهیم. برای این کار ابتدا نمودار قانون زیف را برای ۱۱۴ سوره قرآن رسم نمودیم، همان طوری که انتظار می‌رفت هیچ یک از این نمودارها به صورت خط

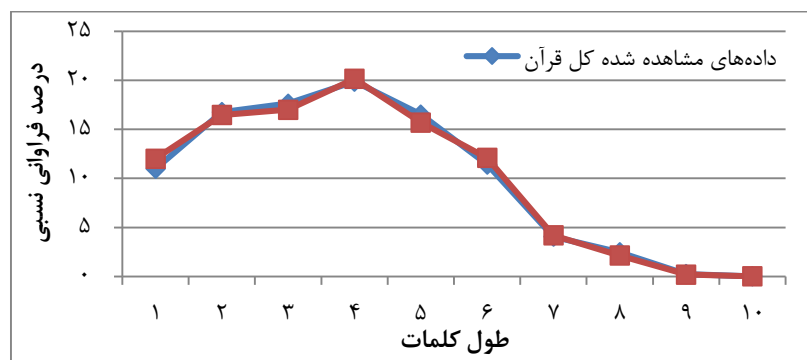
مستقیم نبود. این مطلب نشان می‌دهد که اگر قانون زیف در برخی سور صادق باشد همانند نمودار کل قرآن، با استثنائاتی همراه خواهد بود؛ لذا به بررسی هر یک از این نمودارها می‌پردازیم، مبنای ما برای این کار دو مجموعه داده است:

الف) داده‌های مشاهده شده در کل قرآن؛ ب) داده‌های پیش‌بینی شده کل قرآن از طریق فرمول پیشنهادی.

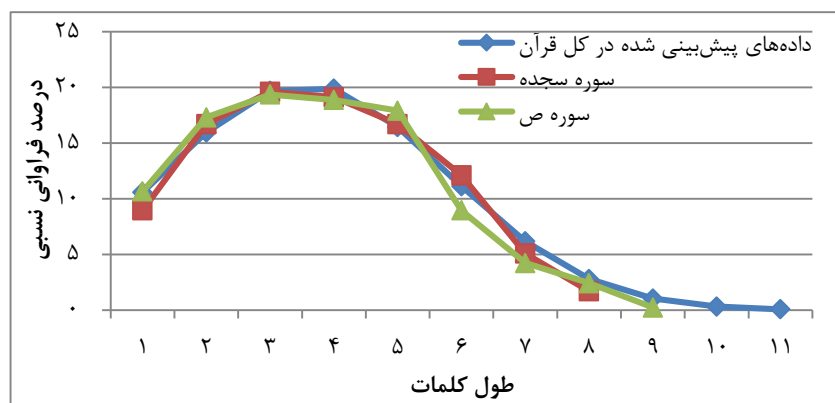
مقایسه نمودار ۱۱۴ سوره با این دو دسته داده به نتایج زیر منجر شد:

۱. همان طوری که در نمودار شماره ۳ مشاهده می‌شود، نمودار سوره بقره بیشترین هماهنگی را با نمودار داده‌های مشاهده کل قرآن دارد.

۲. مطابق نمودار شماره ۴، نمودار سوره (۳۲) سجده بیشترین هماهنگی را با نمودار داده‌های پیش‌بینی شده کل قرآن دارد و بعد از آن سوره (۳۸) ص قرار می‌گیرد.



