

کاربرد پیکره‌های متنی در آموزش زبان آلمانی

حسن پروان*

مربی، گروه زبان آلمانی، دانشکده زبانها و ادبیات خارجی دانشگاه تهران،
تهران، ایران

حسین سرکار حسن خان**

استادیار، گروه زبان آلمانی، دانشکده زبانها و ادبیات خارجی دانشگاه تهران،
تهران، ایران

(تاریخ دریافت: ۹۶/۰۹/۱۰، تاریخ تصویب: ۹۷/۰۶/۲۶، تاریخ چاپ: مهر ۱۳۹۷)

چکیده

در یادگیری زبان بیگانه، زبان‌آموز همواره با چالش‌های مختلفی دست و پنجه نرم می‌کند و به‌ندرت قادر است بدون یاری گرفتن از افراد صاحب‌نظر، پیشرفت قابل‌توجهی در مهارت‌های زبانی کسب کند. مقاله پیش‌رو با در نظر گرفتن این محدودیت، به معرفی پیکره‌های زبانی و چگونگی استفاده از آنها در امر آموزش زبان می‌پردازد. پیکره متنی امروزه به مجموعه‌ای از متون گفتاری و نوشتاری دیجیتال شده گفته می‌شود که از نمونه‌های زبانی جمع‌آوری شده از حوزه‌های مختلف تشکیل شده و در پژوهش‌های زبانی مورد استفاده پژوهشگران قرار می‌گیرند. قصد نگارندگان البته این نیست که با استفاده از پیکره به بررسی تحولات زبانی بپردازد، بلکه سعی بر این است تا نشان دهد، زبان‌آموزان نیز می‌توانند با استفاده از پیکره‌های متنی و مستقل از دیگران بر بسیاری از مشکلات مربوط به یادگیری زبان غلبه کنند. برای تبیین این مسئله، روند تولید یک پیکره، شیوه جستجو در آن با استفاده از پرس‌وجوی پیکره و حل و فصل برخی از مشکلات زبانی شرح داده می‌شوند.

واژه‌های کلیدی: پیکره متنی، زبان پرس‌وجوی پیکره، زبانشناسی پیکره‌ای، آموزش زبان، صفت تفصیلی.

* نویسنده مسئول: E-mail: hparvan@ut.ac.ir

** E-mail: hassankhan@ut.ac.ir

۱- مقدمه

زبان پیوسته دستخوش دگرگونی است؛ خواه این دگرگونی در اثر مراوده و همنشینی با فرهنگ‌ها و زبان‌های بیگانه باشد و خواه در اثر نیاز زمانه به خلق واژگان و مفاهیم نو یا دخل و تصرف در معانی آنچه در واژه‌نامه‌ها یافت می‌شود. علت این امر نیز روشن است: پدیدآمدن شرایط و چارچوب‌های جدید و فناوری‌های نوین یا تراوش‌های ذهنی در حوزه اندیشه‌ورزی، که نیاز به نوآوری در زبان را ایجاب یا تشدید می‌کنند. بدیهی است که همراه شدن با این دگرگونی‌ها و کاربست دستاوردهای جدید در حوزه زبان، شاید برای کاربران زبان مادری امری ساده باشد، اما برای یادگیرندگان زبان بیگانه، اگر نه غیرممکن، به یقین بسیار طاقت‌فرسا و دشوار خواهد بود.

هدف مقاله پیش روی، البته این نیست که چرایی، چگونگی و میزان این‌گونه تحولات را مورد مذاقه و بررسی قرار دهد، بلکه به طرح این مسئله می‌پردازد، که آیا برای افرادی که به طریقی به آموختن زبان بیگانه مشغول‌اند، ابزار و لوازمی وجود دارد که زبان‌آموز با کاربست آن‌ها، در مراحل پیشرفته‌تر آموزش زبان، از وابستگی خود به دیگرانی نظیر معلم و استاد بکاهد؟ به اعتقاد نگارنده، چنین ابزاری قطعاً وجود دارد و می‌توان با استفاده از فناوری‌های نوین، نظیر رایانه و بدون اینکه نیازی به مراجعه به دیگران باشد، مسائل و مشکلاتی را که در حوزه آموزش زبان‌های خارجی گریبانگیر زبان‌آموز می‌شود، برطرف کرد. بدیهی است که

پیماد این‌گونه وابستگی، صرف هزینه و زمان است و همین امر، به‌تنهایی اهمیت بسزای طرح راهکاری را که در این مقاله بیان می‌شود، توجیه می‌کند. در هر صورت، نگارندگان قویاً بر این باور است که کمترین تأثیر منفی غفلت از به‌کارگیری ابزاری نظیر رایانه، به‌ویژه در سال‌های اخیر، طولانی‌تر شدن فرایند آموزش و پژوهش و بروز یکنواختی در شیوه‌های آموزشی است. چنان‌که در عنوان مقاله به چشم می‌آید، جستار حاضر به بررسی یکی از کاربردهای پیکره‌های متنی خواهد پرداخت. پیکره، به مجموعه‌ای از متون مختلفی گفته می‌شود که از اشتراکاتی برخوردارند. اما آنچه بیش از همه، به امر پژوهش در این متون جذابیت و هیجان می‌بخشد، نه محتوای پیکره، بلکه دسترسی به امکاناتی است که نرم‌افزار مدیریت‌کننده این پیکره‌ها برای جست‌وجو و یافتن عبارات و ساختارهای مورد نظر زبانی و تولید داده‌های آماری در اختیار کاربران می‌گذارد. بدیهی است که مقدمه ورود به چنین بحثی، نگاهی اجمالی به مقوله‌هایی نظیر تعریف پیکره، انواع آن و ذکر مطالبی در زمینه کاربردهای مختلف پیکره متنی است. اما، بخش عمده و بنیادین این مقاله، مطالبی را در باب آنچه اصطلاحاً زبان

پرس‌وجوی پیکره^۱ نامیده می‌شود، بیان می‌کند و به تشریح چگونگی استفاده از مجموعه برچسب‌هایی خواهد پرداخت که در دو دانشگاه اشتوتگارت و توینگن آلمان^۲، در توصیف انواع واژه و گروه‌واژه تهیه و تدوین شده‌اند.

۲- بحث و بررسی

پیکره^۳ متنی^۳ به مجموعه‌ای از متن‌ها یا تکه‌متن‌هایی گفته می‌شود که با در نظر گرفتن برخی معیارهای زبان‌شناختی و به منظور استفاده در پژوهش‌های تجربی زبان گردآوری شده‌اند (شیرر، ۲۰۱۴، ص. ۳). مقصود از متن، در تعاریف موجود، نه تنها تولیدات نوشتاری، بلکه محصولات زبان‌گفتار نیز هست. این البته تعریفی است که کمابیش دیگرانی نظیر مک‌انری و ویل‌سن (۲۰۰۱، ص. ۲۹) نیز ارائه داده‌اند. علاوه بر نکات فوق، برای پیکره^۴ متنی ویژگی‌های دیگری نیز ذکر می‌شود؛ برای مثال، بنابر تعریف مک‌انری و ویل‌سن یک متن یگانه و منفرد را نمی‌توان پیکره^۵ متنی نامید (همان‌جا). لازم است پیکره حداقل از دو متن یا بیشتر تشکیل شده باشد. یا بنابر تعاریفی که لم‌نیتسر و سینس‌مایستر (۲۰۱۰، ص. ۴۰)، شیرر (۲۰۱۴، ص. ۴)، کندی (۱۹۹۸، ص. ۳) و جمعی دیگر از صاحب‌نظران ارائه می‌دهند، از جمله ویژگی‌های پیکره‌های متنی، دیجیتالی بودن این‌گونه متن‌هاست. بدیهی است که در روزهای نخستین شکل‌گیری این شاخه از زبان‌شناسی، تولید و استفاده از پیکره‌های الکترونیکی، از جمله به‌علت فراگیر نبودن امکانات رایانه‌ای مرسوم نبوده است، اما امروزه پردازش غیرالکترونیکی داده‌ها در این حوزه در عمل منسوخ شده و تنها در مواردی که آثار دست‌نویس قدیمی محتوای پیکره‌ها را تشکیل می‌دهند، یا بحث هزینه مطرح است، به همان سبک و سیاق قدیم رفتار می‌شود. پیکره‌های متنی چنانکه پیشتر نیز گفته شد، با در نظر گرفتن معیارهایی گردآوری شده‌اند، کارمین شیرر در کتاب *زبان‌شناسی پیکره‌ای* مقبولیت یا به‌عبارتی مشروعیت پیکره^۶ متنی را مشروط به رعایت ۴ اصل می‌داند که عبارت‌اند از: نمایندگی (یا قابلیت تعمیم)^۴، ثبات^۵ داده‌ها (به‌نحوی که بعد از جمع‌آوری و غربال داده‌ها و در نهایت تثبیت پیکره، دخل و تصرفی در

