

از پیکره زبانی تا زبان‌شناسی پیکره‌ای

مصطفی عاصی

پژوهشگاه علوم انسانی و مطالعات فرهنگی

چکیده

بنابر تعریف، پیکره زبانی^۱ عبارتست از مجموعه ای از متن‌های نوشتاری یا گفتاری آوانویسی شده که می‌توان در توصیف و تحلیل زبان از آن بهره گرفت. (Kennedy, 1998:1) گرچه در گذشته، بسیاری از زبان‌شناسان بر اهمیت پیکره زبانی در بیشتر بررسی‌هایشان تأکید کرده‌اند، اما در دوران جدید است که تکیه بر داده‌های واقعی زبانی به صورت گسترده‌ای رواج یافته و شرط اساسی بسیاری از پژوهش‌های نظری و کاربردی مانند نظریه‌پردازی و توصیف ساختمان زبان، گویش‌شناسی، دستورنویسی و فرهنگ‌نگاری به‌شمار می‌آید. امکان سازماندهی، تنظیم، تفکیک، جستجو و دستیابی سریع داده‌های زبانی، افق‌های تازه‌ای در برابر پژوهندگان گشوده و باعث پیدایش شاخه‌ای تخصصی در حوزه زبان‌شناسی رایانه‌ای گردیده است. این شاخه بانام زبان‌شناسی پیکره‌ای^۲ تنها در آخرین دهه‌های قرن بیستم ایجاد شد و در همین زمان کوتاه تبدیل به یکی از فعال‌ترین و پرکاربردترین زمینه‌ها شده است. تالیف کتاب‌های متعدد و ایجاد مجامع علمی، مراکز پژوهشی، نشریه‌های تخصصی و بین‌المللی و سایت‌های اینترنت برای این شاخه، گویای این واقعیت است. در این گفتار، پس از اشاره‌ای به تاریخچه شکل‌گیری این رشته، به رویکردهای نظری و کاربردی زبان‌شناسی پیکره‌ای، انواع پیکره‌های زبانی، نشانه‌گذاری پیکره‌ها، سازماندهی آنها، پایگاه داده‌های زبانی و نقش رایانه در آنها می‌پردازیم. سپس برخی از پیکره‌های مهم زبان انگلیسی و فارسی و نیز چند پروژه مهم فرهنگ‌نگاری را که با بهره‌گیری از پیکره‌های زبانی انجام یافته‌اند، معرفی می‌کنیم.

کلیدواژه‌ها :

پیکره - پیکره زبانی - زبان‌شناسی - زبان‌شناسی پیکره‌ای - رایانه - زبان فارسی - برچسب دهی

پیشگفتار :

در سال ۱۹۹۲ هلیدی، زبان‌شناس نامی، در همایش ویژه‌ای درباره زبان‌شناسی پیکره‌ای می‌گوید: "از نخستین روزهایی که تصمیم گرفتم دستور نویس شوم، همواره می‌اندیشیدم که دستور موضوعی است با مقدار زیادی نظریه و مقدار ناچیزی داده. و از این روی برای دو نکته اهمیت قائل بوده‌ام: اول آنکه برای بررسی دستور نیاز به حجم بزرگی از داده‌های زبانی داریم، چرا که باور دارم دستور را باید به شکلی کمی مطالعه کرد و دیگر آنکه، باید

چگونگی کاربرد روش‌های کمی را برای تعیین درجات ارتباط میان دستگاه‌های گوناگون دستوری نشان داد (کاری که در پایان نامه دکترای خود کرده ام) (Halliday, 1992: 61)

