



واژه‌نامه زبان‌شناسی پیکره‌ای

Corpus Linguistics

پاول بیکر، اندرو هاردی، تونی مک‌کانری

مترجمان: سعیده قندی / محمدحسن ترابی

مقدمه و نظارت علمی: دکتر سیدمصطفی عاصی



واژه‌نامه زبان‌شناسی پیکره‌ای

پاول بیکر، اندرو هاردی، و تونی مک‌اتری

مترجمان:

سعیده فتدی

محمدحسین ترابی

با مقدمه و نظارت علمی دکتر سیدمصطفی عاصی



سازمان اسناد و کتابخانه ملی جمهوری اسلامی ایران

سرشناسه	پاول بیکر، اندرو هاردی، و تونی مک‌انری
عنوان و نام پدیدآور	واژه‌نامه زبان‌شناسی پیکره‌ای / نویسندگان: پاول بیکر، اندرو هاردی، و تونی مک‌انری؛ مترجمان: سعیده قندی، محمدحسن ترابی؛ ویراستار: پرینا بخشنده
مشخصات نشر	تهران: نشر نویسه پارسی، ۱۳۹۹
مشخصات ظاهری	۳۲۷ صفحه
شابک	۹۷۸-۶۲۲-۶۶۴۹-۴۳-۸
وضعیت فهرست نویسی	فیا
یادداشت	عنوان اصلی کتاب: A Glossary of Corpus Linguistics
موضوع	زبان‌شناسی پیکره‌ای -- واژه‌نامه‌ها -- فارسی
موضوع	فارسی -- واژه‌نامه‌ها - انگلیسی
موضوع	زبان انگلیسی -- واژه‌نامه‌ها - فارسی
شناسه افزوده	قندی، سعیده، ۱۳۶۶ -، مترجم
شناسه افزوده	ترابی، محمدحسن، ۱۳۶۶ -، مترجم
شناسه افزوده	بخشنده، پرینا، ۱۳۶۵ -، ویراستار
رده بندی کنگره	QP۲۹۶
رده بندی دیویی	۸۲۳۶/۶۱۲
شماره کتابشناختی ملی	۶۱۱۹۱۴۲



این کتاب ترجمه‌ای است از:

A Glossary of Corpus Linguistics

Paul Baker, Andrew Hardie, and Tony McEnery

© Paul Baker, Andrew Hardie and Tony McEnery, 2006

نویسندگان: پاول بیکر، اندرو هاردی، و تونی مک‌انری

مترجمان: سعیده قندی، محمدحسن ترابی

ویراستار: پریرسا بخشنده

آتلیه نشر نویسه پارسی: STUDIO FIVE

ناشر: نشر نویسه پارسی

دفتر انتشارات: ۰۲۱-۷۷۰۵۳۲۴۶

نماینده فروش: کتابفروشی توس: ۶۶۴۶۱۰۰۷

سامانه پیام کوتاه: ۳۰۰۰۴۵۵۴۵۵۴۱۴۲

وبگاه: www.neveeseh.com

نوبت چاپ: اول، ۱۳۹۹

شمارگان: ۳۰۰ نسخه

شابک: ۹۷۸-۶۲۲-۶۶۴۹-۴۳-۸

چاپ و صحافی: روز

کلیه حقوق محفوظ و منعلق به «نشر نویسه پارسی» است.
تکثیر و انتشار این اثر با قسمتی از آن به هر شیوه، بدون مجوز قبلی و کتبی
ممنوع و مورد پیگیری قانونی قرار خواهد گرفت.

