

نقش پیکره‌های رایانشی در معادل‌یابی و انتخاب حوزه‌های موضوعی در فرهنگ دوزبانه اصطلاحات اسلامی روسی-فارسی

دکتر جواد اکبری

دکترای آموزش زبان روسی، دانشگاه تربیت مدرس،

تهران، ایران

j.akbari@modares.ac.ir

دکتر محمدرضا محمدی ©

دانشیار گروه زبان روسی، دانشگاه تربیت مدرس،

تهران، ایران

mrmoham@modares.ac.ir

واژه‌های کلیدی: زبان‌شناسی رایانشی، معادل‌یابی، فرهنگ‌نگاری پیکره‌ای، اصطلاحات اسلامی.

طبق تعریفی که در فرهنگ زبان‌شناسی و آواشناسی آکسفورد ارائه شده، «پیکره»^۱، مجموعه‌ای از متون نوشتاری یا گفتارهای آوانویسی شده است که از آن می‌توان برای توصیف پدیده‌های زبان‌شناختی یا به‌عنوان وسیله‌ای برای تأیید فرضیه‌های مربوط به یک زبان استفاده کرد. [Crystal, ۲۰۰۸: ۱۱۷]

در این مقاله منظور ما از پیکره، نوع الکترونیکی آن است. متون یا گفتارهای آوانویسی‌شده این پیکره‌ها در پایگاه‌های داده رایانشی ذخیره و به‌وسیله انواع نرم‌افزارهای تحلیل پیکره بررسی می‌شوند. «امروزه پیکره‌های کامپیوتری می‌توانند شامل میلیون‌ها واژه باشند که به‌وسیله برچسب‌دهی (اضافه کردن برچسب‌های شناسایی و طبقه‌بندی واژگان) غنی می‌شوند و همین برچسب‌ها عاملی برای بررسی این پیکره‌ها هستند». [McArthur, ۱۹۹۲: ۲۶۵]

پیکره‌های رایانشی در پژوهش‌ها و فعالیت‌های زبانی، نقش پررنگی ایفا می‌کنند و ابزارهای کلیدی هستند که در خدمت شاخه‌های مختلف زبان‌شناسی قرار می‌گیرند و مسیر رسیدن به اهداف زبان-

^۱.corpora / corpus

شناختی را تسهیل می‌کنند. فرهنگ‌نگاری به‌عنوان شاخه‌ای از زبان‌شناسی کاربردی از این قاعده مستثنا نیست و امروزه پیکره‌ها در مراحل مختلف تدوین فرهنگ لغت نقش اساسی ایفا می‌کنند. از جمله کاربردهای مختلف پیکره‌ها در تدوین فرهنگ لغت می‌توان به نمونه‌های زیر اشاره کرد:

۱. استفاده از پیکره‌ها برای استخراج مثال‌هایی از کاربرد واقعی زبان؛

۲. ارائه و تفکیک معانی مختلف واژگان؛

۳. ارائه اصطلاحات و تعبیرات؛

۴. ارائه اطلاعات دربارهٔ بسامد استفاده از واژه؛

۵. ارائه جملات شاهد؛

۶. ارائه اطلاعات سبک‌شناسی واژگان؛

۷. ایجاد امکان تدوین فرهنگ‌ها و اصطلاح‌نامه‌های تخصصی؛

۸. امکان معادل‌یابی اصطلاحات با استفاده از پیکره‌های موازی [عاصی، ۱۳۹۱].

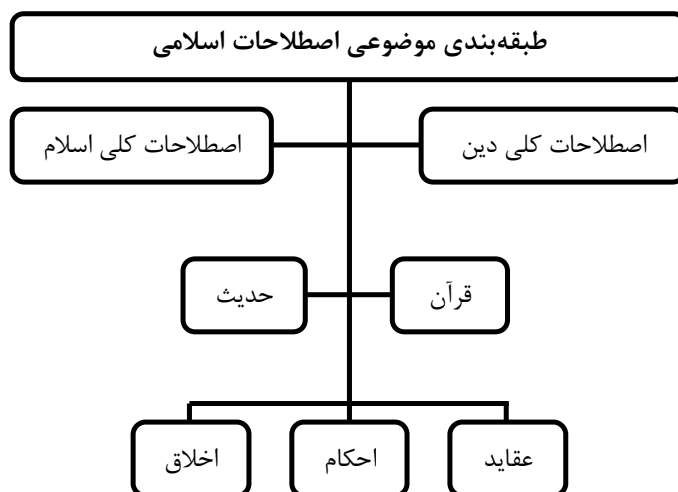
از کاربردهای مختلف پیکره‌ها در تدوین فرهنگ لغت، در این مقاله به بررسی نقش پیکره‌ها در تعیین حوزه‌های موضوعی فرهنگ لغت (تعیین مدخل‌ها و زیرمدخل‌ها) و معادل‌یابی می‌پردازیم. یکی از اصلی‌ترین مشکلات فرهنگ‌نویسان در تدوین فرهنگ لغت، انتخاب حوزه‌های موضوعی آن است، چنانچه ساموئل جانسون، نگارندهٔ فرهنگ استاندارد زبان انگلیسی در قرن ۱۸ میلادی، در طرح توجیهی تدوین فرهنگ لغت خود، در صدر مشکلات خود در تدوین فرهنگ لغت، به مشکل انتخاب مدخل‌ها اشاره می‌کند. [Johnson, ۱۷۴۷: ۴]