بخش بزرگی از بررسی‌های زبان‌شناسان، دیدگاهی همانند هلیدی دارند. همیشه یکی از آرزوهای زبان‌شناسان کاربردی و حتی بسیاری از نظریه پردازان این بوده است که به مقادیر بزرگی از داده‌های زبانی دسترسی داشته باشند، گرچه در گذشته فراهم ساختن چنین منابعی بسیار دشوار، پر زحمت و هزینه بوده است. بسیاری از زبان‌شناسان سرشناس گذشته، انبوه بزرگی یادداشت مربوط به کاربرد واقعی زبان را با رنج و حوصله فراوان گرد می‌آوردند تا براساس آنها قواعد و قوانین زبان را استخراج کنند و چه بسیار از این یادداشت‌ها به دلیل حجم زیاد، پراکندگی و پایان عمر دانشمندان بی‌استفاده رها شده‌اند. (Svartvik, 1992: 7)

نگاهی به پیشینه پیکره‌های زبانی :

نلسون فرانسیس در گفتاری با عنوان "Language Corpus B.C." مروری بر مجموعه‌های داده‌های زبانی دارد که در گذشته به طور دستی فراهم آمده‌اند. یادآور می‌شود که وی با طنز، دوران پیش از پیدایش رایانه را B.C.^۳ می‌نامد. از نخستین مواردی که می‌توان واژه corpus را یافت، قرن ششم میلادی است که امپراتور ژوستینیان Corpus Juris Civilis را به عنوان مجموعه‌ای از قوانین رومی به کار می‌برد. در قرن هشتم فهرستی از واژه‌های مشکل لاتینی با عنوان Corpus Glossary فراهم شد که ترتیبی الفبایی داشت و شاید نخستین فرهنگ دو زبانه لاتین انگلیسی به شمار آید. پس از آن در اروپا پیکره‌هایی به منظورهای گوناگون شکل گرفت که می‌توان آنها را در سه دسته جای داد:

۱- پیکره‌ها و داده‌های فرهنگ نگاری که مهمترین آنها برای زبان انگلیسی به وسیله جانسون در قرن هجده، مورای در قرن نوزده و میریام وبستر در قرن بیستم فراهم شده است.

۲- پیکره‌های گویشی که اغلب در اواخر قرن نوزده و نیمه اول قرن بیستم فراهم آمده است.

۳- پیکره‌ها و مطالب دستوری که مربوط به قرن بیستم است. (Francis, 1992:17)

اصطلاح پیکره را به ویژه زبان‌شناسان ساختگرا به کار می‌بردند و همواره تأکید می‌کردند که توصیف یک زبان یا گویش باید مبتنی بر داده‌های گردآوری شده و تحلیل این داده‌ها باشد و با آنکه فراگیری و بزرگی پیکره عامل تعیین کننده‌ای در افزایش دقت و اعتبار نتایج به شمار می‌آید، اما محدودیت‌های نیروی انسانی و زمان در برابر این عامل قرار داشت و در نتیجه، بیشتر به نمونه‌های برگزیده‌ای اکتفا می‌شد. "تنها از دهه ۱۹۶۰ بود که توانستند با کمک رایانه مجموعه‌های بسیار بزرگی از متن‌های معتبر زبانی فراهم نمایند و آغازگر همه اینها، پیکره زبانی براون بود که نمونه‌ای معیار از زبان انگلیسی امروز برای استفاده در رایانه‌های رقمی به شمار می‌آید." (Svartvik, 1992:7)

این پیکره را نلسون فرانسیس و هنری کوچرا در دانشگاه براون با گردآوری نزدیک به یک میلیون واژه از پانصد نمونه متنی پیوسته ایجاد کردند و کوشیدند انواع مقاله‌ها و گونه‌های زبانی را پوشش دهند. کار از ۱۹۶۱ آغاز گردید و در سال ۱۹۶۴ به پایان رسید. این مدت در مقایسه با کارهای گذشته و با توجه به حجم نسبتاً زیاد داده‌ها در آن زمان بسیار کم به نظر می‌رسید. (Francis and Kučera, 1964)

