



زبان‌شناسی پیکره‌ای در عمل

ویلیام جی. کرافورد
انیکو سزومی

ترجمه دکتر محسن نوبخت

زبان‌شناسی پیکره‌ای در عمل

ویلیام جی. کرافورد
اینکو سزومی

ترجمه
دکتر محسن نوبخت

تهران

۱۳۹۹



سازمان مطالعه و تدوین کتب علوم انسانی دانشگاه‌ها (سمت)
پژوهشکده تحقیق و توسعه علوم انسانی

سرشناسه: کرافورد، ویلیام ج. Crawford, William J.
عنوان و نام پدیدآور: زبان‌شناسی پیکره‌ای در عمل / ویلیام جی. کرافورد، اینیکو سزومی؛ ترجمه محسن نوبخت.
مشخصات نشر: تهران: سازمان مطالعه و تدوین کتب علوم انسانی دانشگاه‌ها (سمت)، پژوهشکده تحقیق و توسعه علوم انسانی، ۱۳۹۹.
مشخصات ظاهری: ده، ۱۹۷ ص.: جدول.
فروست: «سمت»: ۲۳۷۵. زبان‌شناسی: ۴۱.
شابک: ۹۷۸-۶۰۰-۰۲-۲۱۵۸-۴ ۲۳۰۰۰۰ ریال
وضعیت فهرست‌نویسی: فیبا
یادداشت: عنوان اصلی:
یادداشت: واژه‌نامه.
یادداشت: کتابنامه.
یادداشت: نمایه.
موضوع: زبان‌شناسی پیکره‌ای، Corpora (Linguistics)
موضوع: تحلیل زبانی، Linguistic Analysis (Linguistics)
شناسه افزوده: سیسومی، اینیکو، Csomay, Eniko
شناسه افزوده: نوبخت، محسن، ۱۳۵۸-، مترجم.
شناسه افزوده: سازمان مطالعه و تدوین کتب علوم انسانی دانشگاه‌ها (سمت)، پژوهشکده تحقیق و توسعه علوم انسانی.
The Organization for Researching and Composing University Textbooks in the Humanities (SAMT), Institute for Research and Development in the Humanities.
رده‌بندی کنگره: ۱۳۹۹ P ۱۲۸
رده‌بندی دیویی: ۴۲۰/۱۸۸
شماره کتابشناسی ملی: ۷۳۳۹۸۸۱

سازمان مطالعه و تدوین کتب علوم انسانی دانشگاه‌ها (سمت)

پژوهشکده تحقیق و توسعه علوم انسانی



Doing Corpus Linguistics; William J. Crawford, Eniko Csomay. New York: Routledge, c2016.

زبان‌شناسی پیکره‌ای در عمل
ویلیام جی. کرافورد و اینیکو سزومی
ترجمه دکتر محسن نوبخت (عضو هیئت علمی پژوهشکده تحقیق و توسعه علوم انسانی «سمت»)

ویراستار: مریم جابر

چاپ اول: پاییز ۱۳۹۹

تعداد: ۲۰۰

حروفچینی و لیتوگرافی: سمت

چاپ و صحافی: چاپ و انتشارات دانشگاه پیام نور

قیمت: ۲۳۰۰۰۰ ریال. در این نوبت چاپ قیمت مذکور ثابت است و فروشندگان و عوامل توزیع مجاز به تغییر آن نیستند.

نشانی ساختمان مرکزی: تهران، بزرگراه جلال آل احمد، غرب پل یادگار امام (ره)، روبه‌روی پمپ گاز، کد پستی ۱۶۶۳۶، تلفن ۴۴۲۴۶۲۵۰-۲.

www.samt.ac.ir

info@samt.ac.ir

@sazman_samt

هر شخص حقیقی یا حقوقی که تمام یا قسمتی از این اثر را بدون اجازه ناشر، نشر یا پخش یا عرضه یا تکثیر یا تجدید چاپ نماید، مورد پیگرد قانونی قرار خواهد گرفت.

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ



و بدانید که نژاد آریا و عرب از نژاد اروپا و امریکا و شوروی کم ندارد و اگر خودی خود را بیابد و یأس را از خود دور کند و چشمداشت به غیر خود نداشته باشد، در درازمدت قدرت همه کار و ساختن همه چیز را دارد.

صحیفه امام، ج ۲۱، ص ۴۱۷

سخن «سمت»

یکی از اهداف مهم انقلاب فرهنگی، ایجاد دگرگونی اساسی در دروس علوم انسانی دانشگاه‌ها بوده است و این امر مستلزم بازنگری منابع درسی موجود و تدوین منابع مبنایی و علمی معتبر و مستند با در نظر گرفتن دیدگاه اسلامی در مبنای و مسائل این علوم است.

ستاد انقلاب فرهنگی در این زمینه گام‌هایی برداشته بود، اما اهمیت موضوع اقتضا می‌کرد که سازمانی مخصوص این کار تأسیس شود و شورای عالی انقلاب فرهنگی در تاریخ ۶۳/۱۲/۷ تأسیس «سازمان مطالعه و تدوین کتب علوم انسانی دانشگاه‌ها» را، که به اختصار «سمت» نامیده می‌شود، تصویب کرد.

بنابراین، هدف سازمان این است که با استمداد از عنایت خداوند و همت و همکاری دانشمندان و استادان متعهد و دلسوز، به مطالعات و تحقیقات لازم بپردازد و در رشته‌های مختلف علوم انسانی به تألیف و ترجمه منابع درسی اصلی، فرعی و جنبی اقدام کند.

دشواری چنین کاری بر دانشمندان و صاحب‌نظران پوشیده نیست و به همین جهت مرحله کمال مطلوب آن باید به تدریج و پس از انتقادات و یادآوری‌های پیاپی ارباب نظر به دست آید. انتظار دارد که این بزرگواران از این همیاری و راهنمایی دریغ نورزند و با پیشنهادهای اصلاحی خود، این سازمان را در اصلاح کتاب حاضر و تدوین دیگر آثار مورد نیاز جامعه دانشگاهی جمهوری اسلامی ایران یاری دهند.

کتاب حاضر برای دانشجویان رشته زبان‌شناسی در مقاطع کارشناسی و کارشناسی ارشد به عنوان منبعی برای درس «زبان‌شناسی پیکره‌ای» ترجمه شده است. امید است علاوه بر جامعه دانشگاهی، سایر علاقه‌مندان نیز از آن بهره‌مند شوند.

فهرست مطالب

صفحه	عنوان
هشت	فهرست جدول‌ها
ده	فهرست تصویرها
۱	پیش‌درآمد
بخش ۱: درآمدی بر زبان‌شناسی پیکره‌ای و تحلیل سیاق	
۶	فصل اول: زبان‌شناسی، زبان‌شناسی پیکره‌ای و گوناگونی زبانی
۶	۱-۱ زبان و قواعد/نظام‌ها
۹	۲-۱ زبان‌شناسی پیکره‌ای چیست؟
۱۷	۳-۱ رنوس مطالب کتاب
۱۸	منابع
فصل دوم: تحلیل سیاق	
۱۹	۱-۲ چرا سیاق؟
۲۰	۲-۲ تحلیل سیاق چیست؟
۲۲	۳-۲ توصیف ویژگی‌های موقعیتی و تعیین متغیرهای آن
۲۴	۱-۳-۲ ایمیل به دوست و ایمیل به رئیس
۲۸	۲-۳-۲ نوشتار خبری و گفت‌وگوی خبری
۲۹	۳-۳-۲ ارائه در کلاس درس و در همایش
۳۰	۴-۲ ارائه تفسیر نقش‌گرا
۳۳	۱-۴-۲ نوشتار خبری و گفت‌وگوی خبری
۳۳	۲-۴-۲ ارائه در کلاس در برابر ارائه در همایش
۳۴	۵-۲ واحدهای تحلیل و سیاق‌پژوهی
۳۷	۶-۲ تمرین‌های پایان فصل
۳۹	یادداشت‌ها
۴۰	منابع
۴۱	
بخش ۲: جست‌وجو در پیکره‌های موجود	
۴۳	فصل سوم: جست‌وجوی پیکره
۴۴	۱-۳ کلیدواژه‌ها در بافت