۱. Corpus Query Language - CQL

۲. Stuttgart-Tübingen-Tagset

۳. Textkorpus (en: text corpus)

۴. Repräsentativität

۵. Beständigkeit

ترکیب آن انجام نپذیرد)، ظرفیت (حجم)^۱ و محتوا^۲ (شیر، ۲۰۱۴: ۱۰-۵). وی معتقد است از آنجا که ایجاد پیکره‌های متنی، با هدف بررسی کارکردهای زبانی انجام می‌پذیرد، پیکره ایجاد شده بایستی معرف و نشان‌دهنده کلیت حوزه پژوهشی مورد نظر باشد. این اصل همانی است که وی از آن با عنوان نمایندگی (یا قابلیت تعمیم) یاد می‌کند. بررسی زبان قشر جوان جامعه، تهیه پیکره‌ای از مصادیق گفتاری و نوشتاری همین نسل را می‌طلبد، نه گروه‌های سنی مختلف (همان‌جا: ۶). مقصود شیر از محتوا نیز در وهله نخست، تصمیم‌گیری در مورد نحوه جمع‌آوری متن‌ها و طرح این سؤال است که آیا انتخاب متن‌ها باید برحسب تصادف باشد یا هدفمند و حساب شده. وی به درستی معتقد است که تهیه پیکره‌ای از آثار گوته که در قرون ۱۸ و ۱۹ میلادی خلق شده‌اند، برای بررسی اشکال التزامی^۳ در زبان آلمانی معاصر کار عاقلانه‌ای نیست (همان‌جا: ۱۰).

مکانری و ویل‌سن نیز با بیان این مسئله که پیکره در حوزه زبان‌شناسی مدرن، تنها مجموعه‌ای از متن‌های جمع‌آوری شده نیست، بلکه ویژگی‌های دیگری را نیز می‌طلبد، از موارد چهارگانه زیر به‌مثابه ملزومات پیکره نام می‌برند: نمونه‌برداری و نمایندگی (قابلیت تعمیم)^۴، محدودیت حجم^۵، داشتن قالبی که از طریق رایانه قابل پردازش باشد^۶ و قابلیت عرضه شدن به صورت یک مرجع استاندارد^۷ (مکانری/ویل‌سن، ۲۰۰۱، ص. ۳۲-۲۹). تأکید مکانری و ویل‌سن بر نقش داده‌پردازی الکترونیکی، اصلی‌ترین وجه اختلاف میان این دو نظر است.

اما انواع پیکره کدام‌اند؟ در کنار دسته‌بندی‌های موردی پیکره‌ها که از سوی گروه‌های پژوهشی مختلف و به‌منظور معرفی و تبیین محصولات آن‌ها عرضه شده^۸، یا تقسیم‌بندی‌هایی که اولویت رشته‌های علمی خاصی نظیر ترجمه‌شناسی، آموزش زبان یا ادبیات را مدنظر قرار می‌دهند و در آن‌ها به بررسی انواعی از پیکره‌ها نظیر پیکره‌های موازی^۹، تطبیقی (مقایسه‌ای)^{۱۰}

۱. Größe

۲. Inhalt

۳. Konjunktiv

۴. Sampling and representativeness

۵. Finite size

۶. Machine-readable form

۷. A standard reference

۸. برای مثال سایت پروژه لغتنامه دیجیتالی زبان آلمانی (DWDS) از تنها سه نوع پیکره جامع، رسانه‌ای (Zeitungskorpora) و خاص استفاده می‌کند (<https://www.dwds.de/d/korpora>).

۹. Parallelkorpus

۱۰. Vergleichendes Korpus

یا زبان‌آموز^۱ و دادن توضیحات مفصلی درباره آن‌ها بسنده می‌شود، تقسیم‌بندی‌های جامع و فراگیر دیگری نیز وجود دارند که علی‌رغم تفاوت‌هایی در نام‌گذاری، شباهت‌های زیادی به یکدیگر دارند. از جمله این دسته‌بندی‌ها می‌توان به آنچه در نوشته‌های گرم‌کندی (۱۹۹۸، ص. ۱۹-۲۲)، جنیفر پیرسون (۱۹۹۸، ص. ۴۸-۴۳)، ان اوکیف و مایکل مک‌کارتی (۲۰۱۰، ص. ۲۵-۲۰)، نیلادری سینخار دیش (۲۰۱۰، ص. ۳-۶) و دیگران آمده است، اشاره کرد. یکی از کامل‌ترین تقسیم‌بندی پیکره‌های متنی در کتاب *زبان‌شناسی پیکره‌ای* اثر کارمن شیرر ترسیم شده است. وی در کتاب مورد نظر، ابتدا به معیارهایی که در ایجاد دسته‌بندی مورد توجه قرار گرفته‌اند، اشاره کرده و سپس پیکره‌ها را برحسب آن معیارها، به انواع دوگانه‌ای تقسیم می‌کند. شیرر در پایان فصل مربوط، به انواع پیکره‌ها و در جمع‌بندی مطالب، کلیت پیکره‌ها را برحسب دو مقوله کلی شکل و محتوا نیز تقسیم‌بندی می‌کند (شیرر، ۲۰۱۴: ۳۱).
محتوای جدول زیر برگرفته از همین اثر است (همان‌جا: ۱۷):

انواع پیکره	معیار دسته‌بندی
پیکره‌های دیجیتالی و غیردیجیتالی (یا کاغذی) ^۲	بستر ذخیره‌سازی:
پیکره‌های کلی و جزئی ^۳	سلسله‌مراتبی:
پیکره‌های متن کامل و نمونه‌متن ^۴	کامل بودن متن‌ها:
پیکره‌های ایستا و پایشگر ^۵	چگونگی استمرار پیکره:
پیکره‌های نوشتاری و گفتاری ^۶	گفتاری یا شنیداری بودن پیکره:
پیکره‌های نشانه‌گذاری شده و بدون نشانه‌گذاری ^۷	شیوه آماده‌سازی:
پیکره‌های معاصر و تاریخی ^۸	مقطع زمانی تحت پوشش:
پیکره‌های تک‌زبانه و چندزبانه ^۹	تعداد زبان‌ها:
پیکره‌های مرجع و خاص ^{۱۰}	محدوده اعتبار:

۱. Lernerkorpora
۲. Computerlesbares Korpus – nicht computerlesbares Korpus (digitalisiertes – nicht digitalisiertes)
۳. Gesamtkorpus (Nationalkorpus; Referenzkorpus) – Teilkorpus
۴. Volltextkorpus – Probenkorpus
۵. Statisches Korpus – Monitorkorpus
۶. Korpus der geschriebenen Sprache – Korpus der gesprochenen Sprache
۷. Annotiertes Korpus – nicht annotiertes Korpus (پیکره‌هایی که شامل توضیحات دستوری‌اند)
۸. Korpus der Gegenwartssprache – historisches Korpus (synchrone – diachrone Korpora)
۹. Einsprachiges Korpus – mehrsprachiges Korpus
۱۰. Referenzkorpus – Spezialkorpus

پیکره‌ها از کاربردهای مختلفی برخوردارند. از این کاربردها در آثار متعددی سخن به میان آمده، مکائری و ویلسن برای مثال فصلی از کتاب (درآمدی بر) زبانشناسی پیکره‌ای، را به همین موضوع اختصاص داده‌اند و در آن علاوه بر تأکید بر اینکه پیکره‌ها منبعی برای پژوهش‌های میدانی به‌شمار می‌آیند، به تفصیل، به بررسی نقش پیکره‌ها در حوزه‌هایی نظیر: مطالعات واژگانی، دستورزبان، معناشناسی، کاربردشناسی، آموزش زبان، زبان‌شناسی، تحلیل گفتار و غیره می‌پردازند (مکائری/ویل‌سن، ۲۰۰۱: ۱۰۳). در فصل پنجم کتاب راهنمای بین‌المللی زبان‌شناسی پیکره‌ای که شامل سلسله‌مقالاتی در باب همین شاخه از زبان‌شناسی است، تحت عنوان «کاربردها و بهره‌وری از پیکره‌ها» نیز موارد متعددی از کاربردهای پیکره شرح داده می‌شوند. از جمله این مقالات می‌توان به عناوینی نظیر «زبان‌شناسی پیکره‌ای و ساخت‌واژه» اثر هارالد بایان، «زبان‌شناسی پیکره‌ای و گویش‌شناسی» نوشته لیزلوته اندروالد و بندیکت شمترسانی، «پیکره‌ها و ترجمه ماشینی» اثر هارولد سامرس و «پیکره‌ها و ترجمه انسانی» به قلم سیلویا هانزن-شیرا و الکه تایش اشاره کرد (لودلینک/کیتو، ۲۰۰۹).