در بررسی استفاده از پیکره‌ها در انتخاب حوزه‌های موضوعی فرهنگ لغت اصطلاحات اسلامی از نتایج استفاده می‌شود که حاصل بررسی پیکره‌ای از متون عمومی اسلامی متشکل از ۱۶ جلد کتاب (۱۰ عنوان کتاب با حجم بالغ بر یک میلیون واژه) به زبان فارسی است که توسط نگارندهٔ مقاله جمع‌آوری شده است.

استفاده از پیکره‌ها در تعیین مدخل‌ها و زیرمدخل‌های فرهنگ لغت را می‌توان در دو مرحله بررسی کرد: مرحلهٔ تدوین و مرحلهٔ به‌روزرسانی.

در مرحلهٔ تدوین فرهنگ لغت برای انتخاب مدخل‌ها می‌توان از «فهرست بسامدی واژگان»^۱ به‌عنوان یکی از خروجی‌های بررسی پیکره‌ها به‌وسیلهٔ نرم‌افزارهای تحلیل پیکره استفاده کرد. [Summers, ۲۰۰۵: ۲] به‌منظور استفاده از فهرست بسامدی واژگان در تعیین حوزه‌های موضوعی فرهنگ لغت اصطلاحات اسلامی پس از بررسی فهرست بسامدی پیکرهٔ متون اسلامی، ۳۰۰ واژهٔ دینی پربسامد استخراج و از بین واژگان این کتاب‌ها، موضوعات هر یک از واژگان تعیین شده و سپس بر اساس پراکندگی موضوعی واژگان پربسامد، طبقه‌بندی موضوعی اصطلاحات اسلامی به دست آمده است:

^۱.word list by frequency



برای انتخاب زیرمدخل‌های فرهنگ لغت نیز می‌توان از بررسی هم‌آیندها^۱ به‌عنوان نوع دیگری از انواع نتایج بررسی پیکره‌ها به‌وسیله نرم‌افزارهای تحلیل پیکره استفاده کرد. [Summers, ۲۰۰۵: ۳]

امروزه با تحلیل پیکره‌هایی که نماینده مناسبی از زبان باشند، می‌توان فهرست هم‌آیندهای واژگان را به‌سرعت استخراج و در فرهنگ به‌عنوان زیرمدخل استفاده کرد؛ به‌عنوان مثال می‌توانیم به هم‌آیندهای واژه نماز اشاره کنیم که نگارنده مقاله از بررسی پیکره خود به دست آورده است: ~ جماعت، ~ جمعه، ~ باطل، ~ اول وقت، اقامه ~، ترک ~...

در مرحله به‌روزرسانی فرهنگ لغت، می‌توان با تدوین پیکره‌هایی که شامل متون جدیدتری باشند، واژگانی را شناسایی کرد که به‌تازگی وارد زبان شده‌اند یا استفاده از آن‌ها در زبان افزایش یافته است و آن‌ها را به‌عنوان مدخل جدید وارد فرهنگ لغت کرد؛ اما استفاده از پیکره‌ها در به‌روزرسانی مدخل‌ها و زیرمدخل‌ها به این نمونه محدود نمی‌شود. به تعریفی که دکتر عاصی از پیکره‌ها ارائه می‌دهد، نگاهی می‌اندازیم: «پیکره ممکن است به‌شکل برگه‌های یادداشت یا پرونده‌های رایانه‌ای شامل متن‌های کامل یا گزیده‌هایی از آن‌ها یا بخش‌های پیوسته‌ای از متون یا گزیده‌ای از نقل‌قول‌ها و نکات و حتی فهرست‌های واژگانی باشد.» [عاصی، ۱۳۷۹: ۴۸۶]

مطابق با این تعریف از پیکره‌ها، چنانچه فهرست‌های واژگانی را نیز از انواع پیکره‌های زبانی محسوب کنیم و استفاده از پیکره‌ها در تعیین مدخل‌های فرهنگ لغت را نیز محدود به فرهنگ‌های چاپی ندانیم و اگر در فرهنگ‌های الکترونیکی، سامانه فرهنگ لغت به نحوی طراحی شده باشد که بتواند