۴۹	۲-۳ هم آیندها
۵۰	۳-۳ ان - نگاشتی ها
۵۱	۱-۳-۳ یک نگاشتی ها و واژه های منفرد
۵۷	۲-۳-۳ دو نگاشتی ها و هم آیندها
۵۸	۳-۳-۳ سه نگاشتی ها
۶۰	۴-۳-۳ چهار نگاشتی ها و خوشه های واژگانی
۶۳	۵-۳-۳ پنج نگاشتی ها و بالاتر از پنج نگاشتی ها
۶۴	۴-۳ بر حسب مقوله های نحوی
۶۶	۱-۴-۳ مشخص کردن مقوله نحوی واژه
۶۹	یادداشت
۶۹	منابع
۷۱	فصل چهارم: پروژه هایی با استفاده از پیکره های عمومی
۷۳	۱-۴ پروژه های واژه بنیان و عبارت بنیان
۷۸	۲-۴ پروژه های دستور بنیان
۸۷	منابع
	بخش ۳: ساختن پیکره خاص خودتان، تحلیل نتایج کمی خودتان و فهم داده ها
۸۹	فصل پنجم: ساختن پیکره خاص خودتان
۸۹	۱-۵ پیکره های ساخت خودتان
۹۱	۲-۵ تصمیم گیری در مورد پروژه پیکره ای
۹۳	۳-۵ ساختن پیکره
۱۰۱	۴-۵ برنامه های نرم افزاری و پیکره شما
۱۰۲	۱-۴-۵ AntWordProfiler
۱۰۴	۲-۴-۵ AntConc
۱۱۰	یادداشت
۱۱۰	منابع
۱۱۱	فصل ششم: آمار مقدماتی
۱۱۱	۱-۶ چرا تحلیل آماری انجام دهیم؟
۱۱۲	۲-۶ واژه ها، مفاهیم و فرضیه های مقدماتی
۱۱۲	۲-۶ متغیرها و مشاهده ها
۱۱۷	۲-۶-۲ سنجه های گرایش به مرکز و سنجه های تغییر پذیری
۱۲۵	۲-۶-۳ آزمون های پارامتریک و ناپارامتریک، پرسش های پژوهش و فرضیه ها

۱۲۹	۶-۳ نحوه دستیابی به نتایج آماری
۱۳۰	۶-۳-۱ آماده سازی داده ها و آمار توصیفی
۱۳۸	یادداشت
۱۳۹	منابع
۱۴۰	فصل هفتم: آزمون های آماری (آنوا، خی-دو، همبستگی پیرسون)
۱۴۰	۷-۱ آزمون های تفاوت
۱۴۰	۷-۱-۱ آنوای یک طرفه
۱۵۲	۷-۱-۲ آنوای (آنالیز واریانس) دوطرفه
۱۵۶	۷-۲ آزمون های رابطه ای
۱۵۷	۷-۲-۱ خی-دو
۱۶۲	۷-۲-۲ همبستگی
۱۶۶	۷-۳ چگونه نتایج آماری را ارائه دهیم؟
۱۶۶	۷-۳-۱ آزمون های تفاوت
۱۶۷	۷-۳-۲ آزمون های رابطه ای
۱۶۸	یادداشت ها
۱۶۸	منبع
۱۶۹	فصل هشتم: زبان شناسی پیکره ای در عمل
۱۶۹	۸-۱ تحلیل سیاقی پروژه پژوهشی
۱۷۰	۸-۲ تحلیل موقعیتی
۱۷۲	۸-۳ تحلیل زبان شناختی
۱۷۹	۸-۴ تفسیر نقش گرا
۱۸۱	۸-۵ گزارش دادن پروژه
۱۸۱	۸-۵-۱ بخش های مختلف یک مقاله پژوهشی (و پرسش های راهنما)
۱۸۳	۸-۵-۲ ارائه پژوهش
۱۸۴	منابع
۱۸۵	فصل نهم: مسیر پیش رو
۱۸۸	منابع
۱۹۰	واژه نامه انگلیسی - فارسی
۱۹۳	واژه نامه فارسی - انگلیسی
۱۹۶	نمایه

فهرست جدول‌ها

۱۲	جدول ۱-۱ هم‌آیندهای متداول در سمت راست واژه equal و identical در پیکره COCA
۱۵	جدول ۲-۱ مقایسه متن و پیکره
۲۹	جدول ۱-۲ تفاوت‌های موقعیتی کلیدی بین ایمیل به دوست و ایمیل به رئیس
۳۰	جدول ۲-۲ تفاوت‌های موقعیتی بین نوشتار خبری و گفت‌وگوی خبری
۳۲	جدول ۳-۲ تفاوت‌های موقعیتی کلیدی بین ارائه دانشجویان در همایش و سخنرانی استاد در کلاس
۳۴	جدول ۴-۲ متن‌هایی برای نوشتار خبری و یک گفت‌وگوی خبری
۳۵	جدول ۵-۲ متن‌هایی از سخنرانی کلاسی و سخنرانی در همایش
۴۰	جدول ۶-۲ متن‌هایی از ایمیل به دوست و ایمیل به رئیس
۴۶	جدول ۱-۳ توزیع مقوله‌های نحوی بعد از واژه say
۴۷	جدول ۲-۳ توزیع مقوله‌های نحوی بعد از واژه say در گفتمان شفاهی و در نثر دانشگاهی
۴۸	جدول ۳-۳ بسامد خام و نرمال‌شده برای محاسبه said و stated در سه سیاق
۴۸	جدول ۴-۳ توزیع said و stated در دو سیاق
۶۳	جدول ۵-۳ توزیع جایگاه جمله‌ای on the other hand در گفتاری و نوشتاری
۷۲	جدول ۱-۴ اطلاعات مربوط به پیکره‌های موجود در وبگاه byu.edu
۷۹	جدول ۲-۴ نمونه‌ای از فعل باقاعده و بی‌قاعده در زبان انگلیسی
۸۲	جدول ۳-۴ جهت مجهول
۹۶	جدول ۱-۵ نمونه‌هایی از پروژه‌های پیکره‌ای
۹۹	جدول ۲-۵ طرح‌واره رمزگذاری
۱۱۷	جدول ۱-۶ بسامد بندهای متممی
۱۱۷	جدول ۲-۶ بندهای متممی
۱۳۸	جدول ۳-۶ آمار توصیفی برای «عبارات طفره‌آمیز» در سه رشته از طریق SPSS
۱۴۳	جدول ۱-۷ داده‌های مربوط به ضمیر اول‌شخص در فرمت data view 'پنجره داده‌ها' در SPSS
۱۴۶	جدول ۲-۷ محاسبه مجموع مجذورهای درون‌گروهی
۱۴۷	جدول ۳-۷ محاسبه مجموع مجذورهای بین‌گروهی
۱۴۷	جدول ۴-۷ محاسبه مجموع کل مجذورها
۱۴۹	جدول ۵-۷ آمار توصیفی ضمیرهای اول‌شخص
۱۴۹	جدول ۶-۷ نتایج آنوای یک‌طرفه برای ضمیرهای اول‌شخص
۱۵۱	جدول ۷-۷ نتایج آزمون تعقیبی (شفه)
۱۵۴	جدول ۸-۷ توزیع موارد در مجموعه داده‌ها
۱۵۵	جدول ۹-۷ نتایج آنوای دوطرفه برای ضمیرهای اول‌شخص
۱۵۶	جدول ۱۰-۷ نتایج آزمون آنوای دوطرفه درمورد معناداری تعامل
۱۵۸	جدول ۱۱-۷ آزمون خی - دوی یک‌طرفه
۱۵۹	جدول ۱۲-۷ جدول خی - دوی دوطرفه

۱۶۰	جدول ۱۳-۷ جدول بازنگری شده خی - دو
۱۶۰	جدول ۱۴-۷ محاسبات آزمون خی - دو (مجموع ردیف و ستون)
۱۶۰	جدول ۱۵-۷ محاسبه مقادیر مورد انتظار
۱۶۱	جدول ۱۶-۷ محاسبه مقدار خی - دو
۱۶۵	جدول ۱۷-۷ همبستگی محاسبه شده با SPSS
۱۷۰	جدول ۱-۸ پیکره‌ای متشکل از مقالات مربوط به حل مسئله به قلم تایلندی‌های انگلیسی‌زبان
۱۷۱	جدول ۲-۸ ویژگی‌های موقعیتی سیاق‌ها
۱۷۲	جدول ۳-۸ تحلیل موقعیتی پیکره متشکل از مقالات مربوط به حل مسئله
۱۷۷	جدول ۴-۸ پنج مورد از پربسامدترین واژه‌ها در پیکره مقالات مربوط به حل مسئله به قلم ...
۱۷۸	جدول ۵-۸ پنج مورد از پربسامدترین نگاشته‌ها در پیکره مقالات مربوط به حل مسئله به قلم ...
۱۸۳	جدول ۶-۸ راهنمای ارزیابی [و درجه‌بندی] ارائه‌های دانشجویان