لم‌نیتسر و سینس‌مایستر نیز در فصل ششم درآمدی بر زبان‌شناسی پیکره‌ای به تفصیل به بررسی کاربردهای مختلف پیکره‌ها پرداخته‌اند و در آن از جمله، به شرح مواردی نظیر نقش پیکره‌ها در زبان‌شناسی رایانه‌ای و لغت‌نامه‌نویسی می‌پردازند (۲۰۱۰، ص. ۱۶۵-۱۲۴). علاوه بر آثار ذکر شده، که در آن‌ها کلیاتی در باب کاربردهای مختلف پیکره‌ها آمده است، جمعی از دانش‌پژوهان، در آثار خود، به بررسی تنها یک وجه از کاربردهای فوق‌الذکر پرداخته‌اند. از جمله این پژوهشگران می‌توان به جوی‌براتو موکرچی (۲۰۰۲) اشاره کرد که در کتاب زبان‌شناسی پیکره‌ای و درس انگلیسی به تبیین نقش پیکره‌ها در آموزش زبان انگلیسی می‌پردازد یا کارین آیچ‌مر و کتاب وی تحت عنوان پیکره‌ها و آموزش زبان.

همان‌طور که از عنوان مقاله نیز پیداست، هدف نگارنده نشان دادن امکانی به مخاطب برای رفع مشکلات زبانی است، بدیهی است، منظور از مخاطب از جمله کلیه کسانی هستند که به یادگیری یکی از زبان‌های خارجی مشغول‌اند؛ همچنین پیش‌تر بیان شد که از ویژگی‌های اصلی پیکره‌های امروزی، دیجیتالی بودن آن‌هاست. به این ترتیب مخاطب برای آموختن شیوه‌ای که در این نوشتار عرضه می‌شود، بی‌نیاز از ابزار پردازش الکترونیکی داده‌ها نظیر رایانه نخواهد بود. نرم‌افزارهای مورد استفاده در حوزه پیکره‌ها را می‌توان به دو دسته کلی نرم‌افزارهای مبتنی بر وب و دسکتاپ^۱ تقسیم کرد. مزیت استفاده از نرم‌افزارهای تحت وب این است که

۱. Webbased- / Desktop-Anwendungen

داده‌های عرضه‌شده در آن‌ها، در بسیاری از موارد، به‌دست افراد خبره و صاحب‌نظر و با رعایت همه‌جوانب مربوط به مقبولیت یا مشروعیت علمی پیکره‌ها تهیه شده و این نرم‌افزارها، علاوه بر رفع مشکلات زبانی، در دادن آمار و ارقامی که مصارف پژوهشی دارند نیز به‌کار می‌آیند^۱. نرم‌افزارهای مبتنی بر وب و دسکتاپ به دو صورت **متن‌باز**^۲ و **تجاری** عرضه می‌شوند. اما ایجاد پیکره در نرم‌افزارهای دسته دوم، اغلب برعهده کاربران است. مزیت این امکان در این است که کاربر با توجه به نیاز خود، متن‌های پیکره را انتخاب می‌کند و به انتخاب و غربال دیگران در نرم‌افزارهای مبتنی بر وب متکی نیست. نگارنده با این پیش‌زمینه به سراغ یکی از نرم‌افزارهای متن‌باز رفته و تلاش می‌کند با ایجاد یک پیکره تک‌زبان و استفاده از مجموعه برچسب‌های اشتوتگارت-تویینگن به بررسی برخی از مشکلات زبانی و دادن راه‌حلهایی برای آن‌ها بپردازد.

قبل از شروع قسمت عملی، بیان این نکته ضروری است که به اعتقاد نگارنده، در این نوشتار، به علت در اولویت نبودن دادن اطلاعات آماری، نیازی به رعایت جزئیات ملزوماتی نظیر «نمایندگی» (یا قابلیت تعمیم) که بایستی در ایجاد پیکره مورد توجه قرار بگیرند، وجود ندارد.

در میان نرم‌افزارهایی که برای مدیریت و انجام امور پژوهشی در پیکره‌ها تولید و عرضه شده‌اند، برخی به علت کمیت یا کیفیت امکانات، از جایگاه ارزشمندتری برخوردارند. عرضه‌شده‌ای امکانی نظیر فرمت‌های مختلف برای واردکردن^۳ داده‌ها یا افزودن ابزار قوی‌تر جهت جست‌وجو در متن پیکره و نیز امکان تولید نمودارها و جداول مربوط به نتایج پژوهش، در مشخص کردن این جایگاه تعیین‌کننده است. نرم‌افزار انتخابی نگارنده برای استخراج واژگان و ساختارهای موردنظر از متن پیکره، TXM نام دارد. ویژگی‌های این نرم‌افزار یونی‌کد^۴ و متن‌باز

۱. پروژه «سیستم جستجو، مدیریت و تحلیل پیکره - COSMAS2» که از سوی «مؤسسه زبان آلمانی - IDS» در مانهایم ایجاد شده و تعداد زیادی از پیکره‌های موجود در آن، بدون پرداخت هزینه در دسترس کاربران است، یا پروژه «لغتنامه دیجیتال زبان آلمان - DWDS» از جمله نرم‌افزارهای تحت وب‌اند. دسترسی به این پروژه‌ها از طریق لینک‌های زیر امکان‌پذیر است:

COSMAS2 (Corpus Search, Management and Analysis System): <http://www.ids-mannheim.de/cosmas2/web-app/DWDS> (Digitales Wörterbuch der deutschen Sprache): <https://www.dwds.de/>

۲. Open Source

۳. Import

۴. با توجه به یونی‌کد (Unicode) بودن TXM، امکان ایجاد پیکره به هر زبانی فراهم است.

که در دهه ۸۰ میلادی در فرانسه تکامل یافته و آخرین نسخه آن (0.7.8) در ماه‌های اخیر در اختیار علاقه‌مندان و پژوهشگران قرار گرفته، بر اساس توضیحات موجود در صفحه اینترنتی پروژه تکستومتری^۱، به اختصار به شرح زیر است:

- قابلیت نصب و اجرا در سه سیستم عامل ویندوز، مک او اس ایکس و لینوکس؛
- امکان جست‌وجو با استفاده از آنچه اصطلاحاً زبان پرس‌وجوی پیکره یا سی.کوال نامیده می‌شود؛
- قابلیت استفاده از نرم‌افزار مکمل TreeTagger برای نشانه‌گذاری واژگان (با استفاده از برجسب‌های تولیدشده در دانشگاه‌های اشتوتگارت و توینگن موسوم به اس.ت.ت.اس)؛
- عرضه پیکره‌های متن‌باز، از پیش آماده‌شده برای برخی از زبان‌ها؛
- و امکان ارسال داده‌ها به ماژول آماری (R) برای پردازش‌های بعدی.

TXM برای نشانه‌گذاری^۲ واژگان از نرم‌افزار TreeTagger استفاده می‌کند. وظیفه TreeTagger آماده‌کردن متن پیکره برای انجام جست‌وجو است. این نرم‌افزار در یک فرایند خودکار و بدون دخالت کاربر^۳، متونی را که قرار است در داخل TXM به‌عنوان پیکره، مورد استفاده قرار بگیرند، بر پایه برجسب‌های اشتوتگارت و توینگن به اجزای کوچک‌تری تقسیم می‌کند. این اجزای کوچک‌تر، قابل مقایسه با انواع واژه یا واحدهای دستوری‌اند و تعداد آن‌ها برای زبان آلمانی به بیش از ۵۰ عدد می‌رسد. بدیهی است که این رقم، با تعداد انواع واژه از نگاه دستورنویسان سنتی تفاوت آشکاری دارد. وظیفه دیگر این نرم‌افزار، مشخص‌کردن صرف‌های مختلف افعال، صفات و اسامی است؛ به‌نحوی که اگر برای مثال مورد جست‌وجو، یافتن صرف‌های مختلف یک فعل باشد، نمایش تمامی صرف‌های موجود فعل موردنظر در پیکره (به شرط موجود بودن) به‌راحتی امکان‌پذیر باشد. شایان ذکر است که علی‌رغم دقت منحصربه‌فرد TreeTagger در تقطیع اجزای جمله، وقوع خطا در تشخیص اجتناب‌ناپذیر است. اما با نظارت کاربر و حذف یافته‌های اشتباه این مشکل برطرف می‌شود. درعین حال می‌توان فایلی^۴ را که TreeTagger تولید کرده، با استفاده از یک نرم‌افزار پردازش متن ویرایش کرد و

