

آمار واژه‌های زبان فارسی

ب. نیاورانی*



* آمار واژه‌های زبان فارسی

* اکبر قندهاری

* انتشارات رسیپنا، چاپ اول، ۱۳۸۶

- نمونه جدول مربوط به شمارش واژگان فرهنگ فارسی دکتر محمد معین را در تصویر ملاحظه می‌فرمایید.
- فهرست مطالب کتاب بدین شرح است:
۱. مقدمه: هدف، روش کار، حروف الفبای فارسی، تغییر خط فارسی، همسایگان زبانی زبان فارسی، زبان‌های مهم جهان، برآورد جمعیت فارسی‌زبانان (ایران - افغانستان - تاجیکستان)
 ۲. لغت‌نامه‌های مهم فارسی: لغت‌نامه دهخدا، فرهنگ فارسی معین، آمار واژه‌های لغت‌نامه‌های قدیمی، فرهنگ ناظم الاطباء (فرنودسار)، لغت‌نامه برهان قاطع، کتاب‌العین (فرهنگ تازی).
 ۳. تاریخچه فرهنگستان زبان فارسی
 ۴. توان واژه‌سازی زبان فارسی
 ۵. فرهنگ فارسی معین: نحوه شمارش واژه‌های فرهنگ فارسی معین، جدول آمار واژه‌های فرهنگ فارسی معین
 ۶. آمار واژه‌های لغت‌نامه برهان قاطع: (جدول آمار واژه‌های برهان قاطع)
 ۷. تعداد واژه‌های زبان‌های اروپایی در زبان فارسی (جدول تعداد واژه‌های زبان‌های اروپایی در زبان فارسی)
 ۸. واژه‌های عربی در زبان فارسی معاصر (جدول آمار واژه‌های عربی

یکی از مباحثی که در زبان‌شناسی و تدوین فرهنگ لغات هر زبانی اهمیت دارد، بررسی آماری واژگان یک زبان از نظر گاه‌های مختلف است. در زبان فارسی فرهنگ‌های بزرگ و کوچکی برای واژگان این زبان در دست داریم. بررسی واژه‌های گرد آمده در این فرهنگ‌ها، نمایی از کلیت زبان فارسی برای خواننده ترسیم می‌کند و شاید تصویری جدید در او بیافریند. این آمار در تحلیل‌های زبان‌شناختی و مطالعات لغت‌شناسی راهگشاست. کتاب «آمار واژه‌های زبان فارسی» با همین هدف تألیف شده است.

مبنای کار مؤلف در این کتاب، فرهنگ فارسی مرحوم دکتر محمد معین بوده است. نحوه شمارش واژه‌های فرهنگ فارسی دکتر محمد معین (شش جلدی متوسط) به این صورت بوده: «با توجه به اینکه تا تاریخ ۱۹۸۸ م. آغاز جمع‌آوری اطلاعات در مورد برآورد تعداد واژه‌های زبان فارسی، فرهنگ فارسی معین به صورت دیجیتال تهیه نشده بود، بناچار شمارش لغات به صورت دستی انجام شد. برای اجرای این کار، ابتدا جدول‌های مربوط تهیه آمار خام تنظیم و به مرور زمان، لغات بخش‌های مختلف فرهنگ فارسی معین، برحسب حرف اول و دوم هر واژه شمارش و در ستون‌های افقی و عمودی جدول‌ها ثبت شد و پس از کنترل نهایی، در برنامه‌های مربوط رایانه‌ای نوشته شد.»

جدول شماره ۱: آمار واژه‌های (بخش اول) فرهنگ فارسی معین که حرف اول و دوم آنها یکی از حروف الفبای فارسی است (از جلد اول تا چهارم، (آ - ی) از صفحه ۱ تا ۵۲۷۷)

ادامه جدول شماره ۱: آمار واژه‌های (بخش اول) فرهنگ فارسی معین که حرف اول و دوم آنها یکی از حروف الفبای فارسی است (از جلد اول تا چهارم، (آ - ی) از صفحه ۱ تا ۵۲۷۷)

حرف اول کلمه											حرف دوم کلمه											
ا	ب	پ	ت	ث	ج	ح	خ	د	ذ	ر	آ	ا	ب	پ	ت	ث	ج	ح	خ	د	ذ	ر
۱۶۵	۱۶۵	۱۶۵	۱۶۵	۱۶۵	۱۶۵	۱۶۵	۱۶۵	۱۶۵	۱۶۵	۱۶۵	۱۶۵	۱۶۵	۱۶۵	۱۶۵	۱۶۵	۱۶۵	۱۶۵	۱۶۵	۱۶۵	۱۶۵	۱۶۵	۱۶۵