فهرست تصویرها

۲۵	تصویر ۱-۲ متغیرهای موقعیتی
۴۵	تصویر ۱-۳ الگوهای توزیعی واژه «say» در پیکره COCA
۴۵	تصویر ۲-۳ خطوط واژه‌نما برای واژه say در پیکره COCA
۵۳	تصویر ۳-۳ خطوط واژه‌نما برای واژه way در پیکره COCA
۵۴	تصویر ۴-۳ تحلیل متن براساس بسامد واژگانی در Word and Phrase
۵۵	تصویر ۵-۳ واژه‌های دانشگاهی برجسته‌شده متنی در Word and Phrase
۵۶	تصویر ۶-۳ الگوهای بسامدی واژه‌های منفرد در Word and Phrase
۵۷	تصویر ۷-۳ تحلیل متن با استفاده از Word and Phrase
۵۸	تصویر ۸-۳ جست‌وجوی مقوله‌های نحوی در پیکره COCA
۶۱	تصویر ۹-۳ الگوهای توزیعی ۴نگاشتی on the other hand در پیکره COCA
۶۲	تصویر ۱۰-۳ خطوط واژه‌نما برای ۴نگاشتی on the other hand در پیکره COCA
۶۷	تصویر ۱۱-۳ جست‌وجوی بسامدی در Word and Phrase
۱۰۰	تصویر ۱-۵ نمونه‌ای از فایل متنی با اطلاعات سرصفحه
۱۰۱	تصویر ۲-۵ لانه‌گیری برچسب‌های سرصفحه در AntConc
۱۰۳	تصویر ۳-۵ AntConc از فهرست‌های ۳واژه‌ای برای مقایسه بسامد واژه‌ها استفاده می‌کند
۱۰۵	تصویر ۴-۵ جست‌وجو در پیکره خودتان: خطوط واژه‌نما در AntConc برای واژه and
۱۰۵	تصویر ۵-۵ مرتب کردن در AntConc
۱۰۶	تصویر ۶-۵ دیدن فایل در AntConc
۱۰۶	تصویر ۷-۵ توزیع کلیدواژه در متن‌های کامل
۱۰۷	تصویر ۸-۵ کلیدواژه and و هم‌آیندهای آن
۱۰۸	تصویر ۹-۵ هم‌آیندها در خطوط واژه‌نما
۱۰۹	تصویر ۱۰-۵ تعیین ان -نگاشتی‌های مورد نظر تان در AntConc
۱۰۹	تصویر ۱۱-۵ تهیه فهرست واژگانی پیکره خودتان در AntConc
۱۱۹	تصویر ۱-۶ نمودار جعبه‌ای کاربرد اسم‌ها توسط معلم‌ها و دانش‌آموزان
۱۳۲	تصویر ۲-۶ Variable view 'پنجره متغیر' در SPSS
۱۳۲	تصویر ۳-۶ تب‌های مهم در Variable view 'پنجره متغیر' SPSS
۱۳۴	تصویر ۴-۶ برچسب‌های Value 'مقدار' در SPSS
۱۳۴	تصویر ۵-۶ Measures 'سنجه‌ها' در SPSS
۱۳۵	تصویر ۶-۶ Adding 'افزودن' متغیرها در SPSS
۱۳۵	تصویر ۷-۶ Data View 'پنجره داده‌ها' در SPSS
۱۳۷	تصویر ۸-۶ آمار توصیفی از مسیر Explore 'کاوش' در SPSS
۱۶۴	تصویر ۱-۷ همبستگی بین ok و I mean
۱۶۵	تصویر ۲-۷ هم‌پوشی بین «I mean» و «ok»
۱۷۳	تصویر ۱-۸ فهرست واژگانی در پیکره مقالات مربوط به حل مسئله به قلم تایلندی‌های انگلیسی‌زبان
۱۷۵	تصویر ۲-۸ زنجیره‌های ۴نگاشتی در پیکره مقالات مربوط به حل مسئله به قلم تایلندی‌های انگلیسی‌زبان

پیش در آمد

طی تجارب خود از تدریس در کلاس‌های مقدماتی زبان‌شناسی پیکره‌ای^۱، دریافته‌ایم که دانشجویان کارشناسی و تحصیلات تکمیلی زمانی اعتماد به نفس و توانایی تحلیل پیکره‌ای^۲ را می‌یابند که فرصت کار با پیکره^۳‌ها و کسب تجربه کار با ابزارهای پیکره‌ای و تحلیل پیکره‌ای به آن‌ها داده شود. درحالی که فهم اصول، رویکردها و مزایای استفاده از پیکره‌ها شناخت بنیادی از این رویکرد به تحلیل زبان را فراهم می‌سازد، اما جنبه‌ای فنی هم وجود دارد که بهترین راه آشنایی با آن، کار عملی و رویارویی با پیکره است. ما به این نکته پی برده‌ایم که دانشجویان فواید و مزایای استفاده از پیکره‌های زبانی را قبول دارند، اما چالش موجود در این زمینه آموزش نحوه کار با پیکره به آن‌هاست. دانشجویان زمان «کار» با زبان‌شناسی پیکره‌ای باید در مورد جست‌وجوی پیکره و تفسیر نتایج حاصل از جست‌وجوهای پیکره‌ای‌شان تجربه کسب کنند تا بتوانند با استفاده از این اطلاعات توضیح دهند که چرا تحلیل‌ها و یافته‌هایشان مهم است یا چه ارتباطی [با موضوع تحقیقشان] دارد. در این کتاب فرصت‌های مختلفی را برای کار بر روی پروژه‌های پیکره‌ای فراهم می‌آوریم؛ برای نیل به این هدف، ابتدا تمام فصل ۴ را به کار روی پروژه‌های پیکره‌ای کوچک‌تر اختصاص می‌دهیم. سپس در بخش ۳ اطلاعاتی را در اختیار دانشجویان می‌گذاریم تا به ساخت و تحلیل پیکره‌های خودشان بپردازند. دریافته‌ایم که با دادن فرصتی به دانشجویان برای ساخت و تحلیل پیکره خاص خودشان می‌توان تجربه ارزشمندی در زمینه طراحی پیکره در اختیار آن‌ها قرار داد یا حتی آن‌ها را به ساخت پیکره‌های دیگری برای پروژه‌های خارج از کلاس تشویق کرد.

برای آنکه به دانشجویان مجال دهیم تا در «اجرای» زبان‌شناسی پیکره‌ای تجربه کسب کنند، به‌عمد پیکره‌ها و نرم‌افزارهای مورد استفاده در پروژه‌های پیکره‌ای این کتاب را محدود کرده‌ایم. پیکره‌های فراوانی با سطوح دسترسی متفاوت و نیز نرم‌افزارهای متعددی برای تحلیل

1. corpus linguistics

2. corpus analysis

3. corpora

داده‌ها وجود دارد (برخی رایگان‌اند و برخی هم نه). هر پیکره و هر برنامه نرم‌افزاری ویژگی‌های خاص خود را دارد و ما متوجه این نکته شده‌ایم که این پیکره‌ها و نرم‌افزارهای مختلف گاهی باعث سردرگمی دانشجویانی می‌شوند که به پیکره‌ها دسترسی ندارند و/یا تلاش فراوان می‌کنند تا یک برنامه یا «اینترفیس جست‌وجو»^۱ در یک پیکره را یاد بگیرند و سپس ناگزیر می‌شوند به سراغ یادگیری برنامه دیگری بروند. برای حل این مسئله، در تمامی پروژه‌های این کتاب از مجموعه پیکره‌هایی استفاده شده که مارک دیویس^۲ در دانشگاه بریگم یانگ^۳ ساخته است.^۴ بدین ترتیب، دانشجویان می‌توانند از پیکره‌های مختلف این وبگاه استفاده کنند، بنابراین برای کار با یک اینترفیس جست‌وجوی واحد فرصت‌های فراوانی می‌یابند. برای پروژه پیکره‌ای بزرگ‌تر، بر نرم‌افزار رایگان AntConc^۵ تمرکز کرده‌ایم؛ این نرم‌افزار را لارنس آنتونی^۶ در دانشگاه واسدا^۷ تهیه و پشتیبانی کرده است.^۸ مسلماً نرم‌افزارها و پیکره‌های به کاررفته در این کتاب تنها نرم‌افزار و پیکره‌های موجود نیستند، اما به نظر ما این پیکره‌ها و نرم‌افزارها مبنای بسیار محکمی برای تحلیل پیکره‌ای فراهم کرده‌اند و به دانشجویان امکان می‌دهند تا، با استفاده از ابزارهای آسان‌یاب و کاربرپسند، «کار عملی در» حوزه زبان‌شناسی پیکره‌ای را آغاز کنند.