۱. <http://textometrie.ens-lyon.fr/files/software/TXM/0.7.8>

۲. Annotation

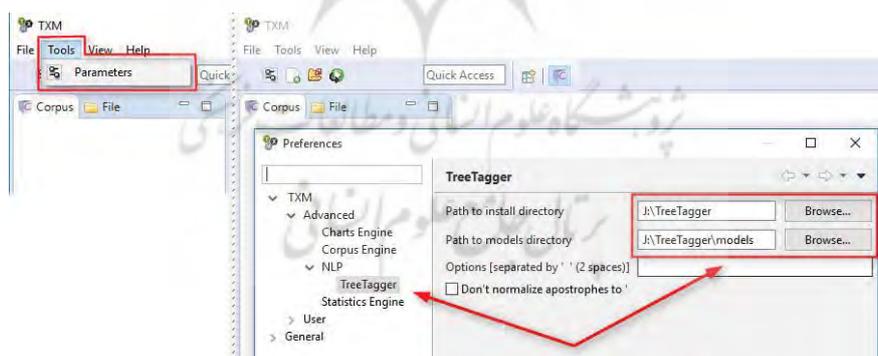
۳. Anwender (= User)

۴. Datei (= File)

اشتباهات را برطرف کرد، اما این فرایند به‌عللی از جمله زمان‌بر بودن و لزوم تسلط کافی به مقوله‌های دستوری به مخاطب مبتدی توصیه نمی‌شود.

نرم‌افزار TreeTagger در دانشگاه اشتوتگارت تهیه شده است؛ بسیاری از زبان‌ها را پشتیبانی می‌کند و به همین علت برای پردازش زبان‌های مختلف، نیاز به فایلی موسوم به فایل پارامتر دارد. این فایل برای هر زبان جداگانه ایجاد شده و مانند TreeTagger^۱ از طریق لینک موجود در پاورقی^۲ قابل دریافت است. مراحل آماده‌سازی نرم‌افزارهای TXM و TreeTagger به اختصار به شرح زیر است:

- دانلود و نصب نرم‌افزار TXM در درایو موردنظر؛
- دانلود TreeTagger و فایل پارامتر از لینک‌هایی که در پاورقی ذکر شده‌اند؛
- خارج کردن پوشه TreeTagger از داخل فایل tree-tagger-windows-3.2.zip، قراردادن آن در درایو موردنظر و ایجاد یک زیرپوشه با عنوان models در داخل آن؛
- خارج کردن فایل german.par از داخل فایل german-par-linux-3.2-utf8.bin.gz، قراردادن آن در زیرپوشه‌ای models که در مرحله قبل ایجاد شد و تغییر نام فایل german.par به de.par؛
- اجرای نرم‌افزار TXM، واردشدن به منوی تنظیمات و واردکردن نشانی پوشه TreeTagger و فایل پارامتر در فیلدهایی که برای این کار پیش‌بینی شده است:



تصویر شماره ۱: منوی تنظیمات TXM

۱. <http://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/data/tree-tagger-windows-3.2.zip>

۲. <http://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/data/german-par-linux-3.2-utf8.bin.gz>

بعد از انجام مراحل فوق، لازم است برای ایجاد پیکره، چند متن در فرمت txt تهیه شود. برای این منظور می‌توان از یک نرم‌افزار ویرایش متن نظیر notepad استفاده کرد، یکی از ویژگی‌های جالب این ویرایشگر ساده، این است که هر متنی به آن ارسال شود، بعد از ذخیره‌سازی یا کپی مجدد، از علائم و فرمت‌های غیرضروری تصفیه می‌شود. تصفیه متن از علائم، نشانه‌ها و فواصل غیرضروری یک گام اساسی در جهت ساخت پیکره‌ای کارآمد است. کیفیت و کمیت متن‌ها، به موضوع پژوهش بستگی دارند. اگر قصد پژوهشگر برای مثال مطالعه در زمینه زبان گفتاری نسل جوان باشد، بدیهی است، اولویت او جمع‌آوری نمونه‌های متن‌های متناسب با این موضوع خواهد بود. یا اگر فرض پژوهش، تحقیق در زمینه ساختارهای نحوی در ادبیات کودک و نوجوان است، لازم است متن‌های جمع‌آوری شده مربوط به همین محدوده موضوعی باشند. در مقاله پیش‌رو همان‌گونه که قبلاً نیز بیان شد، هدف، نشان دادن قابلیت‌های نرم‌افزاری در یاری‌رساندن به زبان‌آموزان آلمانی است، نه رعایت ملزوماتی که در بحث مقبولیت علمی متون پیکره مورد تأکید قرار گرفتند. به همین علت آنچه در انتخاب متن‌ها از نظر نگارنده از اولویت برخوردار است، سادگی آن‌ها، کارآمدی و این ویژگی است که زبان‌آموز بتواند به‌علاوه دقت و ظرافتی که در نگارش متن‌ها به‌کار رفته، از عبارات و ساختارهای موجود در آن‌ها در جایگاه الگوهای قابل اعتماد زبانی استفاده کند.

به اعتقاد نگارنده، همه این ویژگی‌ها تا حد زیادی، در متون استخراج شده از دو وبسایت^۱ آلمانی nachrichtenleicht.de و klexikon.zum.de وجود دارند. سایت نخست، یک سایت خبری است که با هدف ساده‌سازی زبان اخبار سیاسی، فرهنگی و اجتماعی ایجاد شده و سایت دوم نیز به این علت که دایره‌المعارفی برای کودکان و نوجوانان است، از زبان و ساختارهای نحوی ساده‌ای برخوردار است. علاوه‌براین، سایت اخیر از گزینه‌ای برای انتخاب تصادفی مطالب نیز برخوردار است که یکی از شیوه‌های مرسوم در انتخاب متن‌ها برای ایجاد پیکره است. زبان آلمانی مورد استفاده در هر دو سایت، برای بسیاری از زبان‌آموزان آلمانی، به‌راحتی قابل فهم و الگوبرداری است.

مراحل آماده‌سازی پیکره برای پروژه موردنظر این مقاله، به این صورت است که از هر یک از دو سایت نام‌برده، تعداد ۵۰ مقاله کوتاه که به‌طور متوسط از ۱۰ تا ۲۵ خط تشکیل شده‌اند و از موضوعات متنوعی برخوردارند، استخراج و این متن‌ها بعد از انتقال به notepad در فرمت

۱. Webseite (= Website)

txt و با رمزگذاری UTF-8^۱ ذخیره می‌شوند؛ به‌نحوی که در خاتمه دو فایل با نام‌هایی نظیر leicht و klexikon ایجاد شده باشند. نرم‌افزار TXM برای ساخت و نام‌گذاری پیکره، نه از نام خود فایل‌ها، بلکه از نام پوشه‌ای^۲ که دو فایل ایجادشده در آن قرار دارند، استفاده می‌کند. به همین علت لازم است پوشه‌ای در یکی از درایوهای رایانه ایجاد و نام آن متناسب با عنوان پروژه فعلی انتخاب شود.^۳ آدرس این پوشه در مرحله بعد و برای ایجاد پیکره، در مکان مربوط در نرم‌افزار TXM وارد خواهد شد. به‌صورت قاعده‌ای کلی توصیه می‌شود، همیشه یک کپی پشتیبان از کل متن‌های جمع‌آوری‌شده از منابع مختلف، تهیه و در پوشه‌ای مجزا نگهداری شود. مراحل باقی‌مانده از تولید پیکره به شرح زیر است:

- اجرای نرم افزار TXM؛
- انتخاب منوی File > Import > TXT + CSV؛
- در صفحه‌ای باز شده (تنظیمات)، از طریق گزینه ۱، آدرس پوشه‌ای که فایل‌های متنی در آن ذخیره شده‌اند، وارد می‌شود. از طریق گزینه ۲، فقط زبان متن تعیین می‌شود و نیازی به تغییر گزینه‌های دیگر نیست و از طریق گزینه ۳، پیکره متنی، طی فرایندی که زمان آن به حجم متن‌های انتخابی بستگی دارد، ایجاد و نام آن در قسمت چپ رابط کاربر^۴ نرم‌افزار (در فهرست پیکره‌های موجود) ظاهر می‌شود.
- با کلیک بر روی نام پیکره ایجادشده، آیکون‌های مربوط به امکانات نرم‌افزار پدیدار می‌شوند.