یکی از مهم ترین پیکره‌هایی که در همین دهه، اما بدون بهره‌گیری از رایانه فراهم آمد و از این نظر آن را در شمار پیکره‌های پیش از دوران الکترونیک قرار می‌دهند، پیکره بررسی کاربرد زبان انگلیسی یا SEU^۴ است که راندلف کرک^۵ از ۱۹۵۹ در دانشگاه لندن آغاز کرد و گردآوری، آوانگاری و آماده‌سازی آن تا سال ۱۹۸۹ طول کشید، چرا که برای یک میلیون واژه آن یک میلیون برگه با هفده سطر از متن‌های مربوط به سال‌های ۱۹۵۳ تا ۱۹۸۷ با دست نوشته شد. البته در کنار این پیکره نوشتاری، بخش گفتاری آن با عنوان پیکره لندن لوند، بررسی انگلیسی گفتاری (SSE) به وسیله یان سوارتویک در دانشگاه لوند سوئد و از سال ۱۹۷۵ آغاز گردید. از آنجا که این بخش با کمک رایانه فراهم گردید، از همتای نوشتاری خود پیشی گرفت و با حدود نیم میلیون واژه در سال ۱۹۸۰ ارائه گردید. (Kennedy, 1998:32)

پیکره از دیدگاهی نظری : آغاز زبان‌شناسی پیکره‌ای :

در دانش زبان، پیکره، "مجموعه‌ای از متون نوشتاری یا گفتاری آوانویسی شده است که می‌توان آن را به عنوان مبنایی برای تحلیل و توصیف زبانی به کار برد." (Kennedy, 1998:1) پیکره زبانی می‌تواند بسیار بزرگ، فراگیر و نماینده تمامی یک زبان و یا گونه‌ای از آن باشد، به شکل برگه‌های یادداشت یا پرونده‌های رایانه‌ای شامل متن‌های کامل یا گزیده‌هایی از آنها، بخش‌های پیوسته‌ای از متون یا گزیده‌ای از نقل قول‌ها و نکات و حتی فهرست‌های واژگانی باشد. پیکره می‌تواند ویژه بررسی خاصی فراهم شود و یا دربرگیرنده مجموعه عظیم و بی‌ساختاری از متون گوناگون باشد که برای منظوره‌های گوناگون به کار رود.

در آغاز دهه ۱۹۸۰ شاید می‌توانستیم پیکره‌های الکترونیکی را با انگلستان یک دست بشماریم. در آغاز دهه ۱۹۹۰ تعداد طرح‌هایی که برای ایجاد پیکره‌های زبانی فعال بودند به دهها می‌رسید و چند سال بعد در کشورهای مختلف جهان فعالیت‌های بسیاری برای زبان‌های گوناگون آغاز گردیده بود. اکنون پس از گذشت تنها دو دهه، با جستجوی ساده در شبکه جهانی رایانه‌ها (اینترنت) بیش از سیصد پیکره بزرگ و کوچک برای بسیاری از زبان‌های جهان فهرست می‌شوند. علت این گسترش شتابنده در نخستین وهله، راه یافتن رایانه در این حوزه و تغییرات بنیادی در روش‌های پردازش متن و ذخیره‌سازی و دستیابی به آن است. "بدین سان، زبان‌شناسی پیکره‌ای اکنون ناگزیر با رایانه پیوند یافته است و همین وسیله است که سرعت شگفت‌انگیز، شمارپذیری کامل، تکرار و روبرداری دقیق، صحت آماری و امکان به‌کارگیری حجم عظیم داده‌ها را به همراه آورده است." (Kennedy, 1998:5)