فرهنگ‌نامه‌ها	شماره مسلسل انتشارات
۷	۱۰۸

فهرست مطالب

۷	مقدمه
۹	پیشگفتار مترجمان
۱۳	یادداشت‌های مقدماتی
۱۳	آدرس وبگاه‌ها
۱۳	فهرست اختصارات
۲۳	اصطلاحات
۲۵۷	منابع
۲۹۳	واژه‌نامه فارسی-انگلیسی

زبان پدیده‌ای پویا و همواره در حال دگرگونی است و بیشترین تغییرات در واژه‌ها و معنی آن‌ها نمود می‌یابد. عوامل گوناگونی در این دگرگونی‌ها دخالت دارند اما شاید مهم‌ترین آن‌ها را بتوان عامل زمان دانست. هرچه از گذشته به دوران کنونی نزدیک می‌شویم، به دلایل مختلف از جمله افزایش ارتباطات و کاربرد رسانه‌های گوناگون و توسعه فراگیر دانش و فن، سرعت دگرگونی‌ها بیشتر می‌شود و با پیدایش و گسترش برق‌آسای فناوری اطلاعات این سرعت به اوج می‌رسد. در حوزه علم، بیان و توصیف دقیق مفاهیم از مهم‌ترین بایستگی‌هاست که آن‌ها هم به کاربرد مجموعه‌ای از واژه‌ها و اصطلاحاتی^۱ وابسته است که با دقت از سوی دانشوران و کارشناسان تعیین و تعریف می‌شوند. با نگاهی به واژگان تخصصی زبان‌شناسی درمی‌یابیم که بسیاری از واژه‌های این حوزه با معانی و تعریف‌های تازه‌ای در متن‌ها کاربرد یافته‌اند. به‌ویژه در شاخه تازه‌تأسیس و روبه‌رشدی مانند زبان‌شناسی پیکره‌ای که با مفاهیم و فرایندهای نوینی سروکار دارد.

تنها کافی است واژه پیکر (یا در مواردی خاص، پیکره) را در زبان فارسی در نظر بگیریم که تا حدود پنج-شش دهه پیش (زمان انتشار فرهنگ معین) به معنی تصویر، نقش و یا مجسمه و تدیس کاربرد داشت و با آغاز به‌کار رشته دانشگاهی زبان‌شناسی و به‌میان آمدن روش‌های تازه بررسی زبان با بهره‌گیری از مواد گردآوری شده، اصطلاح پیکره^۲ یا پیکره زبانی^۳ برای اشاره به هرگونه داده زبانی با هر حجم و شیوه گردآوری (یادداشت‌برداری، برگه‌نویسی دستی یا ضبط آوایی) بازتعریف شد. با این تعریف، پیکره می‌توانست دربرگیرنده چند جمله، فهرستی از چند واژه یا عبارت، شماری برگه یادداشت‌شده از منابعی پراکنده یا صورت نوشتاری چند ساعت مطلب ضبط‌شده باشد. این پیکره‌ها اغلب کاربردی موردی، شخصی، نادقیق و محدود داشتند و پس از بررسی به کناری نهاده می‌شدند و به اصطلاح یکبارمصرف بودند.

^۱ terminology

^۲ corpus

^۳ linguistic corpus

همین محدودیت‌ها و نارسایی‌ها احساس نیازی فزاینده به پیکره‌هایی بزرگتر، دقیق‌تر و با کارایی بیشتر را در میان پژوهندگان زبان ایجاد کرده بود.

نقطه عطف هنگامی فرارسید که با به‌کارگیری امکانات رایانه‌ای و فناوری اطلاعات، درونداد، ذخیره‌سازی، پردازش، جستجو و بازایی داده‌های زبانی در حجم‌های عظیم، با سرعتی شگفت‌انگیز و دقتی فراتر از انتظار ممکن گشت. امروز واژه پیکره با تعریفی تازه، بار معنایی بسیار متفاوت و گسترده‌ای یافته است و به‌عنوان اصطلاحی فنی و تخصصی در زبان‌شناسی شناخته می‌شود. شاید برای اولین بار با این معنی در نخستین کنفرانس زبان‌شناسی (عاصی، ۱۳۶۹) معرفی شد. از همین جا نام حوزه‌ای جدید یعنی میان‌رشته زبان‌شناسی پیکره‌ای^۱ نیز ساخته شد (عاصی، ۱۳۷۲).