در زبان فارسی معاصر) فرهنگسازان (شفاهی و نوشتاری جوهری و دیژیتالی) را در هر زبان با استفاده از آمار، تجزیه و تحلیل کرد و حدود روند رشد و تکامل و سیر اندیشه در هر زبان را در زمان‌های مختلف مشخص کرد. مثلاً حافظ در غزلیات خود یکصد و چهل و چهار بار واژه «گل» را به کار برده است و به همین منوال می‌توان تعداد، ترکیب و تکرار واژه‌هایی را که فرهنگسازان زبان فارسی در نوشته‌های خود به کار برده‌اند در زمان‌های مختلف محاسبه کرد. همانطور که خانلری (در کتاب زبان‌شناسی و زبان فارسی در صفحه ۲۱۶) می‌نویسد: «فردوسی برای سرودن شاهنامه که یک کتاب و یک شیوه و یک نوع معنی و مطلب در آن بیان می‌شود نزدیک نه هزار (۹۰۰۰) لغت به کار برده است و ما شاید برای انشای هزاران کتاب و هزاران مضمون و معنی همان قدر لغت به کار می‌بریم». تعداد، ترکیب و تکرار به کار بردن واژه‌ها در نوشته‌ها در زمان‌های مختلف نمایانگر ویژگی‌های خاص آن دوران است که مورد استفاده سایر متخصصان قرار می‌گیرد (روان‌شناسی اجتماعی، روانکاوی و ادبیات، جامعه‌شناسی و ...). (آمار واژه‌های زبان فارسی، ص ۱۰۶)

خلاصه نتایجی که نویسنده از استخراج آمارها در منابع پیش گفته به دست آورده، بدین شرح است:

۱. کل واژه‌های فرهنگ فارسی دکتر معین ۹۷/۹۵۸ واژه است. (۷۷/۲ درصد در بخش اول؛ ۴/۱۱ درصد در بخش دوم و ۱۸/۶ درصد در بخش سوم)
۲. در واژه‌های فرهنگ ناظم‌الاطباء ۶۳ درصد واژه‌های عربی است.
۳. در میان واژه‌های اروپایی دخیل در زبان فارسی، ۸۲/۷ درصد،

در زبان فارسی معاصر)

۹. واژه‌های فارسی پیشنهادی احمد کسروی (جدول آمار واژه‌های فارسی پیشنهادی احمد کسروی)
۱۰. تعداد حروف و کلمات قرآن (جدول آمار الفبای عربی در قرآن)
۱۱. نام‌های فارسی (جدول مقایسه تعداد نام‌های ایرانی)
۱۲. تعداد نام‌های عربی که با یکی از حروف الفبای عربی آغاز می‌شوند (جدول تعداد نام‌های عربی - جدول نام‌های مشهور عربی)
۱۳. تعداد غزلیات حافظ (جدول تعداد غزلیات حافظ بر حسب حرف اول هر غزل - تعداد حروف الفبا در غزلیات حافظ - ترکیب حروف اول کوچک با حرف دوم بزرگ - تعداد حروف الفبای فارسی در غزلیات حافظ - مقایسه تکرار برخی از واژه‌ها در غزلیات حافظ - جدول مقایسه آمار برخی از کلمات که در غزلیات حافظ تکرار شده است - تعداد واژه‌های دانشنامه‌مزدیسنا)
۱۴. آمار واژه‌ها در اعلام تاریخی - جغرافیایی تاریخ بیهقی
۱۵. تعداد واژه‌های واژه‌نامه‌های علمی و فنی در زبان فارسی (جدول انواع شکل‌های مختلف حروف نوشتاری الفبای فارسی)
۱۶. نتیجه‌گیری.

در نتیجه‌گیری نویسنده از پژوهش حاضر که استخراج آمارهای متفاوت در جدول‌های متعدد از چند فرهنگ لغت زبان فارسی است، می‌خوانیم: «زبان که یکی از رشته‌های علمی است، آمار نیز کاربرد خاص خود را در این رشته دارد و با توجه به پیشرفت تکنولوژی کامپیوتر و طراحی برنامه‌های مخصوص تجزیه و تحلیل نوشتاری می‌توان اندیشه‌های

متعلق به زبان فرانسه است.

۴. کل واژه‌های قرآن کریم ۷۷/۷۴۵ واژه و تعداد حروف قرآن ۳۳۰/۷۳۳ حرف است. (۳۱/۲۸ درصد آن انواع همزه، الف، م است). تعداد حروف قرآن که توسط ابن عطیه شمارش شده ۸/۸ درصد از این تعداد کمتر است. همچنین عددی که شیخ بهایی از حروف قرآن، به دست آورده ۱/۸ درصد از تعداد مذکور است.

۵. تعداد غزلیات حافظ در نسخ گوناگون: «انجوی شیرازی: ۵۱۹ غزل، خطیب رهبر: ۴۹۵ غزل، سایه: ۴۸۴ غزل، پژمان: ۴۹۷ غزل). تعداد حروف کل غزلیات حافظ ۲۶۵/۶۱۲ حرف است (رقم نهایی نیست) ۶. تعداد واژه‌های واژه‌نامه‌های علمی و فنی مرکز نشر دانشگاهی تا سال ۱۳۷۱، ۷۴/۰۴۰ واژه است.