برای زبان‌شناسان قدیمی‌تر، اصطلاح "زبان‌شناسی پیکره‌ای" اغلب یادآور دوران رونق پیکره زبانی یعنی دهه ۱۹۵۰ است، دوران هریس^۶، فریس^۷، هیل^۸ و دیگر ساختگرایان آمریکایی. اما از آن دوران به بعد است که این اصطلاح کاربرد بیشتری یافته و به ویژه در اوایل دهه ۱۹۹۰ است که به عنوان شاخه‌ای مشخص از زبان‌شناسی، مانند زبان‌شناسی اجتماعی، زبان‌شناسی روانشناختی و زبان‌شناسی متنی که در دهه‌های ۱۹۶۰ و ۱۹۷۰ شکل گرفته‌اند، شناخته شده است. (Leech, 1992:105) اما زبان‌شناسی پیکره‌ای با رشته‌های یاد شده تفاوتی دارد، زیرا نماینده یک حوزه مشخص از مطالعات زبانی نیست بلکه بنیادی روش شناختی برای پژوهش‌های زبانی به

شمار می‌آید. در اصل و عملاً زبان‌شناسی پیکره‌ای به آسانی با شاخه‌های دیگر زبان‌شناسی می‌آمیزد. می‌توان با کمک پیکره به بررسی‌های آوایی، نحوی، اجتماعی یا دیگر زمینه‌های زبان پرداخت و در این صورت می‌گوییم که روش‌ها و فنون زبان‌شناسی پیکره‌ای را با موضوعات آوایی، نحوی و اجتماعی زبان و مانند آن آمیخته‌ایم. (Leech, 1992:106)

تنها رشته دیگر زبان‌شناسی که مانند این رشته، با ابزار و روش‌های مطالعه و نه با موضوعی خاص سروکار دارد، زبان‌شناسی رایانه‌ای است که به عنوان "مطالعه زبان با کمک رایانه" تعریف شده است. و امروزه به نظر می‌رسد که این دو رشته با یکدیگر پیوند یافته‌اند. یعنی می‌توان این حوزه را "زبان‌شناسی پیکره‌ای رایانه‌ای"^۹ یا CCL نامید که در این صورت نه تنها روش‌های نوین بررسی زبان بلکه فعالیت پژوهشی تازه‌ای با رویکردی فلسفی در زبان‌شناسی به شمار می‌آید. لیچ، ویژگی‌های مهم این رشته را چنین برمی‌شمارد:

۱- تمرکز بر کنش زبانی و نه توانش زبانی

۲- تمرکز بر توصیف زبانی و نه بر همگانی‌های زبان

۳- تمرکز بر الگوهای کمی زبانی همانند الگوهای کیفی آن

۴- تمرکز بر دیدگاه‌های تجربی (و نه عقلانی) در بررسی‌های علمی زبان

همانگونه که مشاهده می‌شود، این ویژگی‌ها مجموعه‌ای را به وجود می‌آورد که توجه بیشتری به جنبه‌های رفتاری زبان و بروز طبیعی گفتار و نوشتار دارد و عملاً در مقابل دیدگاه‌های چامسکی و پیروان وی قرار می‌گیرد. (P:107)

تویبرت نیز نگرشی را همسو با لیچ نشان می‌دهد: "زبان‌شناسی پیکره‌ای بر پایه این باور که زبان اساساً پدیده‌ای اجتماعی است بنا نهاده شده است. پدیده‌ای که پیش از هر چیز، می‌توان آن را با داده‌های تجربی آماده، یعنی در کنش‌های ارتباطی، مشاهده و توصیف کرد. متن‌های مورد مشاهده، در اصل کنش‌های ارتباطی گذرا هستند." (Teubert, 1999:1)

