

فصلنامه علمی - پژوهشی زبان پژوهی دانشگاه الزهراء (س)

سال هفتم، شماره ۱۷، زمستان ۱۳۹۴

بررسی تطبیقی کلمات اختصاری در فارسی، انگلیسی، عربی و ترکی استانبولی

علی رضا قلی فامیان^۱

مریم سخنور^۲

تاریخ دریافت: ۹۱/۶/۲۷

تاریخ تصویب: ۹۳/۷/۲۸

چکیده

یکی از فرایندهای عمده در واژه‌سازی، اختصارسازی است که در آن یک حرف یا مجموعه‌ای از حروف جایگزین یک کلمه یا عبارتی طولانی می‌شود. انگیزه اصلی اختصارسازی را می‌توان اصل کم‌کوشی معرفی کرد که بر اساس آن انسان می‌کوشد با صرف کمترین انرژی بیشترین حجم اطلاعات را انتقال دهد. در این پژوهش، ساختمان و بسامد کلمات اختصاری در چهار زبان فارسی، انگلیسی، عربی و ترکی استانبولی بررسی شده است. ابتدا وبگاه‌های اینترنتی چهار کشور جمهوری اسلامی ایران، ایالات متحده آمریکا، مصر و ترکیه در شش حوزه خبر، سیاست، ورزش، تجارت، علوم و فرهنگ در نظر گرفته شده

^۱ استادیار گروه زبان‌شناسی و زبان‌های خارجی، دانشگاه پیام نور (نویسنده مسئول)؛ famian@pnush.ac.ir

^۲ دانشجوی کارشناسی ارشد زبان‌شناسی همگانی، دانشگاه آزاد اسلامی واحد اهر؛ mary2007am@yahoo.com

و برخی متون آنها به طور تصادفی انتخاب شده است. علت انتخاب وبگاه‌های اینترنتی این است که متون درج شده در این رسانه به طور معمول در صفحه نمایشگر به نمایش درمی آید و استفاده از کلمات اختصاری باعث صرفه جویی در فضای مجازی می شود. متون انتخاب شده برای هر زبان شامل ۶۰۰۰۰ واژه می باشد و کل پیکره ۲۴۰۰۰۰ واژه را دربرمی گیرد. یافته‌های تحقیق نشان می دهد که زبان انگلیسی از نظر تعداد کلمه (نوع) و نیز بسامد تکرار کلمات اختصاری در جایگاه نخست است و ترکی استانبولی، فارسی و عربی در رتبه‌های بعدی قرار می گیرند. بررسی ساختار درونی کلمات اختصاری نیز نشان می دهد که بیش از نیمی از کلمات اختصاری پیکره سه حرفی هستند. تأمل در دو عامل الفبا و ملاحظات رده شناختی چهار زبان مورد بررسی نیز نشان می دهد که عامل الفبا و خط نقش مهم تری در گرایش اهل زبان به ساخت و کاربرد کلمات اختصاری ایفا می کند.

واژه‌های کلیدی: کلمه اختصاری، فارسی، انگلیسی، عربی، ترکی

استانبولی

۱. مقدمه

ساخت واژه یا صرف بخشی از زبان شناسی است که به بررسی ساختار تکواژ، واژه و نیز فرایندهای واژه سازی می پردازد (کاتامبا، ۱۹۹۳). یکی از فرایندهای عمده در واژه سازی اختصار سازی است که در آن یک حرف یا مجموعه‌ای از حروف جایگزین یک کلمه یا عبارتی طولانی می شود. انگیزه اصلی اختصار سازی را می توان اصل کم کوشی یا تلاش کمینه^۱ معرفی کرد. این اصل برای نخستین بار توسط زیف در زمینه عملکرد انسان در حوزه علوم ارتباطات و زبان شناسی ارائه شد (زیف، ۱۹۴۹، VII). بر اساس این اصل، انسان و اصولاً هر عامل دیگری می کوشد با صرف کمترین انرژی بیشترین حجم اطلاعات را

^۱ principle of least effort

انتقال دهد. اگر اصل کم‌کوشی را در چارچوب زبان و حوزه ساخت‌واژه در نظر بگیریم، می‌توان ادعا کرد برخی فرایندهای واژه‌سازی در زبان‌ها حاصل عملکرد اصل کم‌کوشی هستند. برای مثال، فرایند آمیزش واژگانی^۱ در واژه «عکاشی» (حاصل ترکیب بخش ابتدایی واژه «عکاسی» و بخش انتهایی واژه «نقاشی») نمونه‌ای از اعمال اصل کم‌کوشی است زیرا یک واژه جایگزین دو واژه شده است. اختصارسازی نیز نمونه‌ای بارز از پیاده‌سازی اصل کم‌کوشی در ساخت‌واژه به شمار می‌آید.

کلمات اختصاری دو نوعند: آغازین‌ها^۲ و سرنام‌ها^۳. منظور از آغازین‌ها صورت‌های کوتاه‌شده‌ای هستند که از ترکیب حروف آغازین واژه‌های تشکیل‌دهنده یک عبارت، بدون توجه به صورت آوایی آن حاصل می‌شوند. برای مثال «ه.ش.» (هجری شمسی) نمونه‌ای از یک آغازین در فارسی است. سرنام نیز به همین شکل ساخته می‌شود با این تفاوت که صورت حاصل به‌عنوان واژه‌ای مستقل در زبان شناخته و تلفظ می‌شود مانند «هما» (هوایمایی ملی ایران). لازم به ذکر است که در این پژوهش تمام صورت‌های اختصاری، اعم از آغازین‌ها و سرنام‌ها، مدنظر قرار گرفته‌اند. صرف‌نظر از این که آغازین‌ها یا سرنام‌ها چگونه ساخته می‌شوند، این گونه کلمات (البته نه تمام آنها) به تدریج و به مرور زمان و با توجه به نیاز اهل زبان به واژه‌های مستقل و متعارف تبدیل می‌شوند تا جایی که گاه اهل زبان واژه‌ای اختصاری را به کار می‌برند بی آن که از اصل و ریشه آن اطلاع داشته باشند. نمونه بارز این مسئله کلمه اختصاری laser در انگلیسی است که با وجود کاربرد گسترده آن، بسیاری از مردم انگلیسی‌زبان از صورت اصلی این کلمه اختصاری آگاه نیستند.