اکنون زبان‌شناسی پیکره‌ای به‌اندازه‌ای گسترش یافته که مجموعه واژه‌ها و اصطلاحات فنی آن واژگانی تخصصی را پدید آورده است و بخشی بنیادی از این شاخه علمی، سودمند و پرکاربرد به‌شمار می‌آید. فرهنگ پیش رو نخستین گام در معرفی واژگان زبان‌شناسی پیکره‌ای و ارائه برابره‌های فارسی آن است. مترجمان برابره‌های مناسبی برای بسیاری از اصطلاح‌ها انتخاب کرده‌اند، با این حال هنوز در عرصه واژه‌گزینی این حوزه تلاش بیشتری باید کرد.

مصطفی عاصی

بهمن ماه ۱۳۹۸

عاصی، مصطفی. ([۱۳۶۹]، ۱۳۷۲). کاربرد کامپیوتر در زبان‌شناسی و فرهنگ‌نگاری. در دبیرمقدم، محمد (گردآورنده)، مجموعه مقالات نخستین کنفرانس زبان‌شناسی نظری و کاربردی (صص. ۱۶۳-۱۷۸). تهران، دانشگاه علامه طباطبائی.

^۱ corpus linguistics

کتابی که در دست دارید، فرهنگ توصیفی زبان‌شناسی پیکره‌ای، ترجمه‌ای است از کتابی نسبتاً قدیمی (نسخه انگلیسی کتاب در سال ۲۰۰۶ منتشر شده است)، اما فعلاً، البته تا جایی که ما اطلاع داریم، تنها فرهنگ توصیفی زبان‌شناسی پیکره‌ای در دسترس است و ما کتاب جدیدتری نیافتیم. ما در اولین تلاش برای انتشار چنین کتابی دست به ترجمه آن زدیم تا راهی باز شود برای تألیف کتابی جدیدتر با تأکید بر پیکره‌ها و ابزارهای در دسترس برای زبان فارسی. اما به هر حال کتاب حاضر هنوز منبع ارزشمندی است که می‌تواند به نیازهای بسیاری از دانشجویان و فعالان حوزه زبان‌شناسی پیکره‌ای و حوزه‌های وابسته به آن پاسخ دهد. کتاب برای خود ما بسیار آموزنده بود و مطمئنیم برای خوانندگان هم آموزنده خواهد بود.

در متن کتاب پیوندهایی به وبگاه‌های مختلف بود، که همان طور که در یادداشت‌های مقدماتی خود کتاب آمده است، ممکن بود آن صفحه‌های اینترنتی دیگر وجود نداشته باشند. ما پیوندها را بار دیگر کنترل کرده و جاهایی که آدرس پیوندها تغییر کرده‌اند، پیوندهای جدید (یعنی تا سال ۱۳۹۸/۲۰۲۰) را جایگزین کردیم. چند مورد از وبگاه‌ها و یا ابزارها کاملاً از دسترس خارج شده بودند، در این صورت پیوند را تغییر ندادیم، فقط در پانویس توضیح دادیم که این صفحه دیگر در دسترس نیست. البته برای اطمینان از اینکه صفحه‌ای را از دست نداده باشیم، آن موضوع خاص را در سه جویشر www.google.com، www.bing.com، و www.ecosia.org جستجو کردیم و اکنون تقریباً مطمئن هستیم که آن چند وبگاه یا ابزار دیگر وجود خارجی ندارند. اما ما نیز، مانند نویسنده‌های اصلی کتاب، پیشنهاد می‌کنیم که هر وقت به یکی از پیوندها مراجعه کردید و آن صفحه را نیافتید، دوباره آن را جستجو کنید، شاید آن صفحه مورد نظر را در آدرسی جدید یافتید. دو نکته دیگر را هم باید اضافه کنیم، یکی اینکه در متن برای ارجاع‌های متقابل بین مدخل‌ها از حروف پررنگ استفاده کرده‌ایم و دوم اینکه برای سهولت دسترسی به اصطلاحات، در پایان کتاب یک واژه‌نامه فارسی-انگلیسی هم قرار داده‌ایم.