از سوی دیگر، وی بررسی این پدیده اجتماعی را مستلزم دانستن چگونگی درک گوینده یا شنونده از مطالب نمی‌داند، زیرا زبان، به عنوان یک پدیده اجتماعی، به صورت متنی متجلی می‌گردد که می‌توان آن را مشاهده، ضبط، توصیف و تحلیل کرد. از آنجا که حجم مواد زبانی (در اغلب زبان‌های امروز جهان) برای گنجاندن در پیکره بسیار زیاد است، در برخی موارد می‌توان پیکره را به صورت نمونه ایجاد کرد. تعیین حدود و گستره پیکره، وظیفه زبان‌شناسان است. برخی از متغیرهایی که در تعیین حدود دخالت دارند عبارتند از زبان، برش یا محدوده‌ای از زمان، منطقه، شرایط و موقعیت، ویژگی‌های بیرونی و درونی متنی و بسیاری دیگر. زبان‌شناسی پیکره‌ای به توصیف تک تک زبان‌های طبیعی می‌پردازد و نه همگانی‌های زبان. از آنجا که نمی‌توان به درون ذهن افراد رخنه کرد، تنها می‌توان قراردادهای زبانی را در کنش‌های ارتباطی و متون یافت. گرچه فرهنگ‌های لغت، کتاب‌های دستور و کتاب‌های درسی زبان نیز جزئی از فضای کلامی هستند، اما نمونه‌های واقعی از فضای کلامی و متن‌ها بهتر می‌توانند واقعیات زبان را نشان دهند. زبان‌شناسی پیکره‌ای با آمیختن سه روش، به فراهم آوردن دانش تجربی زبانی کمک می‌کند:

الف : استخراج خودکار داده‌های زبانی از پیکره‌ها
ب : پردازش برونداد با روش‌های عمدتاً آماری
پ : ارزیابی و تفسیر اینگونه داده‌های پردازش شده
مراحل اول و دوم را می‌توان و باید به طور کامل با برنامه و خودکار انجام داد، اما مرحله سوم نیاز به تصمیم‌گیری و منطق انسانی دارد. (Teubert, 1999:1)

انواع پیکره‌های زبانی :

همانگونه که پیشتر اشاره شد، پیکره‌های زبانی می‌توانند عام و گسترده، نمونه تصادفی از یک زبان و یا تخصصی و نماینده یک گونه خاص زبان باشند. آنها را می‌توان برای منظوره‌های گوناگون به کار گرفت، از جمله برای فرهنگ‌نگاری، معنانشناسی، بررسی‌های دستوری، آموزش زبان و مانند اینها. پیکره‌ها را می‌توان از نظر اندازه و گستره به دسته‌های محدود، متوسط و عظیم تقسیم کرد. براساس یکی از رویکردها (فرم و کارکرد)، می‌توان انواع زیر را برای پیکره‌ها در نظر گرفت :

- پیکره اشباع شده^{۱۰} که با اشاره به عوامل و متغیرهای معین با اطلاعات اضافی پر شده است. اشباع، ویژگی آماری است که با نسبت غنای واژگانی (نسبت شمار سرواژه‌ها به شمار موارد کاربردی type/token) ارتباط دارد. در این مورد، کل متن به پاره‌های مشابه بخش شده است.

- پیکره موقعیتی^{۱۱} که بنابر نیازهای خاص یک بررسی، فراهم می‌گردد.

- پیکره پیشگر^{۱۲} که دارای ابعاد تاریخی است و برای ثبت تغییرات زبانی به کار می‌رود و از برهه‌های گوناگون زمانی نمونه‌های مشابه را دارد.

- پیکره تطبیقی^{۱۳}، پیکره‌ای چندزبانه با نمونه‌های مشابه از هر زبان است.

- پیکره موازی^{۱۴}، پیکره‌ای چندزبانه با متنی اصلی از یک زبان و ترجمه آن در زبان‌های دیگر (یا زبان‌های هدف) است.

- پیکره دو سویه^{۱۵}، پیکره‌ای چندزبانه با متن اصلی به همه زبان‌ها و ترجمه آنها به همه زبان‌ها است. (Teubert, 1999:3).

کندی (۱۹۹۸) پس از دسته بندی پیکره‌ها به پیش از دوران الکترونیک و پیکره‌های الکترونیکی، دسته دوم را نیز به نسل اول و نسل دوم تقسیم می‌کند.