از آنجاکه بخش بزرگی از کار در زمینه زبان‌شناسی پیکره‌ای نیاز به بهره‌گیری از تحلیل کمی داده‌ها دارد، در فصل ۵ قدری اطلاعات آماری مقدماتی و در فصل ۶ آزمون‌های آماری مقدماتی را مطرح کرده‌ایم. همراستا با یکی از اصول هدایت‌گر این کتاب، این اطلاعات مقدماتی را شیوه‌ای برای واداشتن دانشجویان به یادگیری مبانی تحلیل می‌دانیم به این امید که این دانش را در پروژه‌های دیگر به کار بندند یا درباره روش‌های پیشرفته‌تر آماری که در آینده می‌توانند به کار گیرند، بیشتر بدانند.

۱. اینترفیس جست‌وجو (search interface)، یا فرم جست‌وجو (search form) عبارت است از مدخلی که کاربر از طریق آن می‌تواند به محتوای یک دادگان دست پیدا کند. این واسطه به کاربر امکان می‌دهد هر آنچه را در یک دادگان می‌جوید، از طریق این مدخل سؤال کند (منبع: <https://www.igi-global.com/chapter/deep-web-databases-web/20743>). فرهنگستان زبان و ادب فارسی واژه «واسطه» را برای اینترفیس تصویب کرده است - و.

2. Mark Davies
3. Brigham Young University
4. <http://corpus.byu.edu/>
5. Laurence Anthony
6. Waseda University
7. <http://www.laurenceanthony.net/software/antconc/>

چارچوب‌های توصیفی و نظری مختلفی در زبان‌شناسی پیکره‌ای به کار می‌روند. برای هدایت دانشجویان در تفسیر یافته‌های پیکره‌ای‌شان چارچوب خاصی را برگزیده‌ایم. تحلیل سیاق^۱ تأثیر بسیاری بر کار ما داشته است و معتقدیم که این رویکرد به فهم کاربرد زبان به قدری وسیع است که می‌تواند انواع پروژه‌های انتخابی دانشجویی را پوشش دهد. در فصل ۲، مبانی نوعی تحلیل سیاق را مطرح می‌کنیم و سپس از دانشجویان می‌خواهیم که هنگام ساختن پیکره خود و هنگام انجام دادن پژوهش پیکره‌ای‌شان به این چارچوب رجوع کنند. می‌دانیم که برنامه‌های نظری و توصیفی دیگری هم می‌توانند به بحث ما مرتبط باشند، اما احساس می‌کنیم که تمرکز بر یک رویکرد خاص می‌تواند تجربه بیشتری در اختیار دانشجویان قرار دهد تا بتوانند نتایج حاصل از پیکره خود را تفسیر و اهمیت یافته‌های پژوهش خود را تفهیم کنند. به باور ما، بدون دانش و کسب تجربه در استفاده از یک چارچوب خاص، دانشجویان کاملاً جذب آن جنبه‌هایی از زبان‌شناسی پیکره‌ای می‌شوند که صرفاً به «فشاردن دکمه» مربوط اند و بهای آن غافل شدن از قسمت‌های مرتبط با تفسیر داده‌های حاصل از جست‌وجو در پیکره است.

ارائه منسجم گزارش کار نیز برای ما مهم بوده است. به همین منظور، مطالبی را در مورد ویژگی‌های نگارش مقاله پژوهشی و ارائه آن آورده‌ایم (فصل ۸). هدف ما در این فصل، فراهم آوردن دستورالعملی، هم برای دانشجویان و هم برای استادان، در مورد شیوه تبیین و ارائه پروژه‌های پژوهشی خاص است.

در فصل پایانی (فصل ۹)، از دانشجویان می‌خواهیم تا انواع پیشرفته تری از پژوهش پیکره‌ای را مد نظر قرار دهند با این امید که این کتاب در حکم مقدمه‌ای در این زمینه باشد و دانشجویان را به پیگیری این ایده‌ها در سطحی حرفه‌ای تر تشویق کند و شاید حتی به شیوه‌های چشمگیری بر این رشته اثر گذارد.

بخش ۱

درآمدی بر زبان‌شناسی پیکره‌ای و تحلیل سیاق

فصل اول: زبان‌شناسی، زبان‌شناسی پیکره‌ای و گوناگونی زبانی

فصل دوم: تحلیل سیاق

زبان‌شناسی، زبان‌شناسی پیکره‌ای و گوناگونی زبانی

۱-۱ زبان و قواعد/ نظام‌ها

۲-۱ زبان‌شناسی پیکره‌ای چیست؟

۳-۱ رئوس مطالب کتاب

۱-۱ زبان و قواعد/ نظام‌ها

اگرچه همهٔ انسان‌ها برای ارتباط با هم از زبان استفاده می‌کنند، اما توصیف زبان به راحتی به کارگیری عملی آن نیست. یکی از مؤلفه‌های ممیز مطالعهٔ علمی زبان (زبان‌شناسی) عبارت است از توصیف نحوهٔ عملکرد زبان. انگلیسی‌زبانان بومی می‌توانند اسم‌های جمعی بسازند که با آواهای متفاوتی پایان می‌یابند. می‌گوییم batS و bagZ؛ نمی‌گوییم batZ و bagS. همین زبان‌وران می‌توانند صورت جمع واژه‌های بی‌معنایی را بسازند که ما هرگز آن‌ها را نشنیده‌ایم. می‌گوییم bligZ و نه bligS. همچنین، این افراد می‌دانند که هر دو جمله [۱] و [۲] قابل قبول هستند، اما جمله‌های [۳] و [۴] نمی‌توانند به یک اندازه قابل قبول باشند:

[1] We worked out the problem.

[2] We worked the problem out.

[3] We worked it out.

[4] We worked out it.

(مورد [۴] برای بسیاری از زبان‌وران بومی انگلیسی غریب می‌نماید). این حقیقت که می‌توانیم در مورد این جنبه‌های زبان انگلیسی که با تلفظ اسم‌های جمع و ترتیب واژه‌ها مرتبط هستند، توافق کنیم حکایت از این دارد که زبان در بسیاری از جنبه‌ها قابل پیش‌بینی (و نظام‌مند) است. چنین جنبه‌هایی نه تنها با آواها و ترتیب واژه‌ها، بلکه با نحوهٔ به کارگیری آن‌ها در بافت‌های گوناگون و برای اهداف مختلف نیز مرتبط‌اند. برای مثال، هیچ‌گاه برای درخواست از یک

استاد دانشگاه در مورد تمدید وقت تحویل یک تکلیف از جمله «هی بابا! یه کم مهلت بده بینم!»^۱ استفاده نمی‌کنیم. در عوض، برای چنین درخواستی، از این جمله استفاده می‌کنیم: «آیا می‌توانم خواهش کنم به من اجازه دهید تکلیفم را فردا تحویل بدهم؟ آن قدر سرم شلوغ بود که کاملاً فراموشش کردم».^۲

اگرچه توضیح این جنبه‌های خاص زبان انگلیسی سخت است، اما زبان‌وران بومی زبان انگلیسی این «قواعد» را بی‌کم و کاست در گفتار خود به کار می‌گیرند. به عبارت دیگر، یکی از جنبه‌های مهم توصیف زبان‌شناختی آن است که «قواعد» یا الگوهای ناپیدای^۳ زبان را آشکار کنیم (یعنی «دانشی که به کار می‌بریم» را به «دانش قابل توصیف» تبدیل کنیم). با آنکه ممکن است گویشوران زبان نتوانند قاعده‌های زبانی را توضیح دهند، اما این قواعد را بنا به دلایل خاصی به کار می‌بندند (و یا به عمد از آن‌ها پیروی نمی‌کنند). بخش مهمی از مطالعه زبان‌شناختی روی تحلیل زبان و تبیین آن چیزی تمرکز دارد که ممکن است در ظاهر نوعی بی‌نظمی نه‌چندان معنادار به نظر برسد [مثل همین تکواژ جمع ساز s- و z- در زبان انگلیسی].

به نظر خیلی از مردم، قواعد زبان همان دستور زبان و قواعد املائی است که در مدرسه یاد گرفته‌اند. قواعدی همچون «جمله را با حرف اضافه تمام نکنید» یا «جمله را با واژه and آغاز نکنید» قواعدی هستند که بسیاری از افراد آن‌ها را در مدرسه از بر می‌کنند. خیلی از مواقع، افراد اعتقاد راسخی در مورد این نوع قواعد دارند. برای مثال، گزیده زیر، برگرفته از وبگاهی در زمینه دستور [زبان انگلیسی]، در این مورد است که آیا باید از قاعده دستوری «مصدرها را دوتکه نکنید» پیروی کنیم یا نه.