نمودار ۳: داده‌های مشاهده شده کل قرآن و سوره بقره



نمودار ۴: داده‌های پیش‌بینی شده کل قرآن و سوره‌های سجده و ص

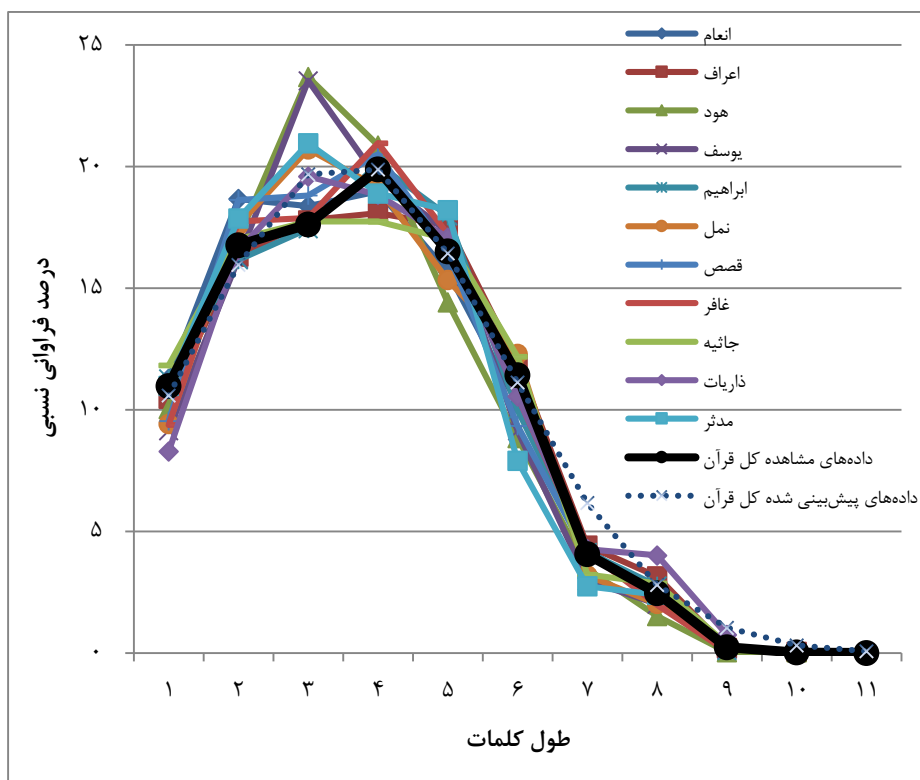
۳. دو دسته سوره زیر به ترتیب دارای بیشترین هماهنگی با هر یک از داده‌های مشاهده شده و پیش‌بینی شده هستند:

دسته سوره اول:

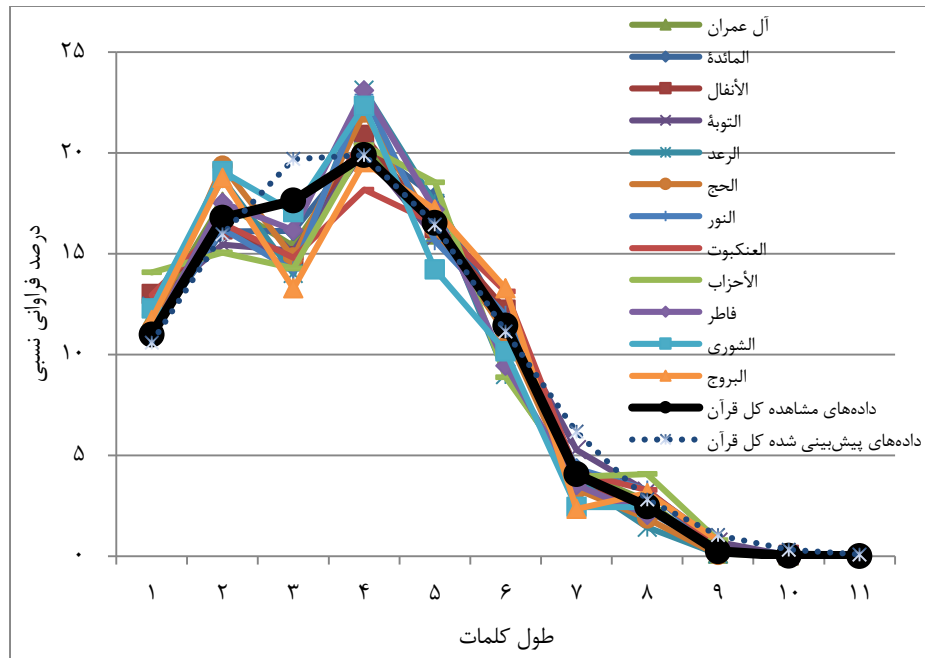
(۴۰) غافر، (۱۴) ابراهیم، (۷) الأعراف، (۶) الأنعام، (۴۵) الجاثیة، (۲۸) القصص، (۲۷) النمل، (۵۱) الذاریات، (۷۴) المدثر، (۱۱) هود، (۱۲) یوسف.

دسته سوره دوم:

(۳) آل عمران، (۵) المائدة، (۶) التوبة، (۸) الأنفال، (۲۲) الحج، (۳۵) فاطر، (۲۹) العنكبوت، (۲۴) النور، (۴۲) الشوری، (۱۳) الرعد، (۸۵) البروج، (۳۳) الأحزاب.