پیکره‌های الکترونیکی نسل اول : این پیکره‌ها با حجم محدود و کاربردهای ویژه و اغلب برای زبان انگلیسی مشخص می‌شوند و مهم‌ترین آنها عبارتند از :

-The Brown Corpus

- The Lancaster-Oslo-Bergen Corpus (LOB)

- The London-Lund Corpus (LLC)

و پیکره‌های تخصصی برای زبان انگلیسی از جمله برای فرهنگ نگاری مانند American Heritage Intermediate Corpus (AHIC) که در آمریکا فراهم شد و هدف آن تدوین فرهنگ مدرسه ای هریتج بود. یا پیکره‌ای که آلفیو^{۱۶} برای بررسی کاربردهای بریتانیایی زبان انگلیسی فراهم آورد و هردوی این پروژه ها مربوط به دهه ۱۹۶۰ می‌شوند. (ص ص ۳۴ و ۳۵). همچنین در جهت دیگر، طرح‌هایی برای بهره گیری از فرهنگ‌های لغت به عنوان پیکره زبانی اجرا گردید. از جمله فرهنگ انگلیسی امروز لانگمن و درونداد فرهنگ بزرگ آکسفورد به رایانه و پردازش‌های بعدی آن بود. پیکره‌های تخصصی دیگر نیز برای بررسی زبان گفتاری انگلیسی و بررسی‌های تاریخی و نیز یادگیری آن طراحی و ایجاد شد. (Kennedy, 1998: 35-40)

پیکره‌های الکترونیکی نسل دوم: نام دیگر این پیکره ها، غول آسا^{۱۷} است و اغلب در دهه ۸۰ و ۹۰ و با بهره گیری از امکانات بسیار پیشرفته رایانه‌های تازه ایجاد شده اند. برخی از آنها را می‌توان تنها نام برد و مشخصات کامل آنها را از منابع یاد شده در این نوشتار به دست آورد:

The COBUILD Project

Bank of English

The Longman Corpus Network

The British National Corpus (BNC)

The International Corpus of English (ICE)

یادآور می‌شود که اکنون بسیاری از پیکره‌های یاد شده، به صورت پیوسته (on-line) و از راه ارتباط با شبکه جهانی رایانه‌ای (اینترنت) قابل دسترسی و استفاده هستند.

نشانه گذاری^{۱۸} پیکره ها :

برای گویاتر شدن پیکره و کاربردهای خاص، کدهای متفاوتی به آن افزوده می‌شود. این نشانه گذاری از یک سو می‌تواند برای ارتباط دادن بخش‌های یک پیکره به ساختار کلی آن باشد، مانند شماره سطر، صفحه، فصل و مانند اینها و یا بافت زبانی را مشخص نماید، مانند شرایط تولید زبانی، گونه زبانی، رسانه و مانند آن. از سوی دیگر نشانه‌گذاری می‌تواند صرفاً زبانی باشد. نشانه گذاری زبانی ممکن است در سه سطح انجام پذیرد :

۱- در سطح صورت‌های کاربردی (tokens) : در این مورد، برای هر صورت کاربردی یک نشانه (کد) در متن افزوده می‌شود.

۲- در سطح پاره‌های زبانی (segments) : که نشانه مربوط به توالی چند صورت کاربردی در متن است.

۳- در سطح تداعی (associative level) : که نشانه‌هایی ارتباط و تداعی میان صورت‌های زبانی و پاره‌های زبانی را در متن مشخص می‌سازد.

به پیکره‌هایی که در سطح صورت‌های کاربردی نشانه گذاری می‌شوند، برچسب دار یا برچسب دهی شده می‌گویند و این نشانه ها می‌توانند نوع دستوری واژه یا برخی ویژگی‌های معنایی آن را نشان دهند. نشانه گذاری در سطح پاره زبانی را برچسب دهی نحوی می‌گویند و می‌توان ارتباط نحوی یا با همایی یا لانه گیری واحدها را با آنها نمایش داد. نشانه گذاری تداعی ها، رواج کمتری دارد.