حتی اگر طرف‌دار توصیفی بودن زبان باشید نه تجویزی بودن آن، دوتکه کردن مصدرها، در بهترین حالت، نوعی بدسلیقگی است و قطعاً کار یک سگ تازی تبیل است. خیلی راحت می‌توان از این اشتباه پرهیز کرد، کافی است فرد کمی تأمل کند و از خود پرسد چرا این امر رایج است. دو راه حل ساده برای این کار وجود دارند.

جمله [۵] را می‌توان به راحتی و با حرکت to به صورت جمله [۶] تصحیح کرد:

[5] The President decided to not attend the caucus.

[6] The President decided not to attend the caucus.

1. Hey, man. Gimme an extension.

2. Would you please allow me to hand in that assignment tomorrow? I have been so busy that I completely forgot about it.

3. implicit patterns

من استدلال خواهم کرد که این شیوه کارآمد است و امتناع از چنین تصحیح ساده‌ای فقط کار یک سنگ تازی تبیل است، اما تنها با کمی تأمل بیشتر می‌توان کلاً مصدرهای دوتکه را کنار گذاشت.

درمورد جمله [۷] چه فکر می‌کنید؟ این جمله چه ایرادی دارد؟

[7] The President has decided he will not attend the caucus.

طرف یک سنگ تازی تبیل بوده. عزت نفس نویسندگان کجا رفته است؟ ([https://](https://gerryellenson.wordpress.com/2012/01/02/to-not-splitinfinitives/))

(gerryellenson.wordpress.com/2012/01/02/to-not-splitinfinitives/)

مثال‌هایی از این دست غیر معمول نیستند. می‌توان تنها با نگاه به نامه‌های ارسالی به سردبیر روزنامه‌ها یا حتی پُست‌های موجود در وبلاگ‌ها، افراد زیادی را یافت که درمورد اهمیت پیروی از قواعد دستوری بخصوص اعتقادی راسخ دارند.

تا اینجا، دیده‌ایم که «قواعد» دو کار متفاوت انجام می‌دهند: (۱) توصیف الگوهای ناپیدای زبان که به صورت طبیعی در زبان یافت می‌شوند و (۲) تجویز صورت‌های زبانی مشخص و مورد قبول جامعه. البته هر دو دیدگاه توصیفی^۱ و تجویزی^۲ ناظر به قواعد زبان هستند؛ قواعد تجویزی دستوراتی را درباره کاربرد زبان دیکته می‌کنند؛ حال آنکه قواعد توصیفی، به دور از هر قضاوتی، اظهاراتی درمورد الگوهای زبانی فراهم می‌کنند. هر دو جنبه تجویزی و توصیفی زبان مفیدند. هنگام نوشتن مقاله دانشگاهی یا نامه‌ای رسمی قراردادهای زبانی خاصی رعایت می‌شوند. قاعده تجویزی می‌تواند رهنمون‌هایی درمورد ارتباط‌های مؤثر در اختیار ما قرار دهد. اما رویکردهای توصیفی می‌توانند در زمینه کشف الگوهای ناپیدای زبانی مؤثر باشند (همانند مثال‌هایی که پیش‌تر توصیف شدند). می‌توان از رویکردهای توصیفی برای بررسی این نکته هم استفاده کرد که کاربران زبان چگونه قواعد تجویزی را رعایت می‌کنند.

مفهوم قواعد زبان پرسش جالب دیگری را مطرح می‌کند: چرا گاهی از این قواعد تبعیت و گاهی از آن‌ها «تخطی» می‌شود؟ به قاعده تجویزی پیش‌گفته درباره مصدر توجه کنید. آیا درست است بگوییم کسانی که عبارت to not attend را به کار می‌برند از قاعده پیروی نمی‌کنند؟ از جهاتی، شاید این درست باشد؛ اما مسئله‌ای دیگر - یا شاید نوعی سوء برداشت - درمورد زبان وجود دارد که نیازمند توجه بیشتری است و می‌تواند مبنایی برای کتاب حاضر باشد: و آن نقش گوناگونی زبانی^۳ است. گوناگونی و تغییر زبان حقیقتی انکارناپذیر است. زبان انگلیسی مردم جزایر بریتانیا با انگلیسی آمریکایی متفاوت است. نوع زبان انگلیسی در جزایر بریتانیا یا ایالات

1. descriptive
2. prescriptive
3. language variation

متحدہ نیز از منطقہ‌ای بہ منطقہ دیگر و در میان مردمانی از طبقہ‌های مختلف اجتماعی - اقتصادی متفاوت است. نوع انگلیسی‌ای کہ ۱۵۰ سال پیش در امریکا بہ کار می‌رفت با نوع انگلیسی امروزی ایالات متحدہ کاملاً متفاوت است. حتی یک فرد ہم دارای گونه‌های زبانی متفاوت است. هدف مطالعہ گونه‌های زبانی رسیدن بہ فہمی درمورد نحوہ تغییر و گوناگونی زبانی بہ دلایل مختلف و در بافت‌های مختلف است. درمورد نحوہ بررسی و شناخت گونه‌های زبانی دیدگاہ‌های مختلفی وجود دارد. دیدگاہ مورد نظر ما، همان‌طور کہ عنوان کتاب مشخص می‌کند، مربوط بہ حوزه‌ای از مطالعہ زبان است کہ «زبان‌شناسی پیکره‌ای» نامیدہ می‌شود.

۱-۲ زبان‌شناسی پیکره‌ای چیست؟

یکی از راه‌های فہم تحلیل زبانی و زبان با استفادہ از رویکرد زبان‌شناسی پیکره‌ای ممکن می‌شود. زبان‌شناسی پیکره‌ای بہ بررسی این مسئلہ می‌پردازد کہ زبان در بافت‌های مختلف عملاً چگونه بہ کار می‌رود و چطور یک گونه زبان از بافتی بہ بافت دیگر تغییر می‌کند. درحالی کہ شناخت گونه‌های زبان و تفاوت‌های بافتی هدف مشترک محققان سایر حوزه‌های پژوهش زبان‌شناسی است، زبان‌شناسی پیکره‌ای با بررسی حجم زیادی از متن‌هایی کہ در شرایط مشابهی تولید شدہ‌اند، گوناگونی و کاربرد زبان را توصیف می‌کند. مفہوم «شرایط»^۱ یا «بافت»^۲ یا «موقعیت»^۳ بستگی بہ نحوہ تعریف ہر محقق از آن‌ها دارد. در مطالعات زبان‌شناسی پیکره‌ای، غالباً تفاوت کلی بین دو 'شیوہ'^۴ مختلف تولید زبان - یعنی زبان نوشتار و زبان گفتار - مد نظر قرار می‌گیرد. در عرصہ زبان نوشتاری، ممکن است فرد بہ بررسی بافت‌هایی نظیر نگارش خبری، پیام‌رسانی متنی، یا نگارش دانشگاهی توجہ داشتہ باشد. در عرصہ زبان گفتاری، ممکن است فرد بہ بررسی گزارش خبری، گفت‌وگوی رو در رو یا سخنرانی‌های دانشگاهی توجہ نشان دہد. اگرچہ ہم پیام‌رسانی متنی و ہم نگارش دانشگاهی نوشتار محسوب می‌شوند، اما ہدف پیام‌رسانی متنی کاملاً با ہدف نگارش دانشگاهی فرق می‌کند. بنابراین، در این دو بافت نوشتاری متفاوت انتظار داریم درجہ‌ای از گوناگونی زبانی وجود داشتہ باشد. ہمین نکتہ درمورد گفت‌وگوی رو در رو و سخنرانی دانشگاهی نیز صدق می‌کند؛ ہر دو گفتاری‌اند، اما اہداف مختلف و در نتیجہ ویژگی‌های زبان‌شناختی متفاوتی

1. circumstances
2. context
3. situation
4. mode

دارند. به بیانی کلی‌تر، می‌توانیم انتظار داشته باشیم که زبان‌گفتاری (با تمام اهداف و بافت‌های مختلفش) با صورت‌های نوشتاری زبان متفاوت باشد. در بسیاری از گونه‌های زبان نوشتاری فرصت برنامه‌ریزی و بازنگری متن را داریم، اما معمولاً در زبان گفتاری چنین فرصتی نداریم. در فصل آتی، به بررسی این نکته خواهیم پرداخت که چگونه شرایط (یا متغیرهای موقعیتی) مختلف می‌توانند بر کاربرد زبان تأثیر بگذارند. اما پیش از این، مایلیم به طور خلاصه منظورمان از پیکره را توضیح دهیم.