در این پژوهش می‌کوشیم پس از مرور پیشینه تحقیق به چند پرسش درخصوص کلمات اختصاری پاسخ دهیم. نخست این که چهار زبان فارسی، انگلیسی، عربی و ترکی از لحاظ کاربرد کلمات اختصاری در فضای مجازی اینترنت چه تفاوت‌ها یا شباهت‌هایی با یکدیگر دارند؟ پرسش دوم این که ساختمان متداول کلمات اختصاری چیست و از میان

^۱ blending

^۲ initialism

^۳ acronym

کلمات اختصاری یک، دو، سه یا چهار حرفی، کدام گروه رایج‌تر است؟ از سوی دیگر با توجه به این که چهار زبان مورد نظر در این پژوهش به رده‌های ساخت‌واژی متفاوتی تعلق دارند، این پرسش مطرح است که رده‌بندی ساخت‌واژی چه نقشی در فرایند اختصارسازی دارد؟

۲. پیشینه تحقیق

در این بخش به معرفی پژوهش‌های انجام‌شده در چارچوب کلمات اختصاری می‌پردازیم.

۲.۱. شقاقی (۱۳۸۹)

شقاقی چهار الگوی اصلی سرواژه‌سازی را به شرح زیر معرفی می‌کند:

(۱) حروف اول چند کلمه: «اتکا» (اتحادیه تعاونی‌های کارکنان ارتش)، «آزفا» (آموزش زبان فارسی)

(۲) حروف اول نام خارجی سازمان یا مؤسسه: «ایرنا» (Islamic Republic News Agency)

(۳) حروف اول ترجمه اسم خارجی: «ساف» (سازمان آزادی‌بخش فلسطین)

(۴) آمیزه‌ای از حروف و هجاها: «توانیر» (تولید و انتقال نیرو)

۲.۲. بیتس و همکاران (۲۰۰۰)

با توجه به نارسایی در شیوه‌های مرسوم و دستی در شناسایی سرنام‌ها، بیتس و همکاران (۲۰۰۰) شیوه‌ای جدید برای استخراج سرنام‌ها در محیط کتابخانه دیجیتال ارائه دادند. در این روش، الگوهای پی.پی.ام.^۱ متعددی برای کدگذاری سرنام‌ها بر مبنای تعریف‌شان به کار رفته است. در این مدل، چهار مشخصه مورد استفاده در کدگذاری هر سرنام عبارتند از: (۱) جایگاه سرنام قبل یا بعد از تعریف، (۲) فاصله بین سرنام و تعریف، (۳) حروف گرفته‌شده از کلمات برای ساخت سرنام و (۴) شمار حروف گرفته‌شده برای تشکیل سرنام. بیتس با استفاده از این مشخصه‌ها و نیز الگوهای پی.پی.ام. توانست مدلی ارائه دهد که

^۱ PPM

برنامه‌ای رایانه‌ای بتواند از رشته‌ای از کلمات به‌طور خودکار سرنام بسازد. یافته‌های تحقیق مذکور نشان داد که این روش از شیوه‌های شمی و دستی کارآمدتر است.

۳.۲. تقوا و گیلبرت (۱۹۹۹)

تقوا و گیلبرت (۱۹۹۹) در جریان طراحی سیستم پس‌پردازش^۱ جهت اصلاح خروجی متن از ابزارهای تشخیص کاراکتر نوری^۲ با مشکلی به نام کلمات اختصاری مواجه شدند. کلمات اختصاری کلماتی بودند که در فرهنگ واژه‌ها درج نمی‌شدند اما به‌وفور در متون به کار می‌رفتند. بخش عمده‌ای از سامانه مذکور درگیر تشخیص کلمات اشتباه بود. در پی بروز این مشکل، تقوا و گیلبرت سامانه‌ای را برای ردیابی این نوع کلمات طراحی کردند. این سامانه در چهار مرحله عمل می‌کند: مرحله آمادگی، فیلتر درون‌داد، تقطیع باقی‌مانده درون‌داد به واژه‌ها، و عملکرد الگوریتم ردیابی کلمه اختصاری.

الف) مرحله آمادگی: در این مرحله، فهرست‌هایی به سامانه ارائه می‌شود: (۱) فهرست کلمات مرزنا؛ مانند the، and و غیره. سامانه باید این قبیل کلمات را از کلمات معمولی تمیز دهد تا الگوریتم بتواند به‌خوبی عمل کند. (۲) فهرستی از کلمات مرجوعی یعنی کلماتی که در متن به‌وفور به کار رفته‌اند اما کلمه اختصاری نیستند؛ مانند کلمه table یا figure، یا اعداد. هر چه این قبیل کلمات بیشتر کنار گذاشته شوند، سامانه با سرعت و دقت بیشتری می‌تواند کلمات اختصاری درون متن را تشخیص دهد. (۳) دادگانی از کلمات اختصاری به‌همراه تعریف‌شان (۴) متنی که قرار است سامانه کلمات اختصاری‌اش را استخراج کند.

ب) مرحله فیلتر: در این مرحله، درون‌داد پیش‌پردازش شده و سطرهایی از متن که با حروف بزرگ درج شده‌اند و معمولاً عنوان اصلی یا فرعی هستند کنار گذاشته می‌شوند. در همین مرحله، سامانه به فهرست کلمات مرجوعی نیز مراجعه می‌کند. اگر سامانه احتمال

^۱ post-processing

^۲ OCR (Optical Character Recognition)

^۳ stopwords

داد که با یک کلمه اختصاری مواجه شده است، ابتدا آن را در فهرست کلمات مرجوعی جستجو می‌کند و اگر آن را در فهرست مذکور نیافت، به کارش ادامه می‌دهد.