کوشش نویسنده در ارائه پژوهشی آماری برای تحلیل‌های زبان‌شناختی در خور تقدیر است، اما به چند نکته باید توجه داشت و توجه به این نکات می‌تواند میزان سودمندی و اعتبار آمارهای استخراج شده را تعیین نماید.

۱. مبنا قرار دادن فرهنگ فارسی مرحوم دکتر محمد معین برای این نوع از تحلیل آماری و نتیجه‌گیری بر اساس آن، غیر علمی است، چون این فرهنگ حدود ۳۰ سال پیش منتشر شده و اکنون فرهنگ کامل و به‌روزی محسوب نمی‌شود. اگرچه فضل تقدم برای آن مرحوم تا زمانی که زبان فارسی زنده است، محفوظ خواهد بود. این فرهنگ تقریباً نخستین کار با روش علمی در زمینه تدوین فرهنگ لغات زبان فارسی بوده است. امروز، نسخه به‌روز شده فرهنگ فارسی مرحوم دکتر محمد معین را «فرهنگ بزرگ سخن» به سرپرستی استاد گرامی دکتر حسن انوری باید تلقی کرد و هر گونه تحلیل آماری باید بر اساس این کتاب ارائه شود. البته آنچه از فحوای کتاب برمی‌آید این است که این پژوهش آماری، در سال ۱۹۹۸ م. یعنی حدود ۱۰ سال پیش انجام شده و اگرچه جامعه آماری این کتاب در زمینه فرهنگ‌های قدیمی زبان فارسی بدون تغییر مانده است، اما این جامعه آماری در زمینه فرهنگ‌ها زبان فارسی امروز و نیز دیوان حافظ و پژوهش‌هایی که اخیراً در زمینه دیوان حافظ منتشر شده، همه تغییر کرده است و آمار ارائه شده در این کتاب را به نوعی باید «آمار تاریخی» تلقی کرد و طبعاً نمی‌توان از این آمار در زمینه تحلیل‌های زبان‌شناسی سود جست. وانگهی نرم‌افزارها و کارهایی که هم‌اکنون در فرهنگستان زبان و

ادب فارسی در زمینه فرهنگ‌نویسی به ویژه تحت نظر استاد گرامی دکتر علی اشرف صادقی در دست انجام است، جایی برای طرح چنین آمارهایی باقی نمی‌گذارد.

۲. هر آماری که از فرهنگی مثل فرهنگ فارسی دکتر محمد معین استخراج می‌شود، اگر بنا باشد این آمار واقعاً مفید و قابل استفاده باشد و مشکلی از مشکلات را حل کند، حداقل باید موارد زیر را از کتاب استخراج کند:

- الف) تعداد واژه‌های کل فرهنگ
- ب) تعداد واژه‌های عربی کل فرهنگ
- ج) تعداد واژه‌های اروپایی به تفکیک زبان مبدأ
- د) تعداد ترکیبات و عبارتهای فعلی و وصفی و ... (به تفکیک)
- ه) تعداد اصطلاحات مربوط به علوم و فنون مختلف، همه به تفکیک (از جمله مثلاً اصطلاحات عرفانی، فلسفه و منطق، نجوم، ریاضیات، پزشکی و ...)
- ز) تعداد کلمات و عبارات مربوط به عوام و کوچه و بازار
- ح) تعداد ضرب‌المثل‌ها و تعبیرات تمثیلی
- ط) تعداد کنایات
- و

صرف دانستن اینکه تعداد کلماتی که با فلان حرف از حروف الفبا شروع می‌شوند، البته اگرچه مفید است اما در مقایسه با حداقل اطلاعات آماری که می‌توان از یک فرهنگ استخراج کرد، بسیار پیش یا افتاده و ناچیز است.

۳. در میان فرهنگ‌هایی که استخراج آمار آنها، مثلاً از کتاب «العین» که یک فرهنگ عربی گمنام است بسیار ضروری‌تر و مفیدتر است، چندین فرهنگ بسیار مهم نادیده گرفته شده است، از جمله مثلاً لغت‌نامه دهخدا، فرهنگ فارسی مرحوم دکتر حسن عمید، فرهنگ معاصر فارسی صدری افشار، و خانم‌ها نسرتین و نسترن حکمی، فرهنگ فارسی عامیانه ابوالحسن نجفی و ...

و البته نکات دیگری هم هست که مجال دیگر می‌طلبید، اما کتاب «آمار واژه‌های زبان فارسی» با همین وضع موجود، در حد خود قابل استفاده است و اطلاعات سودمندی می‌توان از آن به دست آورد.

پی‌نوشت:

کارشناس ارشد زبان و ادبیات فارسی