پیکره عبارت است از مجموعه‌ای [از نمونه‌های زبانی] که بازنمایانگر^۱ زبان هستند و می‌توان از آن‌ها برای اظهار نظر دربارهٔ کاربرد زبان استفاده کرد. زبان‌شناسی پیکره‌ای در پی فهم این مسئله است که افراد چگونه زبان را در بافت‌های گوناگون به کار می‌برند. پیکره مجموعه‌ای از نمونه‌های^۲ زبانی فراوان است؛ این نمونه‌ها (که در زبان پیکره‌ای متن نامیده می‌شوند) ویژگی‌های بافتی یا موقعیتی^۳ مشترکی دارند. این متن‌ها مجموعاً تحلیل می‌شوند تا بتوان به درکی از نحوهٔ کاربرد زبان در این بافت‌های مختلف رسید. نتیجهٔ چنین تحلیلی رسیدن به مجموعه‌ای از الگوهای زبانی تکراری در پیکره است که می‌توانند توضیحی دربارهٔ کاربردهای زبانی ارائه دهند یا مبنایی برای تحلیل‌های زبانی بعدی بشوند. روشی متداول در پژوهش‌های پیکره‌ای توجه به پیرامون [یا بافت هم‌نشین] یک واژه یا عبارت خاص است با این هدف که چه واژه‌های دیگری با واژه مورد نظر می‌آیند (یا با آن هم‌آیند^۴ می‌شوند). برای مثال، از COCA^۵ پیکرهٔ انگلیسی امریکایی معاصر^۶ - مجموعه‌ای در دسترس عموم و متشکل از ۴۵۰ میلیون واژهٔ انگلیسی امریکایی - استفاده خواهیم کرد تا کاربرد دو واژهٔ 'equal' و 'برابر' و 'identical' مشابه را بررسی کنیم.

از خیلی جهات، دو واژهٔ equal و identical می‌توانند به یک چیز (دو چیز شبیه به هم) دلالت کنند و غالباً هم معنا تلقی می‌شوند. برای مثال، این دو واژه در جملهٔ [۸] معنای کلی مشابهی دارند:

[8] These two students are equal/identical in their performance on the exam.

[۸] این دو دانشجو در امتحان خود عملکردی برابر / مشابه داشتند.

1. representative

2. examples

3. collocates

4. Corpus of Contemporary American English

قابل دسترس در <http://corpus.byu.edu/coca/>.

اگر از ما خواسته شود واژه equal را تعریف کنیم، می‌توانیم از واژه identical در تعریف خود استفاده کنیم (و برعکس). با این حال، با دقت در نحوه کاربرد واقعی این دو واژه در پیکره، تصویر متفاوتی ظاهر می‌شود. پیکره COCA به ما نشان می‌دهد که اگرچه این دو واژه گاهی هم‌معنا هستند، اما رفتار بسیار متفاوتی دارند. به احتمال زیاد، از عبارت‌هایی مانند equal opportunity 'فرصت برابر'، equal rights 'حقوق برابر' و equal protection 'حمایت برابر' به جای identical opportunity 'فرصت مشابه'، identical rights 'حقوق مشابه' یا identical protection 'حمایت مشابه' استفاده می‌کنیم. بعید است بگوییم equal twins 'دوقلوهای برابر' یا equal copies 'کپی‌های برابر'، اما در عوض عبارت‌هایی مثل identical twins 'دوقلوهای مشابه' و identical copies 'کپی‌های مشابه' را به کار می‌بریم. بررسی واژه‌هایی که بعد از identical و equal می‌آیند نشان می‌دهد که واژه equal بیشتر مفاهیمی انتزاعی از قبیل rights، opportunities و protection را توصیف می‌کند، حال آنکه identical به احتمال زیاد اسم‌هایی عینی، ملموس و مادی مانند items، twins، houses را توصیف می‌کند. بدون ارجاع به حجم زیادی از متن‌ها، بعید است بتوانیم به چنین یافته‌هایی برسیم. این مثال نشان می‌دهد چگونه زبان‌شناسی پیکره‌ای می‌تواند اطلاعاتی در مورد کاربرد زبان در اختیار زبان‌شناسان قرار دهد تا نحوه کاربرد واقعی زبان در بافت‌های موثق را دریابند.

علاوه بر این، پیکره می‌تواند تفاوت بسامد equal و identical را به ما نشان دهد (رجوع کنید به جدول ۱-۱). پنج هم‌آیند اول واژه equal بین ۹۵۰ و ۴۰۵ بار در پیکره COCA ظاهر شده‌اند؛ و پنج هم‌آیند اول identical بین ۴۱۷ و ۲۰ بار. به عبارت دیگر، درمی‌یابیم equal از identical پربسامدتر است؛ زیرا بسامد هم‌آیندها تفاوت زیادی را بین دو واژه نشان می‌دهد. در واقع، واژه equal ۲۲۴۸۰ بار و واژه identical ۸۰۸۰ بار در این پیکره ظاهر می‌شوند.

پیکره، علاوه بر اطلاعاتی در مورد هم‌آیندها و بسامد، به ما امکان می‌دهد تا ببینیم تا چه اندازه از قواعد تجویزی خاصی پیروی می‌شود. اجازه دهید ببینیم پیکره چه اطلاعاتی در مورد «دو تکه کردن مصدرها» به ما می‌دهد. پیش‌تر در همین فصل دیدیم برخی از افراد چنان [از نقض] این قاعده خشمگین می‌شوند که ضعف‌های شخصیتی جدی‌ای را به نویسندگان ناقض آن نسبت می‌دهند [مثلاً، همان‌طور که پیش‌تر دیدیم، آن‌ها را سگ تازی تنبل می‌خوانند]. پیکره COCA نشان می‌دهد که مثال‌هایی را که در اختیار داریم عبارت‌اند از: better understand (۸۷۴ بار در پیکره) در مقایسه با to understand better (۹۴ بار)؛ و to really get (۳۴۹ بار) در

مقایسه با to get really (۱۵۱ بار) علاوه بر این، زنجیره واژه‌های to better understand بیشتر در متن‌های دانشگاهی ظاهر می‌شوند، حال آنکه زنجیره to get really بیشتر در بافت‌های زبان گفتاری می‌آیند. این شواهد حاکی از آن‌اند که اگر به کاربرد واقعی زبان توجه کنیم، حکم تجویزی متقنی مانند «هرگز مصدرها را دو تکه نکنید» با اشکال‌هایی جدی مواجه می‌شود. در نوشتار رسمی دانشگاهی چند نمونه از کاربرد مصدرهای دو تکه مشاهده می‌شود، اما نمونه‌های دیگر آن در زبان گفتار پربسامدترند. به بیان دیگر، اگرچه در مورد برخی از قواعد کاربرد دستور زبان انگلیسی دیدگاه‌های متقنی وجود دارد، اما بسیاری از این قواعد در مواجهه با مثال‌های موثقی از کاربرد واقعی زبان نقض می‌شوند؛ بدین معنا که «قاعده» «هرگز مصدرها را دو تکه نکنید» در برخی از گونه‌های زبانی رعایت نمی‌شود (یعنی در کاربرد این قاعده گوناگونی وجود دارد).

جدول ۱-۱ هم‌آیندهای متداول در سمت راست واژه equal و identical در پیکره COCA

تعداد	اولین واژه سمت راست	تعداد	اولین واژه سمت راست
۴۱۷	TWINS	۹۵۰	OPPORTUNITY
۲۴۷	TWIN	۸۸۱	RIGHTS
۲۹	COPIES	۷۳۳	PROTECTION
۲۷	ITEMS	۴۸۵	ACCESS
۲۰	HOUSES	۴۰۵	EMPLOYMENT

پیش‌تر با استفاده از چند مثال دیدیم که پیکره چه اطلاعاتی به ما می‌دهد. حال به ویژگی‌های ممیز زبان‌شناسی پیکره‌ای مورد نظر در این کتاب می‌پردازیم. در مفهومی کلی، «پیکره» عبارت است از مجموعه‌ای از متن‌ها که مبنای تحلیل هستند. برای مثال، فردی ممکن است نمونه‌هایی از سرمقاله‌های خبری مربوط به موضوعی خاص را جمع‌آوری کند و به آن «پیکره» بگوید. با این حال، آیا می‌توانیم بگوییم که این فرد در حال «کار در زمینه» زبان‌شناسی پیکره‌ای است؟ بiber، کنراد و ریپن^۱ در کتابشان تحت عنوان *زبان‌شناسی پیکره‌ای: بررسی زبان، ساختار و کاربرد*^۲ (۱۹۹۸)، در تعریف پژوهش پیکره‌ای، این ویژگی‌ها را برای «زبان‌شناسی پیکره‌ای» برمی‌شمرند:

1. Biber, Conrad & Reppen

2. *Corpus Linguistics: Investigating Language, Structure and Use*

۱. زبان‌شناسی پیکره‌ای امری تجربی است و الگوهای واقعی کاربرد در متن‌هایی با زبان طبیعی را بررسی می‌کند.