ج) تقطیع باقی مانده درون داد به واژه‌ها: در این مرحله، برای انواع مختلف کلمات اولویت‌های خاصی در نظر گرفته می‌شود. اولویت‌های مذکور عبارتند از این که کلمه یا مرزنامست، یا کلماتی با خط تیره است، یا کلمه اختصاری است و یا کلمه‌ای معمولی است که در هیچ یک از مقولات بالا نمی‌گنجد. لازم به ذکر است که الگوریتم نمی‌تواند به‌طور کامل کلمات مرزما را نادیده بگیرد زیرا در این صورت بسیاری از کلمات اختصاری تشخیص داده نمی‌شوند. برای مثال اگر سامانه حرف اضافه of را به‌عنوان کلمه‌ای مرزما کنار بگذارد، آن‌گاه کلمه اختصاری (Department of Energy) DOE تشخیص داده نمی‌شود.

د) عملکرد الگوریتم: در آخرین مرحله، الگوریتم توالی آشنا و رایج کلمه اختصاری را شناسایی می‌کند و آرایه اصلی تعریف احتمالی آن را می‌یابد.

۲.۴. هارلی (۲۰۰۴)

هارلی مؤلف مقاله‌ای است با عنوان چرا The CIA اما The NASA*؟ منظور از ستاره در بالای کلمه «ناسا» این است که کاربرد این واژه با حرف تعریف the غیردستوری است. هارلی برای پاسخ به این پرسش ۶۱ سرنام را از وبگاه‌های انگلیسی استخراج کرد و نشان داد که رفتار نحوی صورت‌های کوتاه‌سازی شده و اختصاری قاعده‌مند است. بر اساس تحقیق مذکور، سرنام‌ها مانند اسم‌های خاص عمل می‌کنند و از این رو حرف تعریف the حذف می‌شود در حالی که صورت‌های آغازین عملکرد صورت اولیه خود، یعنی اسم‌های مرکب را حفظ کرده و حرف تعریف را حفظ می‌کنند.

۲.۵. دانلز (۲۰۰۵)

از آنجا که اسناد آنلاین بسیار زیادی در دنیای مجازی وجود دارد، در اغلب موارد یافتن اطلاعات مورد نظر جستجوگر دشوار است. زبان‌شناسان رایانه‌ای کوشیده‌اند با طراحی برنامه‌های رایانه‌ای مختلف، جستجو و یافتن اطلاعات را تسریع ببخشند. دانلز کوشیده است شناسایی کلمات اختصاری در متون پزشکی زبان سوئدی را آزمایش کند. وی با کمک شیوه‌ای قاعده‌بنیاد^۱ این کار را انجام داد و آن را با شیوه‌های مختلف یادگیری ماشین^۲ مقایسه کرد. یافته‌های تحقیق مذکور نشان داد که اگر برنامه رایانه‌ای اطلاعات نحوی کافی داشته باشد و نیز پیش‌پردازش متنی صورت بگیرد، برنامه می‌تواند کلمات اختصاری به کاررفته در متون پزشکی وبگاه‌های سوئدی را تشخیص دهد. از آنجا که در حوزه علم و علوم پزشکی کلمات اختصاری فراوانی برای نامیدن دستگاه‌ها، روش‌ها، بیماری‌ها یا داروها به کار می‌رود، این کار کمک بسیاری به محققان حوزه پزشکی در استخراج سریع اطلاعات می‌کند.

۲.۶. شوارتز و هرست (۲۰۰۳)

در تحقیقی دیگر، شوارتز و هرست کوشیده‌اند الگویی را برای استخراج اطلاعات از چکیده مقالات زیست-پزشکی ارائه دهند. از نظر آنان، حجم گسترده متون علمی چالش بزرگی را پیش روی محققان بازیابی اطلاعات قرار داده است. با توجه به بسامد بسیار زیاد کلمات اختصاری در متون علمی، این دو محقق بر روی استخراج این گونه کلمات متمرکز شدند و الگوریتم ساده و در عین حال سریعی را برای استخراج کلمات اختصاری از متون زیست-پزشکی عرضه کردند. کار آنان در واقع استخراج کلمه اختصاری به همراه تعریف آن در متن بوده است. آنان با این که از رویکردهای یادگیری ماشین در زبان‌شناسی رایانه‌ای استقبال کرده‌اند، معتقدند در این زمینه مشخص، استفاده از الگوریتم ساده مناسب‌تر است. به اعتقاد آنان، مزیت سادگی الگوریتم ارائه‌شده این است که در زمانی کوتاه به هدف مورد نظر دست می‌یابد. با الگوریتم آنان می‌توان در یک ثانیه هزار

^۱ rule-based

^۲ machine learning

چکیده مقاله را پردازش کرد و کلمات اختصاری آنها را استخراج نمود. الگوریتم آنان به زبان جاوا و در ۲۶۰ سطر نوشته شد و کار با آن به تخصص خاصی نیاز ندارد. شوارتز و هرست نیز مانند دانلز به این نتیجه رسیدند که خطاهای موجود در برنامه با افزودن اطلاعات نحوی کم می‌شود. ناگفته نماند که پیش از هر عملیات جستجو توسط این برنامه، بهتر است یک چکیده از پیکره معین به برنامه ارائه شود تا برنامه بتواند به راحتی در چکیده مقالات مشابه به جستجو پردازد.

۲.۲. رضا قلی فامیان (۲۰۰۸)

رضا قلی فامیان (۲۰۰۸) به بررسی کلمات اختصاری در فارسی و انگلیسی و میزان استفاده از آنها در وبگاه‌های اینترنتی پرداخته است. در این پژوهش ادعا شده است که در جوامع توسعه یافته که سرعت زندگی بالاست و گرایش به اقتصاد زبانی وجود دارد، کاربرد کلمات اختصاری نیز بیشتر است. برای ارزیابی این فرضیه، وبگاه‌هایی در دو زبان فارسی و انگلیسی در شش حوزه اخبار، دولت، علوم، تجارت، ورزش و فرهنگ انتخاب شده است و کلمات اختصاری موجود در آنها شمارش شده است. یافته‌های پژوهش فوق نشان می‌دهد که در زبان انگلیسی استفاده از کلمات اختصاری رایج‌تر از زبان فارسی است.