نکته شایسته یادآوری اینکه یک پیکره ممکن است در یک یا چند سطح نشانه گذاری شود. که نمونه آن پیکره LOB است. (Mills, 1996: 247) یکی از معدود کارهایی که در زبان فارسی، برای برچسب دهی پیکره‌های فارسی انجام شده است، طراحی و اجرای برنامه‌ای رایانه‌ای برای برچسب دهی دستوری خودکار متون فارسی است. (Assi and H. Abdolhosseini, 2000)

نگاهی به برخی از طرح‌های پیکره‌ای :

پیشتر اشاره شد که زبان‌شناسی پیکره‌ای ابزارهای بنیادی برای بررسی‌ها، پژوهش‌ها و کاربردهای زبانی متفاوت فراهم می‌آورد. اکنون تنها به برخی از این موارد اشاره می‌کنیم :

- یکی از مهم‌ترین کاربردهای پیکره در پردازش زبان طبیعی است. مهم‌ترین دستاورد این حوزه، درک و بازشناسی گفتار بوده است که تنها با بهره‌گیری از پیکره‌های بزرگ امکان پذیر گشته است.

- اکنون هیچ پروژه فرهنگی نگاری پیشرفته‌ای نمی‌توان یافت که از پیکره‌های زبانی و پایگاه‌های داده‌های زبانی بهره‌گیری نکند. هارتمن که پیشتر نظریه متن‌شناسی مقابله‌ای^{۱۹} را در ۱۹۸۰ ارائه کرده است، در دهه ۱۹۹۰ ادامه و تکامل آن را به سوی پیکره‌های متنی موازی می‌بیند و با بررسی چندین طرح فرهنگی نگاری، نتیجه می‌گیرد که این پیکره‌ها ابزاری کارا و ناگزیر در کار تدوین فرهنگ‌ها به شمار می‌روند. (Hartmann, 1994: 11) نمونه چنین کاربردی در زبان فارسی **واژگان گزیده زبان‌شناسی** است که نرم افزار رایانه‌ای آن نیز با امکانات گسترده‌تر ایجاد شده است. (Assi, 1994)، (عاصی، ۱۳۷۳) و (عاصی و عبدعلی، ۱۳۷۵).

- ایجاد پایگاه‌های داده‌های زبانی نیز جنبه‌ای دیگر از کاربرد پیکره‌های زبانی است که نمونه‌های متعدد آن را هم اکنون در سراسر جهان به صورت پیوسته یا ناپیوسته در اختیار داریم. چنین پایگاهی را برای زبان فارسی نیز نگارنده در پژوهشگاه علوم انسانی ایجاد نموده است. (عاصی، ۱۳۷۶)

- طرح‌های بررسی واژه‌های همایند^{۲۰} در زبان‌های گوناگون با کمک پیکره‌های زبانی اجرا شده است. نمونه مهم و موفق آن **فرهنگ واژه‌های همایند BBI** برای زبان انگلیسی است. هم اکنون در پژوهشگاه علوم انسانی نیز طرحی برای تدوین فرهنگ واژه‌های همایند فارسی براساس پایگاه داده‌های زبان فارسی در دست اجراست.

- برنامه‌های پایشگری زبان برای پی‌گیری و ردگیری تحولات زبانی نیز از امکانات پیکره‌های زبانی سود می‌برند. این گونه پیکره‌ها را پیکره پویا یا پیکره پایشگر می‌نامند. (Kennedy, 1998 : 22)

- همه طرح‌های ترجمه ماشینی به گونه‌ای از پیکره‌های زبانی سود می‌برند بویژه رویکرد بسیار نوین پیکره بنیاد آن، و هنوز بسیاری از زمینه‌های دیگر ناگفته مانده است.