۲. در زبان‌شناسی پیکره‌ای مجموعه بزرگی از متن‌های طبیعی که به‌طور اصولی جمع‌آوری شده‌اند مبنای تحلیل قرار می‌گیرد؛ این مجموعه متن‌ها «پیکره» نامیده می‌شود.

۳. پژوهش پیکره‌ای به‌طور گسترده‌ای از رایانه برای تحلیل بهره می‌برد و آن‌ها را به‌دو روش خودکار و تعاملی به‌کار می‌گیرد.

۴. زبان‌شناسی پیکره‌ای، هم به تکنیک‌های تحلیلی کمی متکی است و هم به تکنیک‌های تحلیلی کیفی (بیر، کنراد و رپین، ۱۹۹۸: ۴).

در بحث پیش، نمونه‌هایی از دو ویژگی اول را دیده‌ایم. زبان‌شناسی پیکره‌ای به بررسی زبانی می‌پردازد که به‌منظور پژوهش زبان‌شناختی تولید نشده است. زبان‌شناسی پیکره‌ای، به جای استفاده از شم زبانی یا قضاوت‌های مربوط به زبان (یا به جای برساختن مثال‌هایی زبانی برای نشان دادن یک نکته بخصوص)، از زبانی استفاده می‌کند که به‌طور واقعی به‌کار می‌رود. از این منظر، زبان پیکره «طبیعی» یا «مؤثق» است؛ زیرا نحوه کاربرد واقعی زبان در یک بافت مشخص را نشان می‌دهد. درمورد ویژگی دوم، پیش‌تر مزایای ارجاع به حجم زیادی از نمونه‌های زبانی را دیده‌ایم؛ الگوهای زبانی بخصوصی هستند که تنها با مراجعه به حجم زیادی از نمونه‌های زبانی به آن‌ها پی می‌بریم. باین‌حال، مهم این است که پیکره فقط شامل آن دسته از متن‌هایی باشد که کاربرد زبان در بافت خاصی را نشان می‌دهند. مثلاً، برای بررسی ویژگی‌های گفت‌وگویی رو در رو، نیازی نیست نمونه‌هایی از زبان گفتار سخنرانی‌های دانشگاهی یا سیاسی را در پیکره خود بیاوریم. هرچند گفت‌وگو و سخنرانی، از این حیث که هر دو گفتاری‌اند، وجه مشترکی با هم دارند (در تقابل با زبان نوشتاری)، اما در ویژگی‌های مهم دیگر (از قبیل هدف گفت‌وگو و مقدار تعامل بین مشارکان) وجه اشتراکی ندارند. بنابراین، یک پیکره بخصوص باید هم حجم زیادی داشته باشد و هم متن‌هایی از یک نوع (یعنی متن‌هایی با ویژگی‌هایی مشترک) را در خود جای داده باشد. مفهوم «متن» در زبان‌شناسی پیکره‌ای ممکن است با برداشت عام شما از متن متفاوت باشد. در زبان‌شناسی پیکره‌ای، متن به هر نمونه‌ای از زبان اطلاق می‌شود که برای هدفی مؤثق به‌کار می‌رود.

از این منظر، متن‌ها عبارت‌اند از: زبان نوشتاری (مانند مقاله‌های روزنامه‌ها، نامه‌ها و داستان) و بازنمایی نوشتاری زبان گفتاری (گفت‌وگوهای رو در رو، [گفت‌وگوهای] سریال‌های طنز و سخنرانی‌های دانشگاهی).

ویژگی‌های ۳ و ۴ زبان‌شناسی پیکره‌ای [که پیش‌تر ذکر شده‌اند] به اهمیت استفاده از رایانه در تحلیل زبان و همین‌طور به رویکردهای تحلیلی مختلف اشاره دارند. مشکل بتوان تصور کرد که چگونه می‌توان بدون کمک رایانه پیکره‌ای ۴۵۰ میلیون واژه‌ای مانند پیکره COCA را به‌منظور شناسایی ویژگی‌های زبانی مشخصی تحلیل کرد. با توجه به پرشماری متن‌ها و سهولت نسبی دسترسی به انبوه نمونه‌ها، تحلیل پیکره‌ای نه تنها مستلزم شمارش برخی از واحدهاست (یعنی تحلیل کمی)، بلکه متکی به یافتن علل این یافته‌های کمی یا ارائه توضیحات آماری درباره آن‌ها هم هست. در فصل‌های ۳ و ۴ برخی از شیوه‌های مشخص «کاوش در پیکره از طریق انواع نرم‌افزارهای تحلیلی» را مطرح خواهیم کرد؛ این نرم‌افزارها امکان تحلیل‌های مختلف را میسر می‌سازند. لازم است یادآوری شود که روش‌های پیکره‌ای صرفاً به استفاده از رایانه‌ها برای یافتن نمونه‌های مناسب محدود نمی‌شوند؛ بلکه این روش‌ها، در عین حال، به منظور ارائه تفسیری کیفی، بر تحلیل و توصیف مثال‌ها هم تأکید دارند.

علاوه بر این چهار ویژگی پیکره، النا توگنینی - بونلی^۱ در کتاب خود تحت عنوان *زبان‌شناسی پیکره‌ای در عمل*^۲ (۲۰۰۱) دیدگاه‌های مفیدی در تعریف زبان‌شناسی پیکره‌ای ارائه می‌دهد. او تفاوت‌های بین خواندن یک متن واحد و استفاده از ابزارهای زبان‌شناسی پیکره‌ای را برای بررسی مجموعه‌ای از متن‌ها (یعنی پیکره) توضیح می‌دهد. برای نشان دادن این تفاوت‌ها، به نوع خاصی از متن، یعنی مقاله روزنامه، اشاره خواهیم کرد. معمولاً مقاله روزنامه از ابتدا تا انتها خوانده می‌شود و خواننده از این طریق محتوای متن را درمی‌یابد و آن را به دیدگاه‌های دیگر درمورد یک موضوع خاص ربط می‌دهد. در انگلیسی، متن به‌صورت افقی و از چپ به راست خوانده می‌شود، و از این رو، تمام متن‌ها را می‌توان کنش یا رویداد ارتباطی منفردی تلقی کرد (رجوع کنید به جدول ۱-۲). اگر فرضاً قرار باشد ۳۰۰ سرمقاله جمع‌آوری و برای تحلیلشان از رویکرد پیکره‌ای استفاده شود، شیوه کاربرد آن‌ها [با شیوه کاربرد متن‌های دیگر] کاملاً فرق می‌کند. متن‌های یک پیکره، برخلاف سرمقاله روزنامه، از ابتدا تا انتها و به‌صورت افقی خوانده نمی‌شوند؛ در عوض، این متن‌ها مجموعه‌ای از رویدادهایی مختلف (اما مرتبط) هستند که نه صرفاً به‌طور کلی، بلکه به‌صورت بخش‌بخش بررسی می‌شوند؛ بدین معنا که بسیاری از نمونه‌های مربوط به یک ویژگی مشخص با یکدیگر مرتبط می‌شوند. از این زاویه دید، پیکره نه به‌صورت افقی بلکه به‌صورت عمودی خوانده می‌شود - یعنی در هر بررسی، نمونه‌های بی‌شماری مربوط به یک مشخصه زبانی خاص مد نظر قرار می‌گیرند.

1. Elena Tognini-Bonelli

2. *Corpus Linguistics at Work*

(Part of) a newspaper text

A corpus of newspaper texts

The site appeared to be working again late Thursday. A spokesman for the Motion Picture Association of America said in an emailed statement that the group's site also had been hacked, but it too appeared to be working later in the evening. If the motion picture and television industry has always been a strong supporter of free speech, the spokesman said. We strongly condemn any attempts to silence any groups or individuals. Matthew Barakat reported from McLean, Virginia.