مطالعه و استفاده از کتاب واژه‌نامه زبان‌شناسی پیکره‌ای را به همه علاقه‌مندان توصیه می‌کنیم، به‌خصوص به فعالان حوزه زبان‌شناسی پیکره‌ای، زبان‌شناسی رایانشی، فرهنگ‌نگاری، واژه‌سازی، آموزش زبان فارسی به فارسی‌زبانان، متخصصین آموزش زبان‌های خارجی، مطالعات ترجمه، و البته کسانی که در حوزه مطالعات زبان اول، روان‌شناسی زبان و زبان کودک پژوهش می‌کنند.

برای انتخاب برابرنهادهای اصطلاحات، از چند کتاب استفاده کردیم که هم به رسم ادب و هم برای آشنایی خوانندگان، فهرست آن‌ها را در اینجا می‌آوریم:

خسروی‌زاده، پروانه، شیخ‌زاده، مارال، مرادی، مهدی، و مواجی، وحید. (۱۳۹۱). فرهنگ توصیفی زبان‌شناسی رایانشی. تهران: پردیس دانش.

قطره، فریبا. (۱۳۹۶). واژه‌نامه توصیفی فرهنگ‌نویسی. تهران: نویسه پارسی.

گروه مؤلفان. (۱۳۹۷). هزار واژه زبان‌شناسی ۱. تهران: فرهنگستان زبان و ادب فارسی.

مندیک، پیت. (۲۰۱۰). کلیدواژه‌های فلسفه ذهن. ترجمه محمدحسن ترابی، ۱۳۹۵. تهران: نویسه پارسی.

میرزائی، آزاده. (۱۳۹۶). آشنایی با زبان‌شناسی پیکره‌ای. تهران: دانشگاه علامه طباطبائی.

همایون، همداخت. (۱۳۷۹). واژه‌نامه زبان‌شناسی و علوم وابسته. تهران: پژوهشگاه علوم انسانی و مطالعات فرهنگی.

در مواردی هم که در سایر منابع برابرنهادی نیافتیم، از استاد گرامی‌مان، دکتر سیدمصطفی عاصی - که نام ایشان برای همه فعالان حوزه زبان‌شناسی پیکره‌ای آشناست - کمک گرفتیم و یا اگر اصطلاحی را خودمان وضع کردیم، نظر ایشان را نیز جویا شدیم. از ایشان به دلیل کمک‌هایشان برای انتخاب برابرنهادها، مقدمه ارزشمندی که بر کتاب نوشتند، و نیز به این دلیل که وقت ارزشمند خود را برای نظارت علمی بر ترجمه این کتاب در اختیار ما قرار دادند صمیمانه سپاسگزاریم. متشکریم از آقای مصطفی کبیری، که اگر کمک‌های ایشان نبود، کتاب حاضر ترجمه نمی‌شد. همچنین قدردانیم از خانم پریسا بخشنده، ویراستار کتاب، که با ریزی و

دقت نظر مثال زدنی خود، ترجمه فارسی این کتاب را خواندنی تر و خوانش پذیرتر کردند. و در نهایت، بی نهایت سپاسگزاریم از آقای امیر احمدی، مدیر نشر نویسه پارسی که با صبر و تلاش ستودنی ایشان، کتاب منتشر شد و در اختیار خوانندگان قرار گرفت.

امیدواریم که ترجمه خوب و قابل قبولی در اختیار علاقه مندان قرار داده باشیم؛ می دانیم که این ترجمه بدون نقص نیست و حتماً کاستی هایی دارد و به عنوان مترجمان اثر، مسئولیت همه این کاستی ها را می پذیریم و قدردان نظرات و انتقادات همه مخاطبان گرامی مان خواهیم بود.