با توجه به کارایی‌های فزاینده پیکره‌های زبانی و گسترش رشته زبان‌شناسی پیکره‌ای، می‌توان گفت که تنها در سه دهه گذشته این رشته رشد اعجاب انگیزی داشته است و به روش‌های نوین، اندیشه‌ها و کارکردهای تازه با کمک دستگاه‌های الکترونیک و رایانه دست یافته است. "با همه اینها، هنوز بزرگترین و مهم‌ترین دستگامی که می‌تواند از میان انبوه واقعیت‌ها قوانین را استخراج و تدوین کند، همان ذهن انسان است." (Svartvik, 1992: 12)

یادداشت ها و پی نوشت ها :

- linguistic corpus. ۱
- corpus linguistics. ۲
- before computer. ۳
- Survey of English Usage. ۴
- Randolph Quirk. ۵
- Harris. ۶
- Feries. ۷
- Hill. ۸
- computer corpus linguistics. ۹
- saturated. ۱۰
- opportunistic. ۱۱
- monitoring. ۱۲
- comparable. ۱۳
- parallel. ۱۴
- reciprocal. ۱۵
- J., Algeo. ۱۶
- mega corpora. ۱۷
- corpus annotation. ۱۸
- contrastive textology. ۱۹
- collocations. ۲۰

مآخذ و منابع :

- عاصی، مصطفی (۱۳۷۳) "طرحی برای تهیه فرهنگ‌های تخصصی با کمک کامپیوتر"، در: *مجموعه مقالات دومین کنفرانس زبان‌شناسی نظری و کاربردی*، تهران، دانشگاه علامه طباطبایی.
- عاصی، مصطفی (۱۳۷۶) "پایگاه داده‌های زبان فارسی"، در: *مجموعه مقالات سومین کنفرانس زبان‌شناسی*، تهران، دانشگاه علامه طباطبایی و پژوهشگاه علوم انسانی و مطالعات فرهنگی.
- عاصی، مصطفی و محمد عبدعلی (۱۳۷۵) *واژگان گزیده زبان‌شناسی*، تهران، شرکت انتشارات علمی و فرهنگی.

Assi, S. M. (1994) "A Computerised English-Persian dictionary of linguistic terms: A project report" Poster paper presented at the *Sixth EURALEX Congress* at Amsterdam.

Assi, S.M. and M.H.Abdolhosseini (2000) "Grammatical tagging of a Persian corpus" in: *International Journal of Corpus Linguistics*, Vol.5, No. 1.

Francis, W N, (1992) "Language corpora B.C." in Svartvik (1992): 17-32.

Francis, W N and H. Kučera (1964) *Manual of Information to Accompany "A Standard Sample of Present-Day Edited American English, for Use with Digital Computers"* (revised 1979), Providence, RI: Department of Linguistics, Brown University.

Halliday, M.A.K. (1992) "Language as system and language as instance: The corpus as a theoretical construct", in: Svartvik (1992): 61-77.

Hartmann, R.R.K. (1997) "From contrastive textology to parallel text corpora: Theory and applications" in: *Language History and Linguistic Modelling*, edited by R. Hickey and S. Puppel, Berlin: Mouton de Gruyter.

Kennedy, Graeme (1998) *An Introduction to Corpus Linguistics*, London: Longman.

Leech, Geoffrey (1992) "Corpora and theories of linguistic performance" in: Svartvik (1992): 105-122.

Mills, Jon (1996) "Computers in applied linguistics" in: *Solving Language Problems*, edited by R. R. K. Hartmann, Exeter: University of Exeter.

Svartvik, Jan (ed.) (1992) *Directions in Corpus Linguistics*, Berlin: Mouton de Gruyter.

Svartvik, Jan (1992) "Corpus linguistics comes of age", in: Svartvik (1992): 1-12.

Teubert, Wolfgang (1999) "Corpus linguistics: A partisan view" in: *International Journal of Corpus Linguistics*, Vol.4, No.1.