۲.۸. زاهدی و شریفی (۱۳۹۰)

زاهدی و شریفی (۱۳۹۰) پس از مقدمه‌ای درباره اهمیت اختصارسازی و ویژگی‌های آن در زبان فارسی کوشیده‌اند سابقه این فرایند واژگانی را در خط میخی فارسی باستان و خط پهلوی نشان دهند. آنان نیز بر این باورند که انگیزه اصلی فرایند اختصارسازی اصل اقتصاد و کم‌کوشی در زبان است. آنان به نقل از تفضلی (۱۳۷۵ و ۱۳۷۶) و امینی (۱۳۸۳) به وجود برخی کلمات اختصاری بر روی سکه‌های دوره ساسانی و اشکانی اشاره کرده‌اند. زاهدی و شریفی استدلال می‌کنند که اختصارسازی در زبان فارسی از قدمت زیادی برخوردار است و در گذشته احتمالاً به دلیل محدودیت در فضای کتیبه‌ها، نگارندگان مایل به استفاده از کلمات اختصاری بوده‌اند.

۳. اختصارسازی در زبان‌های فارسی، انگلیسی، عربی و ترکی استانبولی

پیش از آشنایی با اختصارسازی در چهار زبان فارسی، انگلیسی، عربی و ترکی استانبولی، ابتدا با رده‌زبانی و به‌ویژه رده‌ساخت‌واژی این چهار زبان آشنا می‌شویم.

۳.۱. فارسی

فارسی یکی از زبان‌های ایرانی شاخه‌زبان‌های هندوایرانی است و امروزه به‌طور عمده در ایران، افغانستان و تاجیکستان تکلم می‌شود. این زبان دنباله‌زبان فارسی میانه است که زبان رسمی و ادبی دوران ساسانی بود. فارسی ۱۱۰ میلیون متکلم مادری دارد و در سه کشور فوق‌الذکر زبان رسمی است.

دستور زبان فارسی شبیه بسیاری از زبان‌های هندواروپایی و به‌خصوص زبان‌های خانواده‌هندوایرانی است. زبان فارسی میانه ماهیت تجزیه‌ای^۱ زیادی به خود گرفت زیرا جنس دستوری را از دست داد و وندهای حالت آن به حداقل رسید. فارسی امروزی نیز این خصوصیات را به ارث برده است. به این ترتیب، از نظررده‌شناسی ساخت‌واژی، فارسی معاصر زبانی کم و بیش گسسته^۲ به شمار می‌آید.

همان‌طور که در بخش ۲.۸ اشاره شد، اختصارسازی در فارسی قدمت زیادی دارد و سابقه ساخت کلمات اختصاری به دوران باستان برمی‌گردد. در فارسی معاصر نیز بنا به نیاز سازمان‌ها و نهادهای دولتی و غیردولتی، کلمات اختصاری ساخته شده و می‌شود. در زیر برخی از این کلمات اختصاری را مرور می‌کنیم:

ص: صلی الله علیه و آله و سلم

ص: صفحه

ره: رحمه الله علیه

ه.ش.: هجری شمسی

پ.ت.ت.: پست، تلگراف، تلفن

^۱ analytical

^۲ isolating

هما: هواپیمایی ملی ایران

سیبا: سیستم یکپارچه بانک ملی ایران

شابک: شماره استاندارد بین‌المللی کتاب

مرور مثال‌های فوق نکات جالب توجهی را خاطر نشان می‌سازد. نخست این که کلمات یا نشانه‌های اختصاری در فارسی می‌تواند یک حرفی (ص)، دو حرفی (ه.ش.)، سه حرفی (هما) یا چهار حرفی (سیبا) باشد. دوم این که مانند واژگان معمولی زبان که گاه چند معنا هستند، گاه کلمات یا نشانه‌های اختصاری نیز چند معنا هستند. برای مثال «ص» هم می‌تواند به شماره صفحه اشاره کند و هم نشانه‌ای باشد برای صلوات بر پیامبر اسلام^(ص). نکته دیگر این که گاه حروف اختصاری به لحاظ آوایی با یکدیگر تشکیل یک زنجیره منسجم را نمی‌دهند و لذا کاربرد چندانی پیدا نمی‌کنند (مانند «پ.ت.ت.»). برعکس، کلمات اختصاری، که از ترکیب آوایشان زنجیره‌ای آشنا و متعارف در زبان به وجود می‌آید، به احتمال زیاد راحت‌تر توسط اهل زبان به کار می‌روند (مانند «هما»).

۳.۲. انگلیسی

زبان انگلیسی به خانواده ژرمنی غربی تعلق دارد که ابتدا در انگلستان تکلم می‌شد و امروزه در اقصی نقاط جهان به عنوان زبان اول یا دوم و هم‌چنین به عنوان زبان اول علم، ارتباطات، فناوری و حمل و نقل کاربرد دارد. از نظر جمعیت گویشور مادری بعد از چینی ماندارین و اسپانیایی سومین زبان به شمار می‌آید و نیز زبان رسمی اتحادیه اروپا، کشورهای مشترک‌المنافع و سازمان ملل است. زبان انگلیسی معاصر حاصل ترکیب گویش‌های نزدیک به هم است که امروزه آنها را گویش‌هایی از انگلیسی باستان می‌دانند. بسیاری از کلمات این زبان ریشه لاتین دارند زیرا در گذشته لاتین زبان کمکی کلیسای مسیحی و جامعه روشنفکری اروپا بود.

در دستور زبان انگلیسی برخلاف اغلب زبان‌های هندواروپایی شاهد حداقل تصریف هستیم. برای مثال، در این زبان برخلاف زبان‌های آلمانی و هلندی، جنس دستوری و یا

مطابقه صفت وجود ندارد. از نظر رده‌شناسی، انگلیسی زبان گسسته محسوب می‌شود و به بیان دیگر، معنای جمله در آن نه با ترکیب وندها و ریشه‌های متعدد بلکه از طریق آرایش سازه‌ها و افعال کمکی منتقل می‌شود. پرسش، نفی، ساخت مجهول و وجه استمراری و غیره نه از طریق وندافزایی بلکه به واسطه افعال کمکی انجام می‌گیرد.