در تصویر زیر، آیکونی که در کنارش عدد ۵ درج شده است (☰)، concordance نامیده می‌شود. منظور از این کلمه به اختصار «فهرست یافته‌ها» است. در میحث پیکره‌های متنی، concordance به فهرستی از یافته‌ها (بعد از عمل جست‌وجو) گفته می‌شود که در آن واژه یا گروه‌واژه‌ی مورد جست‌وجو در وسط قرار گرفته و در دو سمت چپ و راست آن، میزان قابل تنظیمی از متون اطراف این واژگان نمایش داده می‌شود (ر.ک بیکر/هاردی/مکانری، ۲۰۰۶:

۱. encoding

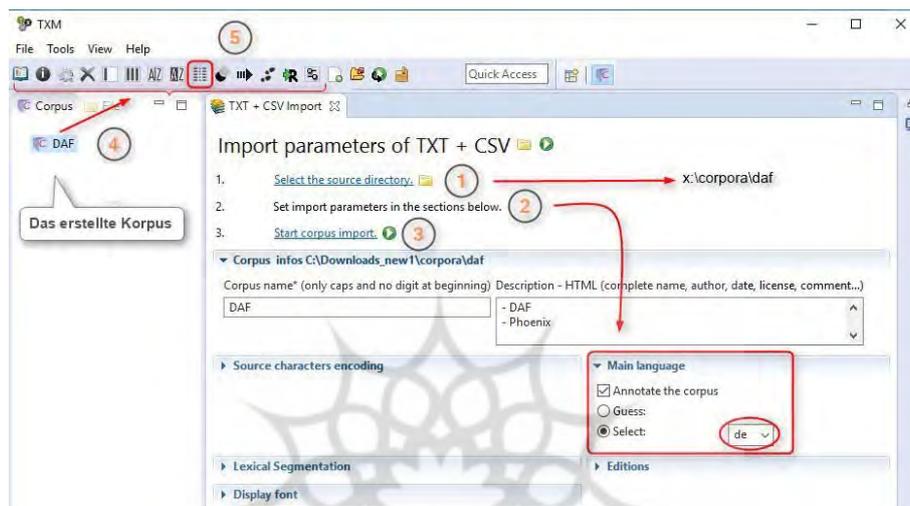
۲. Ordner (= Folder)

۳. برای مثال آدرس فایل‌های متنی می‌تواند c:\corpora\daf برای پروژه مورد نظر این مقاله یا c:\corpora\kjl برای پروژه زبان ادبیات کودک و نوجوان باشد

[daf = Deutsch als Fremdsprache] / [kjl = Kinder- und Jugendliteratur]

۴. Schnittstelle (= Interface)

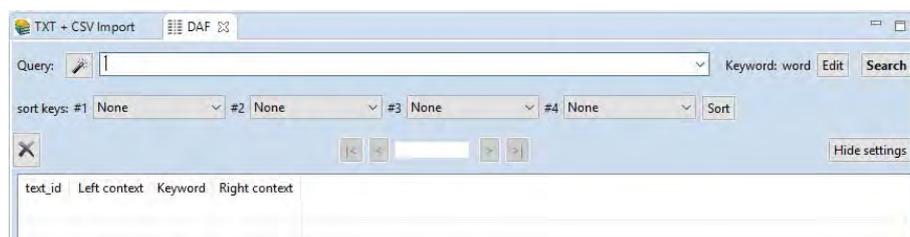
۴۲-۴۳). این فهرست عموماً در جدولی که از حداقل ۳ ستون و ردیف‌هایی (به تعداد کل یافته‌ها) تشکیل شده، ترسیم می‌شود:



تصویر شماره ۲: مراحل ایجاد یک پیکره‌متنی در TXM

LeftContext	Keyword	RightContext
Als der Bösewicht Alberich das Rheingold stiehlt,	hält	er kurz inne, und die Melodie des vorher gesungenen erklingt.
Physiker aus Einsteins Zeit sie zunächst für Unsinn	hielten	. Heute kennt man aber viele Beweise für die Richtigkeit dieser
auf. Er wurde also Herausgeber. Außerdem	hielt	er Vorträge und reiste um die Welt. Das machte er bis
Der Schütze spannt den Bogen. Eine Abzugsvorrichtung	hält	die gespannte Sehne fest. Dann zielt der Schütze wie mit einem
schaffte es, Ägypten stark und unabhängig zu	halten	. Sie wurde die Freundin von Julius Cäsar und hatte auch einen
Tiere wurden aber auch als Schmuck in Parks	gehalten	. Weil sie nicht weglafen, ist das recht einfach. Pfauen

به «فهرست یافته‌ها» (KWIK (keyword in context) نیز گفته می‌شود. با کلیک بر روی آیکن concordance، صفحه‌ای باز می‌شود که در قسمت بالای آن کادری برای جست‌وجو در پیکره و در قسمت پایین آن جدولی برای نمایش یافته‌ها تعبیه شده است:



تصویر شماره ۳: کادر جست‌وجوی متن با استفاده از زبان پرس‌وجوی پیکره در TXM

در TXM نیز مانند همه نرم‌افزارهایی که از زبان پرس‌وجوی پیکره استفاده می‌کنند، برای جست‌وجو، واژه یا برجسب‌ها درون گروه قرار داده می‌شوند:

[word=""]	در این ساختار، واژه موردنظر برای جست‌وجو در داخل علائم نقل‌قول نوشته می‌شود.
[delemma=""]	برای یافتن صرف‌های مختلف یک فعل یا صفت خاص، از ساختار مقابل استفاده می‌شود.
[depos=""]	برای جست‌وجوی مقوله‌های دستوری، برجسب موردنظر در داخل علائم نقل‌قول نوشته می‌شود. فهرست این برجسب‌ها در ضمیمه این مقاله آمده است.

ساختارهای نحوی جست‌وجو با استفاده از زبان پرس‌وجوی پیکره از تنوع زیادی برخوردارند؛ برای مثال می‌توان در کادر جست‌وجو که در سمت چپ آن عبارت (Query) آمده است، از سه عنصر نامبرده به‌صورت ترکیبی استفاده کرد یا از عباراتی برای تعیین فاصله میان عناصر جست‌وجو استفاده کرد. می‌توان با یاری‌گرفتن از ساختارهایی، جست‌وجو را به جمله محدود کرد یا با استفاده از اپراتورهایی نظیر «یا» (I) و «و» (&)، مورد جست‌وجو را به

واژگان بیشتری تعمیم داد. شیوه کاربست این موارد، در خیلی از مقالات یا ویدئوهای آموزشی آمده است.

با این توضیح، شرح مقدمات موردنیاز برای انجام کار پژوهشی با پیکره‌ها خاتمه یافته است و در ادامه به طرح برخی از مسائل مربوط به آموزش زبان و شیوه پاسخ‌یابی با استفاده از پیکره متنی اشاره می‌شود.

مسئله ۱:

زبان‌آموز در صرف صفاتی که میان واژگانی نظیر *viel*, *manch*, *kein* و اسم^۱ قرار می‌گیرند، دچار تردید می‌شود. در توضیحات و مثال‌های ذکر شده در لغتنامه‌ها، غالباً به صرف خود این واژگان بسنده می‌شود و صفتی در کنار آن‌ها آورده نمی‌شود؛ به نحوی که یافتن پاسخ برای زبان‌آموز کار آسانی نخواهد بود.