مترجمان

اردیبهشت ۱۳۹۹

تلاش کرده‌ایم که تا جای ممکن از ارجاع دادن به وبگاه‌ها خودداری کنیم، زیرا دریافتیم که برخی از وبگاه‌هایی که در ابتدای نوشتن کتاب در متن کتاب وارد کردیم، در زمان رسیدن به مراحل پایانی دیگر وجود نداشتند. ما وبگاه‌های برخی از سازمان‌ها، گروه‌ها، پیکره‌ها یا نرم‌افزارها را در صورتی در متن می‌آوردیم که احتمال می‌دادیم بسته نمی‌شوند. اما نمی‌توانیم ماندگاری همه وبگاه‌هایی که در اینجا آورده‌ایم را تضمین کنیم. اگر خوانندگان خواستند که اطلاعات خاصی را در اینترنت دنبال کنند، اما با پیوندی بسته مواجه شدند، عذرخواهی ما را بپذیرند و سپس از یک جویشگر معتبر مانند www.google.com کمک بگیرند (اگر گوگل هنوز وجود داشته باشد!).

فهرست اختصارات

زبان‌شناسی پیکره‌ای رشته‌ای است که اختصارات زیادی دارد. این از نظر وحدت رویه مشکلاتی را به وجود می‌آورد: برخی اصطلاحات با صورت اختصاری‌شان و برخی با نام کاملشان بهتر شناخته می‌شوند. ما می‌خواهیم که در ترتیب‌بندی مدخل‌های فرهنگ وحدت رویه داشته باشیم، اما می‌خواهیم یافتن مدخل‌ها آسان هم باشد؛ بنابراین، تصمیم گرفتیم در ترتیب‌بندی مدخل‌های فرهنگ، صورت کامل همه اختصارات را بنویسیم، اما فهرستی از همه اختصارات را در ابتدای فرهنگ همراه با نام کاملشان بیاوریم. بنابراین، خوانندگانی که می‌خواهند با استفاده از این فرهنگ اطلاعاتی درباره BNC به دست آورند، می‌توانند عنوان کامل آن را در فهرست اختصارات در ابتدای کتاب بیابند، و بعد به مدخل British National Corpus در فرهنگ مراجعه کنند.

برابرنهاد فارسی	مدخل انگلیسی	صورت اختصاری
سیستم برچسب‌دهی معنای واژه برای تحلیل محتوایی خودکار متون گفتاری	Automatic Content Analysis of Spoken Discourse word sense tagging system	ACASD
پیکره انگلیسی استرالیا	Australian Corpus of English	ACE
انجمن رایانه و علوم انسانی	Association for Computers and the Humanities	ACH
انجمن زبان‌شناسی رایانشی	Association for Computational Linguistics	ACL
کارگروه گردآوری داده انجمن زبان‌شناسی رایانشی	Association for Computational Linguistics Data Collection Initiative	ACLDCI
انوتیشن گراف تولکیت	Annotation Graph Toolkit	AGTK
پیکره میانی امریکن هریتیج	American Heritage Intermediate Corpus	AHI
انجمن رایانش ادبی و زبانی	Association for Literary and Linguistic Computing	ALLC
برچسب‌زن نگاشت خودکار در انگاره‌های نشانه‌گذاری واژگانی-دستوری	Automatic Mapping Among Lexico-Grammatical Annotation Models Tagger	AMALGAM
پیکره ملی آمریکا	American National Corpus	ANC
ابزارهای زبان طبیعی آلوی	Alvey Natural Language Tools	ANLT
درخت‌بانک آسوشیتدپرس	Associated Press Treebank	AP
چاپ‌خانه آمریکا برای درخت‌بانک نابینایان	American Printing House for the Blind Treebank	APHB
پیکره نماینده سیاق‌های تاریخی انگلیسی (پیکره آرچر)	Representative Corpus of Historical English Registers Corpus	ARCHER
کد استاندارد آمریکا برای تبادل اطلاعات	American Standard Code for Information Exchange	ASCII
پیکره کنترل ترافیک هوایی	Air Traffic Control Corpus	ATC