به دلیل گستردگی کاربرد انگلیسی در سرتاسر جهان، کالاها و خدمات بیشماری با نام انگلیسی روانه بازارهای جهانی می‌شوند و بدیهی است نام‌گذاری این قبیل کالاها و خدمات مسئله‌ای جالب توجه در زبان است. شرکت‌ها و نهادهای اقتصادی و خدماتی نام‌گذاری محصولات خود را یکی از مراحل تبلیغات می‌دانند و معتقدند نام مناسب یکی از رموز موفقیت تجاری‌شان است. یکی از ویژگی‌های نام خوب در دنیای تبلیغات کوتاه بودن آن است و همین نکته باعث شده است که اهل زبان انگلیسی به اختصارسازی روی بیاورند. در واقع، خصلت جهان‌شمول بودن زبان انگلیسی باعث شده است هم‌زمان با واژه‌سازی به طرق مختلف (مانند قرض‌گیری^۱، ابداع^۲، ترکیب و غیره)، اختصارسازی نیز پرطرفدار باشد. در جدول زیر چند نمونه از کلمات اختصاری زبان انگلیسی را مرور می‌کنیم.

جدول ۱. نمونه‌هایی از کلمات اختصاری انگلیسی

کلمه اختصاری	عبارت کامل	معادل فارسی	
C	centigrade	سانتی‌گراد	۱
U.S.	United States	ایالات متحده	۲
et al.	et alii	و دیگران	۳
BBC	British Broadcasting Corporation	بی‌بی‌سی	۴
NATO	North Atlantic Treaty Organization	ناتو	۵
laser	light amplification by stimulated emission of radiation	لیزر	۶
Interpol	International Criminal Police Organization	پلیس بین‌الملل	۷

¹ borrowing

² coinage

با مرور اطلاعات جدول بالا درمی‌یابیم در انگلیسی کلمه اختصاری تک حرفی (C)، دو حرفی (US)، سه حرفی (BBC)، چهار حرفی (NATO)، پنج حرفی (laser)، و حتی هشت حرفی (Interpol) وجود دارد. دوم این که بعضی کلمات اختصاری از کلمات لاتین گرفته شده‌اند، مانند et al (جالب این است که خود این کلمه مرکب لاتین در انگلیسی کاربرد چندانی ندارد اما کلمه اختصاری آن به‌ویژه در متون علمی کاربرد فراوان دارد). نکته دیگر این که قاعده کاملاً ثابتی برای ساخت کلمه اختصاری وجود ندارد. برای مثال، در کلمه اختصاری BBC، اولین حروف عبارت سه‌کلمه‌ای British Broadcasting Corporation انتخاب شده است در حالی که در Interpol چنین اسلوبی رعایت نشده است.

۳.۳. عربی

عربی مهم‌ترین زبان خانواده زبان‌های سامی^۱ است که در خاورمیانه و آفریقای شمالی کاربرد دارد. به دلیل گستردگی جغرافیایی آن، گویش‌های مختلفی در این زبان دیده می‌شود. با وجود تنوع و تفاوت آوایی و واژگانی این لهجه‌ها و گویش‌ها، به دلایل سیاسی و قومی، همگان آنها را نمونه‌هایی از یک زبان واحد می‌دانند. گفتنی است عربی کلاسیک قرآنی در ایجاد چنین انسجامی تأثیر بسیار دارد. اگر تمام گونه‌های عربی را در حکم یک زبان واحد در نظر بگیریم، این زبان ۲۸۰ میلیون نفر گویشور مادری دارد و اگر گونه‌های آن را از هم تفکیک کنیم، گونه عربی مصری به‌عنوان بزرگ‌ترین گونه ۹۵ میلیون نفر گویشور دارد.

زبان نوشتاری عربی از زبان قرآن نشأت گرفته است که به‌طور گسترده در مدارس و دانشگاه‌ها تدریس می‌شود. گونه نوشتاری عربی با الفبای عربی و از راست به چپ نوشته می‌شود، هر چند در برخی نقاط گونه‌های شفاهی این زبان با الفبای لاتین نیز نوشته می‌شود.

^۱ Semitic

برخلاف اغلب زبان‌ها، در عربی برای ساخت واژه‌ها کمتر از وندافزایی استفاده می‌شود و کلمات جدید بیشتر از طریق تغییر و تحول ریشه و ستاک مشتق می‌شوند. به این ترتیب از منظر رده‌شناسی، عربی در ردهٔ زبان‌های آمیخته^۱ جای می‌گیرد.

الغینانی (۲۰۰۷) معتقد است کاربرد کلمات اختصاری در عربی متداول نیست و لغت‌شناسان عرب چه در گذشته و چه در زمان حاضر ترجیح می‌دهند کلمات را به صورت کامل به کار ببرند. ساخت کلمات ترکیبی در زبان عربی نیز چندان رایج نیست. با این حال تعدادی کلمهٔ اختصاری ساخته شده است که در زیر به معرفی آنها می‌پردازیم:

د: دکتر

ج.م.ع.: جمهوری عربی مصر

کم: کیلومتر

ق.م.: قبل از میلاد

همان‌طور که در بالا می‌بینیم، کلمات و نشانه‌های اختصاری در عربی می‌تواند یک حرفی (د)، یا دو حرفی (کم) باشد، و یا بیش از دو حرف داشته باشد (ج.م.ع.).

۳. ۴. ترکی استانبولی

زبان ترکی استانبولی از شاخهٔ جنوب غربی زبان‌های ترکی ریشه گرفته است. این زبان رایج‌ترین زبان از شاخهٔ زبان‌های ترکی است و جغرافیای آن شامل ترکیه و قبرس شمالی و بخش‌هایی کوچک از عراق، یونان، بلغارستان، جمهوری مقدونیه، کوزوو، و آلبانی است. این زبان در قرون متمادی شاهد تحولات بسیاری بوده است و یکی از این تحولات که در دوران آتاتورک رخ داد، تغییر الفبای عثمانی به الفبای لاتین بود. در همان دوران، کلمات عربی بسیاری از زبان ترکی حذف شد (لويس، ۱۹۹۹). از نظر رده‌شناسی، ترکی زبان پیوندی^۲ است و بسیاری از مفاهیم دستوری یا واژگانی از طریق پیوند و ندها به یکدیگر منتقل می‌شود.