یافتن پاسخ از طریق پیکره متنی:

A1: [word="viel.*"%c] [depos="ADJA"] [depos="NN"]

A2: [word="(viel|manch|kein).*"%c] [depos="ADJA"] [depos="NN"]

توضیح: با استفاده از ساختار A1 از نرم‌افزار درخواست می‌شود، در پیکره ایجاد شده کل واژگانی را بیابد که چهار حرف اول آن‌ها *viel* است و بلافاصله بعد از آن‌ها، به ترتیب یک صفت صرف شده (برچسب ADJA) و یک اسم (برچسب NN) قرار گرفته باشند. نشانه (%c) بعد از علامت نقل قول به این معنی است که در یافته‌ها هر دو بسامد کوچک و بزرگ‌نویسی واژه *viel* نمایش داده شوند. تفاوت ساختار دوم (A2) با اول فقط در این است که جست‌وجو به کلمه *viel* محدود نمی‌شود، بلکه نرم‌افزار ترکیبات دیگری را نیز خواهد یافت، که در آن‌ها *manch* و *kein* در کنار یک صفت صرف شده و اسم بعد از آن قرار گرفته‌اند:

۱. Nomen

LeftContext	Keyword	RightContext
. Später hat er auch ohne Dean Martin	viele erfolgreiche Filme	gemacht. Zwei besonders bekannte Filme
. Die Internet-Firma sagt: Hacker haben	viele geheime Informationen	von Yahoo gestohlen. Die Hacker haben jeden
. Sie ist so unglaublich, dass selbst	viele große Physiker	aus Einsteins Zeit sie zunächst für Unsinn
niemandem ihre Haare. So tun es auch	viele muslimische Frauen	. Sie zeigen ihre Haare nur anderen Frauen
der Mehrheit. Äthiopien ist die Heimat für	viele verschiedene Pflanzen	und Tiere. Von der Savanne über immergrüne
selbst Geld, aber nicht sehr viel.	Viele arme Kinder	leben nur mit einer Mutter oder einem Vater
in einer armen Familie. Der Staat gibt	vielen armen Familien	Geld, damit sie sich zum Beispiel eine,
. Ein Buch besteht aus unterschiedlich	vielen bedruckten Seiten	. Die zeigen Texte oder Bilder und sind

نتیجه: با نگاهی به نتیجه جست‌وجو می‌توان دریافت: همهٔ صفاتی که میان واژه صرف‌شده viel و اسامی جمع بسته‌شده قرار می‌گیرند، از صرف قوی برخوردارند. گاهی ممکن است در میان یافته‌ها، صفتی با صرف ضعیف نیز پیدا شود؛ برای مثال صرف alt در عبارت "vieles alte Wissen". اما این اشکال نیز در عین استثناء بودن از قاعده‌ای تبعیت می‌کنند^۱. علاوه بر این، زبان آموز با تأمل در اشکال مفرد و جمع اسامی، تفاوت موجود را درمی‌یابد و یقین خواهد کرد که با انتخاب صرف قوی برای صفت، قطعاً به خطا نرفته است.

۱. در canoo.net که در شمار معروف‌ترین سایت‌های اینترنتی در زمینه آموزش دستور زبان آلمانی است، این قاعده آمده است: صفات بعد از دو واژه vieles و vielem اغلب از قواعد صرف ضعیف تبعیت می‌کنند:
<http://www.canoo.net/services/OnlineGrammar/Wort/Adjektiv/Deklinationstyp/Schwankend.html>

مسئله ۲:

در محیط‌های دانشگاهی، زبان‌آموزان علاوه بر یادگیری قواعد زبان و استفاده عملی از آن در حوزه گفتار و نوشتار، اصول و مقدمات ترجمه را نیز می‌آموزند. اصلی‌ترین چالش در بحث ترجمه، ابتدا فهم مطلب زبان مبدأ و سپس استفاده از معادل‌های مناسب در متن ترجمه است. در زبان آلمانی ترکیباتی وجود دارند که از منظر ساختاری یا در زبان فارسی وجود ندارند یا اگر وجود داشته باشند، شیوه بیان آن‌ها به گونه دیگری است. زبان‌آموز در چنین مواردی با این سؤال مواجه است که آیا این ترکیبات در قرینه‌های متنی دیگر هم تکرار می‌شوند یا فقط گوینده یا نویسنده خاصی از آن استفاده کرده است. پاسخ به سؤالاتی از این قبیل، احاطه و اشراف بر کاربردهای زبانی را می‌طلبد. بدیهی است که زبان‌آموز از چنین توانی برخوردار نیست و به همین علت، رجوع به پیکره‌های متنی وی را در یافتن پاسخ به چنین سؤالاتی یاری می‌دهد. از جمله این موارد، عبارت آلمانی "von ... über ... bis" است. سؤال این است که آیا می‌توان به ترکیباتی که شامل این سه کلمه‌اند، به مثابه یک شابلون زبانی نگریست؟ اگر چنین است، آیا می‌توان برای ترجمه آن از معادل ثابتی در زبان فارسی استفاده کرد.

یافتن پاسخ از طریق پیکره متنی:

B1: [word="von|vom"%c] []{0,2} [word="über|übers"] []{0,2} [word="bis"]

B2: [word="vo[nm]"%c] []{0,2} [word="übers?"] []{0,2} [word="bis"]

B3: [word="vo(n|m)"%c] []{0,2} [word="übers?"] []{0,2} [word="bis"]

توضیح: ساختار اول (B1) از نرم‌افزار پیکره می‌خواهد، کل ترکیباتی را بیابد که در آن‌ها اولاً هر سه کلمه von, über و bis وجود داشته باشند و ثانیاً این کلمات در یک فاصله متغیر، در کنار یکدیگر قرار گرفته باشند. این فاصله ($\{0,2\}$) صفر تا دو کلمه تا واژه بعدی است. در ساختار دوم (B2)، علامت سؤال فقط اشاره به حرف ماقبل خود دارد و در عبارت über? bis به این معنی است که اگر در پیکره، علاوه بر کلمه über? bis، کلمه über نیز پیدا شد، مورد جست‌وجو تحقق یافته است. سه عبارت اول در ساختارهای پیشنهادی، از لحاظ نحو اندکی تفاوت دارند ولی نتیجه استفاده از همه آن‌ها یکسان است:

LeftContext	Keyword	RightContext
zu finden . Die Auswahl reicht	von Fußball über Radsport bis	Tennis und bietet nahezu jedem
Spezialitäten sind hier zu haben ,	von Pasta über Pizza bis	zu Fisch- und Fleischgerichten . Seit
viele schöne Rollen . Ich habe schon	vom Griechen über Spanier bis	zum Russen alles gespielt . Aber
, die man sich leisten kann -	vom Privatzimmer über die Ferienwohnung bis	zum Hotel . Auf dem immer beliebteren
und hochinteressant . Die Zeit	von der Antike über das Mittelalter bis	heute bietet einen unerschöpflichen
dürfen . Der Zitatenschatz reicht	von Laotse über Martin Heidegger bis	zu Niklas Luhmann "Kapitel"
Hier besiedeln sie weite Gebiete	von China über die Mongolei bis	nach Sibirien . Sie haben sich dort

نتیجه: یافته‌های مندرج در جدول فوق تنها بخشی از کل یافته‌هاست. کلیت این نتایج بیان‌کننده این مسئله است که ترکیب آلمانی "von ... über ... bis"، در بسیاری از متون گفتاری و شنیداری، مورد استفاده قرار می‌گیرد و به همین علت، یک شابلون زبانی به‌شمار می‌رود. در زبان فارسی، آنچه کاربرد بیشتری دارد، ترکیبات دو قسمتی نظیر «از شیر مرغ تا جان آدمیزاد» است، ولی استفاده از ترکیبات سه‌قسمتی مانند «از بغداد گرفته تا دمشق و بیروت» نیز رایج است؛ به‌این ترتیب، با توجه به معانی جملات آلمانی فوق و فهم این نکته که در کل این مثال‌ها، گستردگی شمول مسئله‌ای به‌توسط گوینده یا نویسنده بیان می‌شود، می‌توان از عبارت «از ... و ... (گرفته) تا ...» یا «از ... گرفته تا ... و ...» به‌عنوان معادل در ترجمه‌های فارسی استفاده کرد. نکته قابل‌توجه در معادل فارسی این عبارت، تغییر کاربردی معنای «حروف اضافه»^۱ über و bis است.

۱. Präposition

مسئله ۳:

گرچه ترتیب قرارگرفتن کلمات در جمله، در زبان آلمانی آسان به نظر می‌رسد، اما ممکن است زبان‌آموز در بدیهی‌ترین موارد نیز دچار تردید شود؛ برای مثال اگر یکی از «کلمات ربط» همچون "denn" یا "aber" در اول جمله قرار بگیرد، آیا فعل در جای دوم می‌نشیند یا عضو دیگری از جمله در این مکان قرار می‌گیرد؟ پاسخ به این پرسش شاید ساده باشد، اما اگر بخواهیم از ترتیب قرارگرفتن اجزای جمله در ساختارهایی نظیر "je ... desto"، "je ... desto" و "je ... umso" سردرپیآوریم، مسئله قدری پیچیده می‌شود.