نمودار ۵: دسته سوره اول و داده‌های پیش‌بینی شده و مشاهده شده کل قرآن



نمودار ۶: دسته سور دوم و داده‌های پیش‌بینی شده و مشاهده شده کل قرآن

با توجه به نمودارهای شماره ۵ و ۶ تفاوت این دو دسته سور در این است، که دسته سور اول دارای یک نقطه عطف است، که این نقطه عمدتاً در کلمات چهار حرفی واقع شده است؛ در حالی دسته دوم دارای دو نقطه عطف بوده که در کلمات چهار حرفی و دو حرفی واقع شده است. همچنین بی‌نظمی‌های مشاهده شده در بازه چهار حرفی تا یازده حرفی، در دسته اول کمتر از دسته دوم می‌باشد.

با توجه به اینکه از یک طرف نمودارهای این دو دسته سور در کلمات ۴ حرف و بیشتر، اختلاف قابل توجهی با یکدیگر ندارند و از طرف دیگر همان‌طور که در قسمت پیشین مقاله بیان شد قانون زیف در کل قرآن برای کلمات ۴ حرفی و بیشتر صادق است، پس تفاوت بین این دو دسته در بازه کلمات ۱ تا ۳ حرفی خللی بر صدق قانون زیف وارد نمی‌کند؛ پس می‌توان گفت قانون زیف در این ۲۶ سوره همانند کل قرآن صادق است.

تعیین فراوانی کلمات بر اساس طولشان در دو دسته سور مذکور

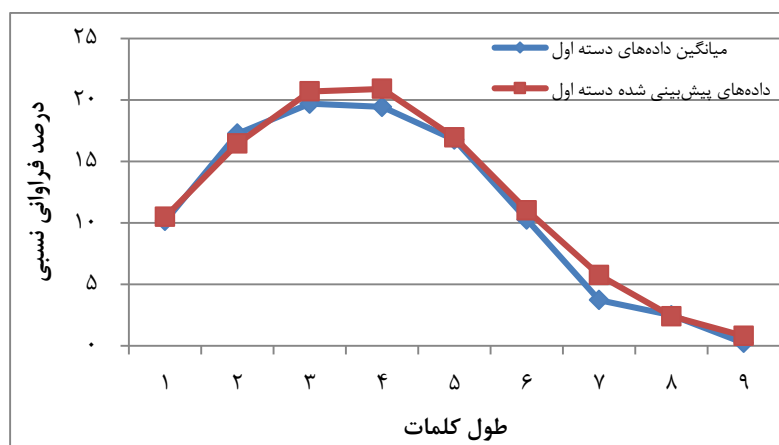
فرمول شماره ۳ فقط در کل قرآن کاربرد دارد، پس برای هر یک از دسته سوره‌های دوگانه فوق‌الذکر یک فرمول جداگانه مورد نیاز است. برای این کار هم می‌توان داده‌های هر یک از دسته سور را به منظور یافتن

یک فرمول جدید مورد تجزیه و تحلیل قرار داد و هم می‌توان با بروز رسانی فرمول شماره ۲ مقدار متغیرهای سه گانه موجود در آن فرمول را محاسبه نمود؛ ما در این تحقیق روش دوم را انتخاب می‌کنیم. از آنجایی که هر یک از دسته‌ها دارای سوره‌هایی با درصدهای فراوانی نسبی متفاوتی هستند، جهت ارائه یک فرمول واحد برای هر یک از این دسته‌ها، ابتدا میانگین درصد فراوانی‌های نسبی را برای کلمات هم طول در هر یک از این دسته‌ها محاسبه کرده و سپس با استفاده از داده‌های حاصل از مرحله پیشین در نرم‌افزار Data Fit، متغیرهای سه‌گانه موجود در فرمول شماره ۲ را محاسبه می‌کنیم. در جدول زیر مقدار متغیرهای سه گانه در فرمول‌های مربوط به هریک از دسته‌ها جاگذاری شده است. مقدار R^2 نیز نشان دهنده میزان همبستگی بین نتایج حاصل از فرمول‌های پیشنهادی و میانگین داده‌های هر یک از دسته سور است:

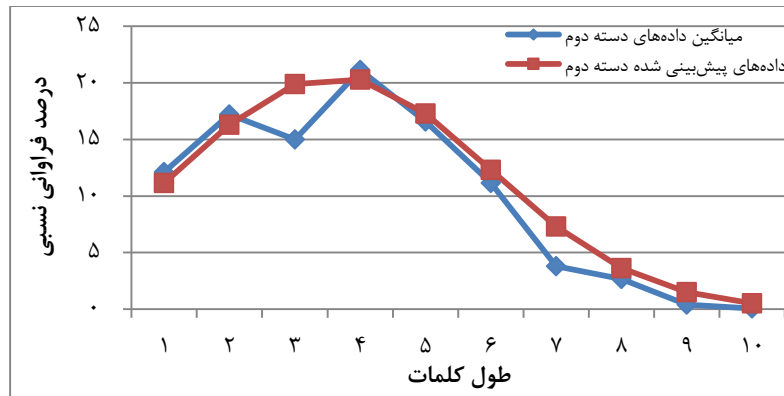
جدول ۳: فرمول‌بندی دسته سور دوگانه

شماره فرمول	شماره دسته سور	فرمول	R^2
۴	یک	$\exp(1.68 + 0.78*L - 0.11*L^2)$	۰,۹۹
۵	دو	$\exp(1.87 + 0.64*L - 0.09*L^2)$	۰,۹۴

همان‌طور که مشاهده می‌شود، میزان R^2 در فرمول شماره ۵ کم‌تر از دیگری است، با توجه به اینکه دسته سور دوم نسبت به دسته دیگر، در کلمات ۳ حرفی دارای بی‌نظمی‌های بیشتری است، پس کاهش میزان R^2 کاملاً طبیعی است. همان‌طور که نمودارهای شماره ۷ و ۸ نشان می‌دهد میزان R^2 مربوط به هریک از این دو فرمول، در حد قابل قبولی است.



نمودار ۷: مقایسه میانگین داده‌های مشاهده شده با داده‌های پیش‌بینی شده در دسته سور اول



نمودار ۸: مقایسه میانگین داده‌های مشاهده شده با داده‌های پیش‌بینی شده در دسته سور دوم

نتیجه‌گیری

در تحقیق حاضر، مشخص شد رابطه عکس بین فراوانی واژگان قرآنی و طول آنها، که به قانون زیف معروف است، برقرار است. این قانون در کل قرآن و در ۲۶ سوره آن، برای کلمات بلندتر از ۴ حرف صدق می‌کند.

با بررسی واژگان کل قرآن و ۲۶ سوره مذکور، فرمول زیر برای پیش‌بینی درصد فراوانی نسبی کلمات قرآنی بر اساس طولشان، به دست آمد:

$$F\% = \exp(a + b*L + c*L^2)$$

این فرمول نسبت به فرمول‌های مشابه از کارایی بیشتری برخوردار بوده و برای تمام کلمات قرآن با هر طولی قابل استفاده است؛ متغیرهای سه گانه موجود در این فرمول با توجه به سوره مورد بررسی در محدوده زیر قرار دارند:

$$1.85 \geq a \geq 1.68 \quad 0.78 \geq b \geq 0.65 \quad -0.09 \geq c \geq -0.11$$

از میان ۲۶ سوره مذکور، نمودار سوره (۲) بقره دارای بیشترین شباهت با نمودار قانون زیف در کل قرآن است؛ و فرمول پیشنهادی مربوط به کل قرآن، داده‌های دو سوره (۳۲) سجده و (۳۸) ص را به صورت دقیق‌تری پیش‌بینی می‌کند.

در پایان، ذکر این نکته لازم است که اثبات یک قانون بشری در متن قرآن هرگز با وحیانی بودن این کتاب مقدس تعارض ندارد، زیرا قرآن کریم به این مسئله اشاره نموده و می‌فرماید: «وَمَا أَرْسَلْنَا مِنْ رَسُولٍ إِلَّا بِلِسَانٍ قَوْمِهِ لِيُبَيِّنَ لَهُمْ» (ابراهیم: ۴)؛ (و هیچ فرستاده‌ای را، جز به زبان قومش نفرستادیم) یعنی علاوه بر اینکه این کتاب باید به زبان عربی که زبان قوم پیامبر اسلام (ص) بوده است، نازل می‌گشت می‌باید در جزئیات این زبان و نه الزاماً در تمام جزئیات، با آن همراهی کند.

منابع

- Altmann, Gabriel (2002), "Zipfian linguistics", *Glottometrics*, vol. 3.
- DataFit version 9.0.59, Oakdale Engineering, 1995-2008.
- Excel 2010, Microsoft Corporation.
- Prüin, Claudia & Zipf, Robert (2002) "Biographical notes on G. K. Zipf", *Glottometrics*, vol. 3.
- Rousseau, Ronald (2002), "George Kingsley Zipf: life, ideas, his law and informetrics", *Glottometrics*, vol. 3.
- Sigurd Bengt, Eeg-Olofsson Mats & Weijer, Joost van de (2004) "Word Length, Sentence Length and Frequency".
- Zipf, G.K. (1936), *The psycho-biology of language; an introduction to dynamic philology*, London, George Routledge & Sons Ltd.