برخی کلمات اختصاری رایج در زبان ترکی به شرح زیرند:

^۱ fusional

^۲ agglutinative

Ü: Üniversite	(دانشگاه)
bÖl:bölüm	(بخش)
EAT: Eski Anadolu Türkçesi	(ترکیه آناتولی باستان)
Erm.: Ermenice	(ارمنی)
Hız: Hızret	(حضرت)

۴. گردآوری داده‌ها

به منظور گردآوری داده‌ها و ایجاد پیکره پژوهش، محققان ابتدا شش مقوله یا حوزه را شامل خبر، سیاست، علوم، فرهنگ، ورزش، و تجارت انتخاب کرده‌اند. در مرحله بعد، وبگاه‌های اینترنتی چهار زبان مورد نظر پژوهش در چهار کشور انتخاب شده است. پیکره مربوط به وبگاه‌های زبان فارسی از کشور جمهوری اسلامی ایران، انگلیسی از ایالات متحده آمریکا، عربی از کشور مصر و ترکی استانبولی نیز از کشور ترکیه استخراج شده است. صفحات مربوط به خبر از وبگاه‌های ایرنا، آسوشیتدپرس، یاهو، الاهرام و خبر ترک استخراج شده است. در خصوص وبگاه‌های سیاسی نیز به تارنماهای دولتی مانند وزارت خانه‌های چهار کشور مورد نظر مراجعه شده است. در مورد علوم، به وبگاه‌های دانشگاه‌ها و انجمن‌های علمی، در خصوص فرهنگ به وبگاه موزه‌ها و مراکز فرهنگی، در مورد ورزش، به وبگاه باشگاه‌های ورزشی و در خصوص تجارت نیز به وبگاه شرکت‌های بزرگ مراجعه شده است. در وبگاه‌های مذکور به طور عمده به صفحات درباره ما^۱، تاریخچه^۲ و نظایر آن مراجعه شده است. البته در مواردی به خصوص در وبگاه‌های خبری و ورزشی، مقالات اصلی وبگاه‌ها نیز مورد توجه قرار گرفته است.

در این پژوهش با دو اصطلاح نوع^۳ و شمار^۴ نیز سروکار داریم. منظور از نوع، وجود یک کلمه اختصاری در پیکره و منظور از شمار، تکرار آن در پیکره مذکور است. در این پژوهش، هم نوع و هم شمار کلمات اختصاری مد نظر است و اطلاعات آماری مربوط به

¹ About Us

² History

³ type

⁴ token

هر دو مورد ارائه می‌شود. حجم مورد نظر برای هر یک از مقولات شش گانه ۱۰۰۰۰ کلمه بوده و بنابراین برای هر زبان پیکره‌ای شامل ۶۰۰۰۰ کلمه گردآوری شده است. به این ترتیب کل پیکره مربوط به چهار زبان شامل ۲۴۰۰۰۰ واژه می‌باشد.

۵. یافته‌های پژوهش

پس از ایجاد پیکره، کلمات اختصاری موجود در متون شناسایی و شمارش شدند. در پیکره فارسی ۱۷ نوع کلمه اختصاری با شمار ۷۶ مورد به کار رفته است. چهار کلمه اختصاری «ره» (رحمه‌الله‌علیه) با شمار ۱۴، «ه. ش.» (هجری شمسی) با شمار ۱۲، «ناجا» (نیروی انتظامی جمهوری اسلامی ایران) با شمار ۶، و «ع» (علیه‌السلام) با شمار ۵ پربسامدترین کلمات اختصاری پیکره فارسی هستند.

در پیکره انگلیسی ۳۳۰ نوع کلمه اختصاری با شمار ۱۴۴۹ مورد مشاهده شد. چهار کلمه اختصاری پرشمار در پیکره انگلیسی عبارتند از: US (United States) با شمار ۲۵۰ مورد، p.m. (post meridiem) با شمار ۵۰، e (electronic) با شمار ۵۰، و a.m. (ante meridiem) با شمار ۳۵.

مرور وبگاه‌های مربوط به زبان عربی نشان داد تنها چهار کلمه اختصاری در وبگاه‌های مشاهده شده به کار رفته است. چهار کلمه مورد نظر عبارتند از: ق.م. (قبل از میلاد) با شمار ۲۱، ت (تلفن) با شمار ۱۰، کم (کیلومتر) با شمار ۶ و الخ (الی آخریه) با شمار ۴. به این ترتیب شمار کلمات اختصاری پیکره عربی ۴۱ مورد می‌باشد.

در وبگاه‌های زبان ترکی استانبولی نیز ۶۸ نوع کلمه اختصاری با شمار ۳۷۹ مشاهده شد. چهار کلمه اختصاری پرشمار در پیکره ترکی به شرح زیرند: YETKM (Yunus Emre Türk Kültür Merkezi) با شمار ۱۱، Hz (Hazret) با شمار ۸، ODTÜ (ÜOrta DuGu Teknik Üniversitesi) با شمار ۶، و TSK (Türk Silâhli Kovvetleri) با شمار ۴. جدول زیر اطلاعات آماری مربوط به نوع و شمار کلمات اختصاری در کل پیکره را نشان می‌دهد.

جدول ۲. اطلاعات آماری نوع و شمار کلمات اختصاری پیکره

شمار	نوع	زبان	
۱۴۴۹	۳۳۰	انگلیسی	۱
۳۷۹	۶۸	ترکی	۲
۷۶	۱۷	فارسی	۳
۴۱	۴	عربی	۴
۱۹۴۵	۴۱۹	مجموع	

بر اساس اطلاعات جدول فوق، از نظر تعداد کلمات اختصاری (نوع) انگلیسی در رتبه اول و ترکی استانبولی در جایگاه دوم است. فارسی و عربی نیز با فاصله‌ای نسبتاً زیاد در جایگاه سوم و چهارم قرار می‌گیرند. مقایسه شمار کلمات اختصاری در مقولات شش‌گانه پژوهش نیز حائز اهمیت است. جدول زیر اطلاعات آماری مربوط به شمار کلمات اختصاری را در مقولات شش‌گانه نشان می‌دهد.