یافتن پاسخ از طریق پیکره متنی:

C1: [word="je"] []{0,5} [word="je|desto|umso"]

C2: [word="je"] [word!="und|\.\|":-"]{1,2} [word="je|desto|umso"]

C3: [word="je"] [word!="und|\.\|":-"]{3,10} [word="je|desto|umso"]

توضیح: اگر پیش فرض در طرح مسئله این باشد که زبان‌آموز هیچ‌گونه اطلاعی درباره این‌گونه ترکیب‌ها ندارد، توصیه این است که از ساختار اول (C1) استفاده شود. با این ساختار، کاربر از نرم‌افزار درخواست می‌کند، کل ترکیباتی را بیابد که در آن‌ها عنصر دوم در یک فاصله متغیر (حداقل صفر و حداکثر پنج کلمه) نسبت به عنصر اول قرار گرفته باشد. با بررسی اولین یافته‌ها، زبان‌آموز درمی‌یابد که در اکثریت مطلق نمونه‌ها یک «صفت تفصیلی»^۱ بعد از je، desto و umso قرار گرفته و نتیجه می‌گیرد که اولاً حداقل فاصله باید یک کلمه باشد، نه صفر؛ ثانیاً بهتر است کلمات غیرضروری موجود در میان دو عنصر جست‌وجو (مانند und در "je und je" که در فهرست یافته‌ها آمده است) با استفاده از قواعد زبان پرس‌وجوی پیکره حذف شوند. ساختار C2 این هدف را تحقق بخشیده و با کاربست آن، فقط ترکیباتی پیدا می‌شوند که در آن‌ها ابتدا je، سپس یک صفت تفصیلی و در انتها یکی از سه عبارت je، desto و umso آمده است. علت انتخاب عدد دو به‌عنوان حداکثر فاصله به این علت است که علائم جمله نظیر نقطه و ویرگول در زبان پرس‌وجوی پیکره کلمه (word) محسوب می‌شوند و به همین علت، بایستی در متونی که قواعد نگارش را رعایت کرده‌اند، میان je و عنصر دوم جست‌وجو، به تعداد حداقل دو کلمه فاصله ایجاد شود:

۱. Konjunktion

۲. Komparativ

LeftContext	Keyword	RightContext
sehr klein sein , es gilt das Prinzip :	je kleiner , desto	effektiver . Sollte es also der US-Armee
Grunde nicht . Wie des öfteren gilt jedoch :	je früher desto	besser . Da in unserem Verein hauptsächlich
Frau aus den Arm , er will mich loswerden ,	je schneller , desto	besser . Er redet viel und laut , berichtet
gibt Kurse ab dem 6 Lebensjahr , es gilt :	je früher , desto	besser Marina : Ist Orientalischer Tanz
. Mit den Werten ist es wie mit Kaviar :	je rarer , desto	kostbarer . Schon Mitte der 90er Jahre

ساختار سوم (C3) برای یافتن عباراتی است که در آنها، ترکیب عنصر اول جست‌وجو (je) و صفت تفصیلی در قالب یک جمله کامل آمده است. برای این منظور از فاصله «حدافل سه کلمه» استفاده شده است. با این تنظیم، زبان پرس‌وجوی پیکره مجبور به یافتن این‌گونه عبارات خواهد شد:

LeftContext	Keyword	RightContext
. Dafür ist die Mauer da , und	je schneller sie gebaut wird , umso	besser . Sie schafft getreu der
! Was ist nur passiert ? Ganz einfach :	je höher die Temperaturen steigen , umso	weniger Klamotten haben die Leute an
daraus ? 1. Fliege nie zu hoch , denn	je höher Du aufsteigst , desto	tiefer ist der Fall . 2. Nicht jeder ,
. " Käthe " , ruft er und rennt . Und	je schneller er rennt , desto	weiter entfernt sie sich von ihm . "
Und je entschlossener wir auftreten ,	je mehr wir wagen , desto	geringer ist die Gefahr , desto weniger
Klebstoffen benutzt . Benzin sparen :	Je schneller man fährt , umso	mehr Benzin verbraucht ein Auto . Es ist
lag der Gipfel in Wolken , aber	je näher wir kamen umso	mehr lösten sich die Wolken auf . Nach

نتیجه: زبان‌آموز با ایده‌گرفتن از شیوه جست‌وجوی فوق و نتایج به‌دست‌آمده، قادر خواهد بود، به راحتی همه ساختارهای نحوی به‌کاررفته در ترکیبات "je ... je"، "je ... desto"، و "je ... umso" را بیابد. در ادامه به برخی از این ساختارها اشاره می‌شود:

- **je + Komparativ + [,] + je (oder umso oder desto) + Komparativ + [.]**
Je höher, desto besser.
- **je + Komparativ + Subjekt + Prädikat + [,] + je (oder umso oder desto) + Komparativ + [.]**
Je höher du aufsteigst, desto besser.
- **je + Komparativ + Subjekt + Satzergänzungen + Prädikat + [,] + je (oder umso oder desto) + Komparativ + [.]**
Je schneller man die Arbeit erledigt, desto besser.
- **je + Komparativ + Satzergänzungen + Subjekt + Prädikat + [,] + je (oder umso oder desto) + Komparativ + [.]**
Je mehr Material du findest, desto besser.
- **je + Komparativ + Subjekt + Prädikat + [,] + je (oder umso oder desto) + Komparativ + Prädikat + Subjekt + [.]**
Je schneller er rennt, desto weiter entfernt sie sich von ihm.
- **je + Komparativ + Subjekt + Satzergänzungen + Prädikat + [,] + je (oder umso oder desto) + Komparativ + Prädikat + Subjekt + [.]**
Je mehr man die Katze streichelt, desto höher trägt sie den Schwanz."
- **je + Komparativ + Satzergänzungen + Subjekt + Prädikat + [,] + je (oder umso oder desto) + Komparativ + Prädikat + Subjekt + [.]**
Je mehr Freude wir empfinden können, desto verwundbarer sind wir.

ساختارهای جست‌وجو در زبان پرس‌وجوی پیکره از تنوع زیادی برخوردارند؛ به نحوی که یافته‌های مسائل ۱-۳ نه فقط از طریق موارد پیشنهادی، بلکه از طریق تولید ساختارهای دیگر پرس‌وجو نیز قابل دستیابی است. در خاتمه مطلب، به چند ساختار دیگر برای یافتن داده‌های مورد نیاز اشاره می‌شود:

D: [depos="ADJA*" & delemma=".*(ig|lich|isch)"]

با استفاده از ساختار فوق کلّ مواردی یافت می‌شوند که اولاً به‌عنوان صفت تعریف شده‌اند؛ ثانیاً به مجموعه حروف ig, lich و isch ختم می‌شوند.

E: [depos="PTKVZ" & word="nach"]

این ساختار تمام عباراتی را می‌یابد که در آن‌ها فعل جمله، دارای یک پیشوند جداشدنی است. برچسب PTKVZ از نرم‌افزار پیکره می‌خواهد، این‌گونه پیشوندها را که در زبان آلمانی، به‌طور معمول در آخر جمله قرار می‌گیرند، بیابد. اگر هدف از جست‌وجو، یافتن افعال با یک پیشوند مشخص باشد، عنوان دقیق پیشوند با استفاده از عنصر word تعریف می‌شود.

F: [depos="PPER" & word="[IDESW].*"] [depos="V.FIN"]

با استفاده از این ساختار می‌توان کلّ جملاتی را پیدا کرد که در آن، فاعل جمله یکی از ضمائر شخصی (برچسب PPER) و عبارت بعد از آن یکی از انواع افعال نظیر کمکی، معین یا غیره است. نرم‌افزار پیکره، حروفی که در داخل گروه قرار گرفته‌اند و علامت * را I.*, D.*, E.*, S.* و W.* تفسیر می‌کند. این حروف با توجه به قید برچسب PPER فقط می‌توانند حروف اول ضمائر باشند.

۳- نتیجه‌گیری

کلیه کسانی که به یادگیری زبان‌های خارجی اهتمام می‌ورزند، با مشکلات متعدد و متنوعی روبه‌رو می‌شوند. گستره این مشکلات، شامل حوزه مختلفی است. عدم اطلاع از مسائل فرهنگی که باعث استفاده نادرست از آموخته‌های زبانی می‌شود، عدم تسلط به جنبه‌های مختلف زبانی نظیر دستورزبان یا قواعد نوشتار و در سطوح بالاتر، عدم اشراف به ویژگی‌های زبان‌های تخصصی از جمله مواردی اند که موجب وابستگی نسبی یا مطلق زبان‌آموز به آموزگار یا استادان یا دیگر اهالی فن می‌شوند. نگارنده معتقد است، اگر نه تمامی این مشکلات، به یقین بسیاری از آن‌ها را می‌توان به‌شرط استفاده صحیح از ابزار رایانه‌ای موجود مرتفع ساخت. در همین راستا، مقاله پیش‌رو اهمیت یکی از همین ابزار رایانه‌ای، یعنی پیکره‌های متنی را یادآوری و تلاش کرد بعد از طرح مقدماتی نظیر دادن تعریف، بیان شاخصه‌ها و توضیحاتی پیرامون انواع پیکره‌ها که دربردارنده مصادیق عینی کاربردهای زبانی‌اند، شیوه کار با این‌گونه

متون را تشریح کند. چنانکه در متن مقاله نیز بیان شد، پژوهش در متون، مستلزم آشنایی با ابزارهای جست‌وجو در نرم‌افزار مدیریت پیکره است. به همین علت نیز نگارنده تلاش کرد در قسمت پایانی مقاله نحوه استفاده از زبان پرس‌وجوی پیکره را تشریح کند و با استفاده از آن، به حل و فصل برخی از مشکلات زبان‌آموزان بپردازد.