برابرنهاد فارسی	مدخل انگلیسی	صورت اختصاری
برچسبزن سیستم نشانه گذاری متنی خودکار	Automatic Text Annotation System Tagger	AUTASYS
آرشیو باواریایی برای سیگنال های گفتاری	Bavarian Archive for Speech Signals	BAS
پیکره انگلیسی گفتاری دانشگاهی بریتانیایی	British Academic Spoken English Corpus	BASE
پیکره ملی بریتانیا	British National Corpus	BNC
بانک زبان انگلیسی	Bank of English	BoE
زبان آموزی رایانه یار	Computer Assisted Language Learning	CALL
آرشیو رایانه ای از متون انگلیسی جدید	Computer Archive of Modern English Texts	CAMET
پیکره گفتاری کمبریج و ناتینگهام از زبان انگلیسی	Cambridge and Nottingham Corpus of Discourse in English	CANCODE
پیکره مکاتبات انگلیسی متقدم	Corpus of Early English Correspondence	CEEC
پیکره گنجینه الکترونیک زبان ولزی	Cronfa Electroneg o Gymraeg	CEG
دادگان رابطه ای مرکز اطلاعات واژگانی	Centre for Lexical Information Relational Database	CELEX
استاندارد کدگذاری پیکره	Corpus Encoding Standard	CES
مرکز متون الکترونیکی حوزه علوم انسانی	Centre for Electronic Texts in the Humanities	CETH
سیستم کد برای تحلیل انسانی ترانوشته ها	Codes for the Human Analysis of Transcripts System	CHAT
سیستم تبادل داده های زبانی کودک	Child Language Data Exchange System	CHILDES
فرهنگ همکارانه بین المللی برای زبان انگلیسی	Collaborative International Dictionary of English	CIDE

صورت اختصاری	مدخل انگلیسی	برابرنهاد فارسی
CLAN	Computerized Language Analysis System	سیستم تحلیل زبان رایانه‌ای شده
CLAWS	Constituent Likelihood Automatic Word-tagging System	سیستم برچسب دهی واژگانی خودکار احتمال سازه
CLEC	Chinese Learner English Corpus	پیکره زبان‌آموز انگلیسی از گویشوران چینی
CLR	Consortium for Lexical Research	کنسرسیوم پژوهش‌های واژگانی
CMU SLM	Carnegie Mellon University–Cambridge Statistical Language Modeling Toolkit	جعبه‌ابزار مدل‌سازی زبانی آماری دانشگاه کارنچی ملون-کمبریج
Coconut	Cooperative, Coordinated Natural Language Utterances Corpus	پیکره پاره‌گفتارهای زبان طبیعی همکارانه و همپایه
COLT	Bergen Corpus of London Teenage English	پیکره برگن از انگلیسی نوجوانان لندن
CRATER	Corpus Resources and Terminology Extraction	پیکره استخراج منابع پیکره‌ای و اصطلاحات
CSAE	Corpus of Spoken American English	پیکره انگلیسی آمریکایی گفتاری
CSLU	Centre for Spoken Language Understanding Speech Corpora	مرکز پیکره‌های گفتاری ایجادشده به منظور درک زبان گفتاری
CSTR	Centre for Speech Technology Research	مرکز تحقیقات فناوری گفتار
CWBC	Corpus of Written British Creole	پیکره نیم‌زبان بریتانیایی نوشتاری
DAT	Dialogue Annotation Tool	ابزار نشانه‌گذاری گفتگو

برابرنهاد فارسی	مدخل انگلیسی	صورت اختصاری
پیکره در زمانی از انگلیسی گفتاری امروز	Diachronic Corpus of Present-day Spoken English	DCPSE
تعریف سندی	document type definition	DTD
شعبه اروپایی انجمن زبان شناسی پیکره‌ای	European Chapter of the Association for Computational Linguistics	EACL
گروه مشاورین نخبه برای استانداردهای مهندسی زبان	Expert Advisory Group on Language Engineering Standards	EAGLES
کارگروه پیکره اروپایی	European Corpus Initiative	ECI
برچسب‌زن زبانی یودیکو	Eudico Linguistic Annotator	ELAN
شبکه فعالیت زبانی اروپا	European Language Activity Network	ELAN
آژانس ارزیابی‌ها و توزیع منابع زبانی	Evaluations and Language Resources Distribution Agency	ELDA
انجمن منابع زبانی اروپا	European Language Resources Association	ELRA
قطب علمی اروپایی فناوری های زبان انسان	European Network of Excellence in Human Language Technologies	ELSNET
پیکره فعال‌سازی مهندسی زبان اقلیت	Enabling Minority Language Engineering Corpus	EMILLE
تجزیه‌گر دستور محدودیت برای زبان انگلیسی	Constraint Grammar Parser of English	ENGCG
داده‌بانک زبان دوم بنیاد علوم اروپا	European Science Foundation Second Language Databank	ESFSLD
پیکره انگلیسی بریتانیایی فرایبورگ-لوب (FLOB)	Freiburg-LOB Corpus of British English	FLOB
دادگان زبان میانی فرانسوی	French Interlanguage Database	FRIDA
پیکره فرایبورگ-براون از انگلیسی آمریکایی	Freiburg-Brown Corpus of American English	FROWN