جدول ۳. شمار کلمات اختصاری پیکره در مقولات شش‌گانه

مجموع شمار	فرهنگ	ورزش	تجارت	علوم	سیاست	خبر		
۱۴۴۹	۳۳۰	۲۳۷	۲۱۸	۱۸۱	۲۹۹	۱۸۴	انگلیسی	۱
۳۷۹	۵۶	۴۰	۳۱	۸۸	۷۳	۹۱	ترکی	۲
۷۶	۱۸	۱۲	۹	۱۱	۱۲	۱۴	فارسی	۳
۴۱	۳	۰	۰	۳۲	۶	۰	عربی	۴
۱۹۴۵	۴۰۷	۲۸۹	۲۵۸	۳۱۲	۳۹۰	۲۸۹	مجموع	

مرور ارقام جدول نشان می‌دهد هر یک از زبان‌ها از منظر مقولات و حوزه‌های شش‌گانه رفتاری کم و بیش متفاوت نشان می‌دهند. در پیکره زبان انگلیسی، مقوله فرهنگ بیشترین و مقوله علوم کمترین کلمات اختصاری را به خود اختصاص می‌دهند. مقوله خبر

در انگلیسی نیز آمار پایینی را نشان می‌دهد، در حالی که در پیکره زبان ترکی استانبولی، مقوله خبر بیشترین آمار کلمات اختصاری را به خود اختصاص داده است. تفاوت میان پیکره ترکی استانبولی و انگلیسی محدود به این نیست. برخلاف انگلیسی که در آن فرهنگ پربسامدترین مقوله به شمار می‌آید، در پیکره ترکی استانبولی این مقوله چهارمین مقوله محسوب می‌شود. در پیکره فارسی، فاصله معنی‌داری میان مقولات دیده نمی‌شود، هرچند در این زبان مقوله فرهنگ با ۱۸ مورد کمی از سایر مقولات بالاتر است و از این جنبه شبیه پیکره زبان انگلیسی است. پیکره فارسی از لحاظ پایین بودن بسامد کلمات اختصاری مقوله تجارت با پیکره ترکی استانبولی شباهت دارد. از دیگر نکات قابل تأمل در خصوص جدول بالا این است که در زبان عربی ۳۲ مورد (حدود ۷۸٪) از کل کلمات اختصاری این زبان در وبگاه‌های مقوله علوم مشاهده شده‌اند. این در حالی است که در حوزه‌های خبر، تجارت و ورزش هیچ موردی از کلمات اختصاری دیده نشد. به‌طور کلی می‌توان نتیجه گرفت که در کل پیکره این پژوهش، مقوله فرهنگ بیشترین و مقوله تجارت نیز کمترین شمار کلمات اختصاری را به خود اختصاص داده‌اند.

در این پژوهش به ساختمان کلمات اختصاری از نظر تعداد حروف تشکیل‌دهنده نیز توجه شده است. جدول زیر اطلاعات آماری مربوط به ساختار کلمات مورد بررسی را ارائه می‌دهد.

جدول ۴. تعداد حروف کلمات اختصاری کل پیکره

شمار حروف	یک	دو	سه	چهار	پنج	بیش از پنج
انگلیسی	۱۶	۶۳	۱۷۲	۶۰	۱۷	۲
ترکی	۵	۹	۴۱	۱۱	۱	۰
فارسی	۲	۴	۵	۳	۲	۱
عربی	۱	۲	۱	۰	۰	۰
مجموع	۲۴	۷۸	۲۱۹	۷۵	۲۰	۳
درصد	۵/۷٪	۱۸/۶٪	۵۲/۲٪	۱۷/۸٪	۴/۷٪	۰/۷٪

اطلاعات جدول فوق نشان می‌دهد در سه زبان انگلیسی، ترکی استانبولی و فارسی تعداد کلمات اختصاری سه‌حرفی بیش از موارد دیگر است. زبان عربی در این میان استثنا به شمار می‌آید، هر چند با توجه به آمار بسیار پایین کلمات اختصاری در این زبان نمی‌توان تحلیلی در این زمینه ارائه داد. با توجه به مجموع اطلاعات، کلمات اختصاری سه‌حرفی بیش از نیمی از کل کلمات اختصاری کل پیکره را تشکیل می‌دهند. کلمات دو‌حرفی و چهارحرفی نیز رتبه‌های بعدی را به خود اختصاص می‌دهند. کلمات اختصاری بیش از پنج کلمه بسیار معدودند و برای مثال کلمه اختصاری «هوانیروز» در پیکره فارسی نوعی استثنا به شمار می‌آید.

۶. نتیجه‌گیری

در این پژوهش پیکره‌ای شامل ۲۴۰۰۰۰ واژه را در وبگاه‌های اینترنتی چهار زبان فارسی، انگلیسی، عربی و ترکی استانبولی تهیه کرده و کلمات اختصاری به کاررفته در آنها را شناسایی کردیم. یافته‌های تحقیق نشان داد زبان انگلیسی هم از نظر تعداد کلمات اختصاری (نوع) و هم از نظر بسامد با فاصله‌ای نسبتاً زیاد در جایگاه نخست قرار دارد و پس از آن زبان ترکی استانبولی رتبه دوم را به خود اختصاص می‌دهد. زبان‌های فارسی و عربی نیز در رتبه‌های بعدی هستند. بررسی ساختار درونی کلمات اختصاری نیز نشان می‌دهد که بیش از نیمی از کلمات اختصاری کل پیکره سه‌حرفی هستند و کلمات دو و چهارحرفی در رتبه‌های دوم و سوم هستند. گرایش به کلمات اختصاری سه‌حرفی بر اهمیت چنین ساختاری در فرایند اختصارسازی تأکید می‌کند. به نظر می‌آید کلمات اختصاری سه‌حرفی در ذهن اهل زبان به الگوی واقعی کلمات طبیعی نزدیک‌ترند و این نکته باید در فرایند واژه‌گزینی مد نظر قرار گیرد. نکته دیگر این که در میان مقولات شش‌گانه بررسی شده در این پژوهش، مقوله فرهنگ دارای کلمات اختصاری بیشتری است. گفتنی است وبگاه‌های مقوله فرهنگ که توسط محققان بررسی شد، اغلب کارکردی هنری داشته‌اند. ناگفته پیداست هنر عرصه خلاقیت است و ساخت کلمه اختصاری نیز فرایندی خلاقانه در واژه‌سازی به حساب می‌آید و این که بسامد کلمات