زبان پرس‌وجوی پیکره ابزار قدرتمندی است که با استفاده از آن، می‌توان نه تنها مشکلات مربوط به یادگیری معانی و کاربرد واژگان و اصلاحات، مقوله‌های دستوری و نگارش و غیره را مرتفع ساخت، بلکه پیچیده‌ترین ساختارهای نحوی در زبان را از پیکره‌های متنی استخراج کرد و مورد پژوهش و بررسی قرار داد. چنانکه در مقاله پیش‌رو نیز ملاحظه شد، آموختن این زبان و استفاده از پیکره‌های آماده تحت‌وب یا تولید پیکره به شیوه‌ای که شرح داده شد، زبان‌آموز را در موقعیتی قرار می‌دهد که در یادگیری زبان، وابستگی خود به دیگران را به حداقل برساند.

۴- منابع

- Aijmer, Karin (2009). *Corpora and language teaching*. Amsterdam: Benjamins (Studies in corpus linguistics, 33).
- Baker, Paul; Hardie, Andrew; McEnery, Tony (2006). *A glossary of corpus linguistics*. Edinburgh: Edinburgh Univ. Press (Glossaries in Linguistics EUP).
- Dash, Niladri Sekhar. *Corpus Linguistics: A General Introduction*. CIIL, Mysore, 25th August 2010. URL: <http://www.ldcil.org/download/Corpus%20Linguistics.pdf>
- Kennedy, Graeme D. (1998). *An introduction to corpus linguistics*. London: Longman (Studies in language and linguistics).
- Lemnitzer, Lothar; Zinsmeister, Heike (2010). *Korpuslinguistik. Eine Einführung*. 2., durchges. und aktualisierte Aufl. Tübingen: Narr (Narr-Studienbücher).
- Lüdeling, Anke; Kytö, Merja (2009-). *Corpus linguistics. An international handbook* / edited by Anke Lüdeling and Merja Kytö. Berlin: Walter de Gruyter (Handbücher zur Sprach- und Kommunikationswissenschaft = Handbooks of linguistics and communication science, Bd. 29.1-).
- McEnery, Tony; Wilson, Andrew (2001). *Corpus linguistics. An Introduction*. 2nd ed., Edinburgh: Edinburgh University Press (Edinburgh textbooks in empirical linguistics).
- Mukherjee, Joybrato (2002). *Korpuslinguistik und Englischunterricht. Eine Einführung*. Frankfurt am Main, New York: P. Lang (Sprache im Kontext, Bd. 14).

O'Keeffe, Anne; McCarthy, Michael (2010). *The Routledge handbook of corpus linguistics*. London: Routledge (Routledge handbooks).

Pearson, Jennifer (1998). *Terms in context*. Amsterdam: John Benjamins (Studies in corpus linguistics, 1388-0373, v.1).

Scherer, Carmen (2014). *Korpuslinguistik. 2., aktualisierte Aufl.* Heidelberg: Winter (Kurze Einführungen in die germanistische Linguistik, 2).



ضمیمه: مجموعه برجسب دانشگاه‌های اشتوتگارت و توبینگن برای توصیف انواع کلمه

Das Stuttgart-Tübingen Wortarten-Tagset^۱

Tag	Definition	Beispiel
ADJA	attributives Adjektiv	[das] große [Haus]
ADJD	adverbiales oder prädikatives Adjektiv	[er fährt] schnell [er ist] schnell
ADV	Adverb	schon, bald, doch
APPR	Präposition; Zirkumposition links	in [der Stadt], ohne [mich]
APPRART	Präposition mit Artikel	im [Haus], zur [Sache]
APPO	Postposition	[ihm] zufolge, [der Sache] wegen
APZR	Zirkumposition rechts	[von jetzt] an
ART	bestimmter oder unbestimmter Artikel	der, die, das, ein, eine, ...
CARD	Kardinalzahl	zwei [Männer], [im Jahre] 1994
FM	Fremdsprachliches Material	[Er hat das mit " A big fish [" übersetzt]
ITJ	Interjektion	mhm, ach, tja
ORD	Ordinalzahl	[der] neunte [August]
KOUI	unterordnende Konjunktion mit "zu" und Infinitiv	um [zu leben], anstatt [zu fragen]
KOUS	unterordnende Konjunktion mit Satz	weil, daß, damit, wenn, ob
KON	nebenordnende Konjunktion	und, oder, aber
KOKOM	Vergleichskonjunktion	als, wie

۱. فهرست این برجسب‌ها کامل نیست و برخی مقوله‌های دستوری نماینده‌ای در این فهرست ندارند. اما نظر به اینکه مجموعه کامل‌تری از این برجسب‌ها برای زبان آلمانی وجود ندارد، استفاده از آن برای مقاصد مورد نظر در این مقاله توصیه می‌شود.

Tag	Definition	Beispiel
NN	normales Nomen	Tisch, Herr, [das] Reisen
NE	Eigennamen	Hans, Hamburg, HSV
PDS	substituierendes Demonstrativpronomen	dieser, jener
PDAT	attribuierendes Demonstrativpronomen	jener [Mensch]
PIS	substituierendes Indefinitpronomen	keiner, viele, man, niemand
PIAT	attribuierendes Indefinitpronomen ohne Determiner	kein [Mensch], irgendein [Glas]
PIDAT	attribuierendes Indefinitpronomen mit Determiner	[ein] wenig [Wasser], [die] beiden [Brüder]
PPER	irreflexives Personalpronomen	ich, er, ihm, mich, dir
PPOSS	substituierendes Possessivpronomen	meins, deiner
PPOSAT	attribuierendes Possessivpronomen	mein [Buch], deine [Mutter]
PRELS	substituierendes Relativpronomen	[der Hund ,] der
PRELAT	attribuierendes Relativpronomen	[der Mann ,] dessen [Hund]
PRF	reflexives Personalpronomen	sich, einander, dich, mir
PWS	substituierendes Interrogativpronomen	wer, was
PWAT	attribuierendes Interrogativpronomen	welche [Farbe], wessen [Hut]
PWAV	adverbiales Interrogativ- oder Relativpronomen	warum, wo, wann, worüber, wobei
PAV	Pronominaladverb	dafür, dabei, deswegen, trotzdem
PTKZU	"zu" vor Infinitiv	zu [gehen]
PTKNEG	Negationspartikel	nicht
PTKVZ	abgetrennter Verbzusatz	[er kommt] an, [er fährt] rad
PTKANT	Antwortpartikel	ja, nein, danke, bitte
PTKA	Partikel bei Adjektiv oder Adverb	am [schönsten], zu [schnell]
SGML	SGML Markup	

Tag	Definition	Beispiel
SPELL	Buchstabierfolge	S-C-H-W-E-I-K-L
TRUNC	Kompositions-Erstglied	An- [und Abreise]
VVFIN	finites Verb, voll	[du] gehst, [wir] kommen [an]
VVIMP	Imperativ, voll	komm [!]
VVINFIN	Infinitiv, voll	gehen, ankommen
VVIZU	Infinitiv mit "zu", voll	anzukommen, loszulassen
VVPP	Partizip Perfekt, voll	gegangen, angekommen
VAFIN	finites Verb, aux	[du] bist, [wir] werden
VAIMP	Imperativ, aux	sei [ruhig !]
VAINFIN	Infinitiv, aux	werden, sein
VAPP	Partizip Perfekt, aux	gewesen
VMFIN	finites Verb, modal	dürfen
VMINFIN	Infinitiv, modal	wollen
VMPP	Partizip Perfekt, modal	gekonnt, [er hat gehen] können
XY	Nichtwort, Sonderzeichen enthaltend	3:7, H2O, D2XW3
\\$,	Komma	,
\\$.	Satzbeendende Interpunktion	. ? ! ; :
\\$(sonstige Satzzeichen; satzintern	- [,] (

پژوهشگاه علوم انسانی و مطالعات فرهنگی
رتال جامع علوم انسانی