برابرنهاد فارسی	مدخل انگلیسی	صورت اختصاری
تکه‌های درخت نامعین	Fuzzy Tree Fragments	FTF
معماری کلی برای مهندسی متن	General Architecture for Text Engineering	GATE
پیکره نفت گوانگزو	Guangzhou Petroleum English Corpus	GPEC
مرکز پژوهش‌های ارتباطات انسانی	Human Communication Research Centre	HCRC
پیکره دانشگاه علم و فناوری هنگ کنگ	Hong Kong University Of Science And Technology Corpus	HKUST
فناوری زبان انسان	human language technology	HLT
زبان نشانه‌گذاری ابرمتن	Hypertext Markup Language	HTML
آرشیو رایانه‌ای بین‌المللی از انگلیسی میانه و معاصر	International Computer Archive of Modern and Medieval English	ICAME
پیکره بین‌المللی انگلیسی	International Corpus of English	ICE
پیکره بین‌المللی از برنامه به‌کارگیری پیکره انگلیسی	International Corpus of English Corpus Utility Program	ICECUP
پیکره بین‌المللی انگلیسی آموز	International Corpus of Learner English	ICLE
میزکار پیکره‌ای موسسه پردازش ماشینی زبان	Institut für Maschinelle Sprachverarbeitung	IMS
پیکره تعاملی زبان گفتاری برای آموزش	Interactive Spoken Language Education Corpus	ISLE
پیکره تغییرات آهنگی در زبان انگلیسی	Intonational Variation in English Corpus	IviE
کلیدواژه در متن	key word in context	KWIC
پیکره لنکستر از ماندارین چینی	Lancaster Corpus of Mandarin Chinese	LCMC

برابرنهاد فارسی	مدخل انگلیسی	صورت اختصاری
پیکره پروژه نویسی کودکان لنکستر	Lancaster Corpus of Children's Project Writing	LCPW
دادگان زبانی	Linguistic DataBase	LDB
کنسرسیوم داده‌های زبانی	Linguistic Data Consortium	LDC
پیکره فراگیری ویژگی‌های نوایی زبان خارجی	Learning the Prosody of a Foreign Language Corpus	LeaP
دادگان بین‌المللی از زبان انگلیسی میانه گفتاری لووین	Louvain International Database of Spoken English Interlanguage	Lindsei
پیکره لندن-لوند	London-Lund Corpus	LLC
پیکره لنکستر-اوسلو/برگن	Lancaster-Oslo/Bergen Corpus	LOB
پیکره انگلیسی گفتاری رایانه‌خوانا	Machine-Readable Spoken English Corpus	MARSEC
برچسب‌زن حافظه‌بنیاد	Memory Based Tagger	MBT
پیکره اندازه‌گیری تکرار متن	Measuring Text Reuse Corpus	METER
پیکره همیشگان از انگلیسی	Michigan Corpus of Academic Spoken English	MICASE
پروژه برچسب‌دهی مونستر	Münster Tagging Project	MTP
برچسب‌زن اجزای کلام آنتروپی بیشینه	Maximum Entropy Part-of- Speech Tagger	MXPOST
پیکره الکترونیکی انگلیسی تاینساید نیوکاسل	Newcastle Electronic Corpus of Tyneside English	NECTE
شبکه متون انگلیسی اوایل قرن هجدهم	Network of Early Eighteenth Century English Texts	NEET
پیکره ترانویسی شده گفتاری ایرلند شمالی	Northern Ireland Transcribed Corpus of Speech	NITCS
پردازش زبان طبیعی	natural language processing	NLP
برنامه واژه‌نمای آکسفورد	Oxford Concordance Programme	OCP
بازشناسی نوری کارکتر	optical character recognition	OCR