اختصاری در مقوله فرهنگ کل پیکره بیش از سایر مقولات باشد، امری توجیه شدنی است. از سوی دیگر، مقوله تجارت با ۲۵۸ مورد کلمه اختصاری پایین ترین جایگاه را اشغال می کند. این امر نشان می دهد که وبگاه های مربوط به تجارت مانند وبگاه کارخانه ها، صنایع و بانک ها کمتر از وبگاه های سایر مقولات از کلمات اختصاری استفاده می کنند. همان طور که پیش تر اشاره شد، زبان های انگلیسی و فارسی از نظر رده شناسی زبان های گسسته به شمار می آیند، ترکی استانبولی زبانی پیوندی است و عربی نیز در رده زبان های آمیخته قرار دارد. شاید پیش بینی اولیه مان این باشد که دو زبان انگلیسی و فارسی از آنجا که به یک رده ساخت واژی مشترک (گسسته) تعلق دارند، از نظر فرایندهای واژه سازی مانند اختصار سازی مشابه عمل می کنند، و یا دست کم در مقایسه با ترکی استانبولی که زبانی پیوندی است، شباهت بیشتری با یکدیگر دارند. یافته ها نشان می دهد که چنین نیست و اتفاقاً از این لحاظ انگلیسی با ترکی استانبولی شباهت بیشتری دارد. دو زبان انگلیسی و ترکی استانبولی با این که به رده های ساخت واژی متفاوتی تعلق دارند، به دلیل استفاده از الفبای لاتین رفتاری کم و بیش مشابه در قبال اختصار سازی نشان می دهند. از سوی دیگر، فارسی و عربی از الفبایی استفاده می کنند که در بسیاری از موارد حروف به یکدیگر متصل می شوند و همین امر می تواند به عنوان مانعی برای اختصار سازی مطرح شود. نتیجه نهایی پژوهش حاضر این است که یکی از دلایل کاربرد فراوان کلمات اختصاری در انگلیسی و ترکی استانبولی استفاده از الفبای لاتین و جدا بودن حروف در نوشتار است. به بیان دیگر در فرایند اختصار سازی، عنصر نسبتاً حاشیه ای الفبا نقش پررنگ تری نسبت به رده ساخت واژی ایفا می کند.

منابع

- امینی، امین (۱۳۸۳). هرمرز چهارم: دوازده سال سلطنت، سیزده سال سکه. یزد: شرکت تبلیغاتی طلوع یزد.
- تفضلی، احمد (۱۳۷۵). زبان پهلوی- ادبیات و دستور آن. تهران: معین.
- _____ (۱۳۷۶). تاریخ ادبیات ایران پیش از اسلام. تهران: سخن.
- شقایق، ویدا (۱۳۸۹). مبانی صرف. تهران: سمت.

- Al-Qinani, J. (2007). "Abbreviation and Acronymy in English Arabic Translation". *Meta:Translators' Journal*. Vol. 52. pp. 368-375.
- Dannells, D. (2005). "Recognizing Swedish Acronyms and their Definitions in Biomedical Literature". *Research Report*. Department of Swedish Language. Gotenborg University. Sweden.
- Harley, H. (2004). "Why is the CIA But Not *The NASA? Acronyms, Initialisms, and Definite Descriptions". *American Speech*. 79(4). pp.368-399.
- Lewis, G. (1999). *The Turkish Language Reform: A Catastrophic Success*. Oxford University Press.
- Katamba, F. (1993). *Morphology*. Macmillan Press LTD.
- Reza Goli Famian, A. (2008). "A Cross-Cultural Study of Acronyms in English and Persian". *The Perspectives of Linguistics in the 21st Century*. Jong Yurl Yoon & Kyoung Ae Kim (eds.). Hankook Munhwasa. pp. 471- 479.
- Schwartz, A.S. & M.A. Hearst (2003). "A Simple Algorithm for Identifying Definitions in Biomedical Text". *Pacific Symposium on Biocomputing*. 8. pp. 451-462.
- Taghva, K. & J. Gilberth (1999). "Recognizing Acronyms and their Definitions". *International Journal on Document Analysis and Recognition (IJ DAR)*.1. pp.191-198.
- Yeats, S., D. Bainbridge & I. H.Witten (2000). "Using Compression to Identify Acronyms in Text". In. *Proceedings of the Conference on Data Compression*. IEEE Computer Society, Washington, DC.USA. <<http://www.sigmod.org/publications/dblp/db/conf/dcc/dcc2000.htm>>
- Zipf, G.K. (1949). *Human Behaviour and the Principle of Least Effort*. Cambridge.

A Comparative Analysis of Acronyms in Persian, English, Arabic and Turkish¹

Ali Reza Gholi Famian²

Maryam Sokhanvar³

Received: 2011/09/17

Accepted: 2014/10/20

Abstract

Acronym formation is a major word-formation process where one or a series of letter(s) substitute one word or a longer expression. In this study, we explored the structure and frequency of acronyms in four languages, i.e. Persian, English, Arabic and Turkish. First, the websites related to Islamic Republic of Iran, United States of America, Egypt, and Turkey are selected in six topics including news, politics, sports, business, science, and culture. The pages selected randomly for each language includes six thousand (6000) words, and thereby the research corpus includes twenty four thousand (24000) words. The findings show that English is top in both the number of acronyms and their frequency, Turkish is the second

¹(DOI): 10.22051/JLR.2016.2148

² Assistant professor, Department of Linguistics and Foreign Languages, Payame Noor University, (Corresponding Author);famian@pnush.ac.ir

³ MA. General Linguistics, Department of Linguistics, Islamic Azad University, Ahar Branch mary;2007am@yahoo.com

followed by Persian and Arabic. The study of acronyms' internal structure reveals that more than fifty percent of acronyms have three letters. Considering alphabet and morphological typology, it is maintained that the former plays a more significant role in making and using acronyms.

Keywords: *acronym, Persian, English, Arabic, Turkish*