^۱. collocation

جست‌وجوهای کاربران و رفتار آن‌ها در استفاده از فرهنگ را ذخیره کند، پیکره‌ای متشکل از جست‌وجوهای کاربران در فرهنگ لغت به دست می‌آید. از نتایج تحلیل این پیکره می‌توان در به‌روزرسانی فرهنگ لغت بهره جست. با استخراج لیست بسامدی جست‌وجوی کاربران، چنانچه نتیجه جست‌وجوی آن‌ها در لیست مدخل‌های فرهنگ لغت وجود نداشته باشد، فرهنگ‌نگار می‌تواند به تکمیل هر چه بیشتر این مدخل‌ها بپردازد که کاربران توجه ویژه به آن می‌کنند و اگر جست‌وجوهایی که بسامد بالایی دارند در لیست مدخل‌های فرهنگ لغت وجود نداشته باشند، فرهنگ‌نگار در صورت همسو بودن این مدخل با اهداف فرهنگ لغت می‌تواند نسبت به ورود آن‌ها به فرهنگ لغت اقدام کند. افزایش توجه کاربران به موضوعات جدید می‌تواند نتیجه عوامل مختلفی نظیر رویدادهای خبری، اتفاقات فرهنگی، سیاسی و غیره باشد یا شامل واژگانی شود که به مفاهیم نوظهور اشاره می‌کنند.

پس از اتمام مرحله انتخاب حوزه‌های موضوعی و مدخل‌ها، در مرحله معادل‌یابی اصطلاحات می‌توان از «پیکره موازی»^۱ به‌عنوان ابزاری برای معادل‌یابی مدخل‌های انتخاب‌شده استفاده کرد. [Frankenberg-Garcia, ۲۰۰۹: ۶۵] برای تبیین نقش پیکره‌های موازی در معادل‌یابی اصطلاحات اسلامی در این مقاله از پیکره‌ای استفاده می‌شود که شامل ۱۱ عنوان کتاب در حوزه‌های موضوعی ذکر شده به زبان فارسی همراه با ترجمه روسی آن‌ها (با حجم کلی بالغ بر ۱.۵ میلیون واژه) است. به این منظور از نتایج حاصل از بررسی «فهرست‌های واژه‌نما»^۲ برای تبیین معانی مختلف واژه در بافت‌های مختلف و انتخاب معادل مناسب استفاده می‌شود.

۱۰ واژه از صدر فهرست پربسامدترین واژگان اسلامی و معادل آن‌ها به زبان روسی به‌عنوان نمونه‌ای از نتایج حاصل از بررسی پیکره موازی متون اسلامی در مرحله معادل‌یابی در جدول زیر آمده است:

ردیف	واژه فارسی	معادل روسی	بسامد واژه در پیکره
۱.	خدا	Бог	۶۵۳۹
۲.	اسلام	Ислам	۳۶۹۸
۳.	آیه	аят	۲۹۶۵
۴.	قرآن	Коран	۲۹۳۸
۵.	دین	религия	۲۵۱۳
۶.	حضرت	его/ее светлость	۲۲۳۴
۷.	الله	Аллах	۲۱۸۰
۸.	پیامبر	Пророк	۲۱۷۱
۹.	امام	Имам	۱۹۶۳
۱۰.	سوره	сура	۱۷۴۴

^۱.parallel corpus

^۲.concordance

از نتایج حاصل از این تحقیق که برای نخستین بار با استفاده از پیکره‌های زبان فارسی و روسی به صورت موازی به دست آمده است، می‌توان در نمونه‌های زیر استفاده کرد:
_تبیین و تدوین سرفصل درس ترجمه متون اسلامی در مقطع کارشناسی زبان روسی؛
_تهیه فرهنگ تخصصی دوسویه روسی به فارسی، اصطلاحات اسلامی پرکاربرد برای مخاطبین فارسی‌زبان و روسی‌زبان.

References

منابع

۱. عاصی، مصطفی. تدوین و کاربرد پیکره‌های زبانی در آموزش زبان و فرهنگ‌نگاری (سخنرانی)، پژوهشکده علوم انسانی و مطالعات فرهنگی، ۱۳۹۱.
۲. عاصی، مصطفی، از پیکره زبانی تا زبان‌شناسی پیکره‌ای، مجموعه مقالات دانشگاه علامه طباطبایی، ۱۳۷۹، شماره ۹۱، ص. ۴۸۴-۴۹۵.
3. McArthur, Tom. (ed.). *The Oxford Companion to the English*. Oxford & New York: Oxford University Press. 1992.
4. Crystal, David. *A Dictionary of Linguistics and Phonetics*. – 6th ed. Oxford: Blackwell. 2008.
5. Johnson, Samuel. *The plan of a dictionary of the English language: addressed to the Right Honourable Philip Dormer, Earl of Chesterfield*. London: Printed for J. and P. Knapton, T. Longman, and T. Shewel. 1747.
6. Summers, Della. *Corpus lexicography – The importance of representativeness in relation to frequency*. 2005. Available from: <<http://www.pearsonlongman.com/dictionaries/pdfs/Corpus-lexicography.pdf>> [Accessed on February 13, 2019]
7. Frankenberg-Garcia, A. Compiling *and using a parallel corpus for research in translation*. Babel: international journal of translation, XXI (1). 2009. pp. 57-71.