صورت اختصاری	مدخل انگلیسی	برابرنهاد فارسی
OLAC	Open Language Archives Community	جمعیت فعال در آرشیو زبان باز
OTA	Oxford Text Archive	آرشیو متن آکسفورد
POS	part-of-speech tagging	برچسبدهی اجزای کلام
POW	Polytechnic of Wales corpus	پیکره پلی تکنیک ولز
SARA	SGML-Aware Retrieval Application	اپلیکیشن بازیابی SGML آگاه
ScoSE	Saarbrücken Corpus of Spoken English	پیکره انگلیسی گفتاری ساربروکن
SEC	Lancaster/IBM Spoken English Corpus	پیکره انگلیسی گفتاری لنکستر/IBM
SEU	Survey of English Usage Corpus	پیکره بررسی کاربرد زبان انگلیسی
SGML	Standard Generalised Markup Language	زبان نشانه گذاری عمومی استاندارد
SPAAC	Speech Act Annotated Corpus for Dialogue Systems	پیکره نشانه گذاری شده کارگفت برای سیستم های گفتگو
SUSANNE	Surface and Underlying Structural Analyses of Naturalistic English Corpus	پیکره تحلیل ساختاری روساختی و ژرف ساختی انگلیسی طبیعت گرایانه
TEI	Text Encoding Initiative	کارگروه نشانه گذاری متن
TELC	Thai English Learner Corpus	پیکره زبان آموز تایلندی انگلیسی
TESS	Text Segmentation for Speech Project	پروژه پاربندهی متن برای گفتار
TLFi	Trésor de la Langue Française Informatisé	گنجینه رایانشی زبان فرانسه
TLG	Thesaurus Linguae Graecae	گنجینه زبان یونانی

برابرنهاد فارسی	مدخل انگلیسی	صورت اختصاری
برجسبزن تراگرام ن نگز	Trigrams*n'Tags	TnT
پیکره ابزارهای پیکره تحلیل نحوی (توسکا)	Tools for Syntactic Corpus Analysis Corpus	TOSCA
پیکره زبان گفتاری و نوشتاری دانشگاهی تافل ۲۰۰۰	TOEFL 2000 Spoken and Written Academic Language Corpus	T2K-SWAL
درختبانک زبان اسپانیایی دانشگاه آزاد مادرید	Universidad Autónoma de Madrid Spanish Treebank	UAM
مرکز دانشگاهی پژوهش‌های پیکره‌ای رایانه‌ای زبان	University Centre for Computer Corpus Research on Language	UCREL
سیستم تحلیل معنایی UCREL	UCREL Semantic Analysis System	USAS
پیکره انگلیسی تجاری ولورهمپتون	Wolverhampton Business English Corpus	WBE
پیکره ولینگتون از انگلیسی گفتاری نیوزلندی	Wellington Corpus of Spoken New Zealand English	WSC
پیکره ولینگتون از انگلیسی نوشتاری نیوزلندی	Wellington Corpus of Written New Zealand English	WWC
معماری نمایه‌سازی و بازیابی XML آگاه	XML Aware Indexing and Retrieval Architecture	Xaira
زبان نشانه‌گذاری گسترش یابنده	Extensible Markup Language	XML
پیکره نثر انگلیسی باستان یورک-تورتو-هلسینکی	York-Toronto-Helsinki Corpus of Old English Prose	YCOE
پیکره روزنامه‌های انگلیسی زبان زوریخ	Zürich English Newspaper Corpus	ZEN