2012	NEWS	Assoofness	A B C	occasion and watching dozens of NFL games every season with a hope in his lap always during special teams pla	
1	2012	NEWS	Assoofness	A B C	working later in the evening. # "The motion picture and television industry has always been a strong supporter of the
1	2012	NEWS	Assoofness	A B C	generate massive numbers of downloads, according to the indictment. Such content was almost always copyright prot
1	2012	NEWS	Assoofness	A B C	generate massive numbers of downloads, according to the indictment. Such content was almost always copyright prot
1	2012	NEWS	Assoofness	A B C	working later in the evening. # "The motion picture and television industry has always been a strong supporter of the
1	2012	NEWS	Assoofness	A B C	who plan to stop down in 2014, also said the university "must always be mindful of the need for institutional humil
1	2012	NEWS	Assoofness	A B C	he's handled it has really made us handle our situation better. He's always been a great leader for us. # "Those who
1	2012	NEWS	Assoofness	A B C	expects that this year, and Beethoven is hoped it won't. # "It's always easier to win than lose. When you're losing, it
1	2012	NEWS	Assoofness	A B C	generate massive numbers of downloads, according to the indictment. Such content was almost always copyright prot
0	2012	NEWS	Assoofness	A B C	working later in the evening. # "The motion picture and television industry has always been a strong supporter of the
1	2012	NEWS	Assoofness	A B C	have been slight improvements but the situation hasn't changed fundamentally. # "It always has been hard for small
2	2012	NEWS	Assoofness	A B C	might offer occasional behind-the-scenes advice on nutrition matters, Sobel says. "We were always thought of as, G
3	2012	NEWS	Assoofness	A B C	Ed Angel's, "the definitive book on Kersh's disappearance. # "Americans always thought of themselves, and this ac
4	2012	NEWS	Assoofness	A B C	up analyzing it," she says. "I, as a professor, always accepted the role of law. If a court or jury made a finding,
5	2012	NEWS	Assoofness	A B C	she wrote that she was fighting for her daughter. # "Our family will always remember today as the day truth was de
6	2012	NEWS	Assoofness	A B C	now friends. But Cain still struggles with her conscience. The guilt will probably always plague her, she says. She told
7	2012	NEWS	Assoofness	A B C	# As for Sobel's, "the only concern that she has, and has always had, is the financial well-being of Arctic comics,
8	2012	NEWS	Assoofness	A B C	in the hands of your vehicle driver. # "It won't always drive. # In the 1980s, Honda and Toyota domi

نکته پایانی در بررسی پژوهش پیکره‌ای به دیدگاه‌های مختلف پژوهشگران در مورد زبان‌شناسی پیکره‌ای مرتبط است. النا توگنینی - بونلی (۲۰۰۱) بین پژوهش پیکره‌بنیان^۱ و پژوهش مشتق‌ازپیکره^۲ فرق قائل شده است. در رویکرد پیکره‌بنیان، یافته‌های پیکره‌ای پیشین و یا مسائل مشخصی در مورد کاربرد زبان پژوهشگر حوزه زبان‌شناسی پیکره‌ای را در مسیر پژوهش هدایت می‌کنند، بدین معنا که محقق پیش از آنکه آن ویژگی زبانی را جست‌وجو کند می‌داند که به دنبال چیست. پیش‌تر، در این مورد مثال «مصدرهای دوتکه‌ای» را مطرح کردیم. در این مورد، به اهم نکات دیدگاه یک مقاله درباره این نوع مصدر دوتکه اشاره شد و این دیدگاه مبنایی برای پیکره‌پژوهی ما در خصوص چگونگی کاربرد واقعی مصدرهای دوتکه‌ای قرار گرفت. باتوجه به این قاعده تجویزی در مورد پرهیز از دوتکه کردن مصدر [در انگلیسی]، تصمیم گرفتیم ببینیم که یک پیکره - یعنی نمونه‌های متنی از یک گفتمان طبیعی - چه نکاتی را در مورد نحوه به کارگیری این قاعده توسط کاربران زبان به ما نشان خواهد داد. مثال‌های دیگر شامل هر ویژگی زبان‌شناختی دیگری می‌شود که می‌خواهیم آن را جست‌وجو کنیم و می‌دانیم به دنبال چه هستیم، مانند یک واژه مشخص یا یک مقوله دستوری مشخص. باین حال، در تمام این موارد از قبل (پیش از شروع تحقیق) می‌دانستیم که در پیکره به دنبال چه هستیم.

رویکرد دیگری که توگنینی - بونلی آن را در زمینه آگاهی از کاربرد زبان معرفی کرده، استفاده از شیوه مشتق‌ازپیکره است. برخلاف روش پیش، در شیوه مشتق‌ازپیکره، محققان بدون استفاده از مجموعه‌ای از معیارهای ثابت و از پیش تعیین شده دست به تحلیل پیکره‌ای می‌زنند؛ آن‌ها با استفاده از شیوه‌های خاصی در پی این هستند که خود پیکره چه نوع الگوهای زبانی‌ای را آشکار می‌سازد. پژوهشگران این الگوها را از متن‌ها استخراج و آن‌ها را مستند می‌کنند؛ سپس به توصیف و تفسیر آن‌ها می‌پردازند. مطالعات مربوط به هم‌آیندها و «خوشه‌های واژگانی»^۳ نمونه‌هایی از پژوهش‌هایی هستند که با این رویکرد انجام می‌شوند. خوشه‌های واژگانی عبارت‌اند از پربسامدترین توالی‌های ۴ واژه‌ای در یک سیاق. در بسیاری از مواقع نمی‌دانیم که پربسامدترین توالی‌ها کدام‌اند [یعنی نمی‌دانیم با چه نتایجی روبه‌رو خواهیم شد]. بنابراین، برای استخراج این الگوها از برنامه‌های رایانه‌ای بخصوصی استفاده می‌کنیم.

1. corpus-based
2. corpus-driven
3. lexical bundles

سپس پژوهشگران این الگوها را به لحاظ دستوری (بیر و همکاران، ۱۹۹۹) و نقشی (بیر و همکاران، ۲۰۰۴) تحلیل می‌کنند. نه تنها واحدهای واژگانی بلکه ویژگی‌های دستوری را هم می‌توان از این طریق جست‌وجو کرد. نوعاً پژوهشگران به دنبال الگوهای دستوری هم‌وقوع^۱ در متن‌ها هستند تا از این طریق بتوانند سیاق‌ها را مشخص کنند. در فصل آخر کتاب، در مورد «مسیر پیش‌روی» کار با زبان‌شناسی پیکره‌ای، به طور خلاصه این شیوه را توصیف خواهیم کرد. به هر روی، در این نوع مطالعات، پیش از شروع تحقیق، نمی‌دانیم که [در پیکره] به دنبال کدام ویژگی زبانی هستیم.

۳-۱ رتوس مطالب کتاب

کتاب حاضر به سه بخش اصلی تقسیم می‌شود. در بخش ۱ به معرفی مفهوم پیکره می‌پردازیم و زبان‌شناسی پیکره‌ای را به عنوان رویکردی به مطالعه زبان و مرتبط با تحلیل زبان واقعی و مؤتق مطرح خواهیم کرد، و با استفاده از حجم زیادی از متن‌ها روی گوناگونی زبانی تمرکز می‌کنیم (فصل ۱). در ادامه، چارچوب تحلیل سیاق را برای تفسیر یافته‌های پیکره‌ای مطرح می‌کنیم (فصل ۲). در بخش ۲ روی نحوه استفاده از پیکره‌های موجود تأکید می‌کنیم. از این رو، مجموعه‌ای از ابزارهای جست‌وجو و واحدهای زبانی را معرفی می‌کنیم که می‌توانند مبنایی برای تحلیل پیکره‌های از پیش موجود باشند (فصل ۳). سپس به منظور معرفی مبانی نحوه کار با پیکره‌ها، تفسیر داده‌ها و ارائه یافته‌های پژوهش، ۱۲ پروژه مختلف را مطرح می‌کنیم که در آن‌ها از پیکره‌های موجود برخط استفاده می‌شود (فصل ۴). پس از بررسی مبانی مربوط به تحلیل پیکره و چارچوب تحلیلی، خوانندگان آماده می‌شوند پیکره خود را طراحی کنند و مطالعه پژوهشی خود را انجام دهند. در بخش ۳، مراحل ساختن پیکره آموزش داده می‌شود (فصل ۵) و شیوه‌های آماری (فصل‌های ۶ و ۷) در تفسیر داده‌ها بررسی می‌گردد. فصل ۸ مراحل گام به گام نوشتن و ارائه پژوهش را معرفی می‌کند. از آنجا که این کتاب مقدماتی، در حکم مبنایی نحوه انجام پژوهش پیکره‌ای است، بسیاری از مسائلی را که امروزه در زبان‌شناسی پیکره‌ای بررسی می‌شوند مطرح نخواهیم کرد. در فصل ۹، در مورد برخی از این مسائل بحث می‌کنیم با این امید که این کتاب به قدر کافی پژوهش پیکره‌ای را به شما آموزش دهد تا بتوانید به پژوهش‌های پیشرفته‌تری در این زمینه بپردازید.

منابع

- Biber, D., S. Conrad & R. Reppen (1998) *Corpus Linguistics: Investigating Language, Structure and Use*. Cambridge: Cambridge University Press
- Biber, D., S. Conrad & V. Cortes (2004) 'If you look at . . . : Lexical Bundles in university teaching and textbooks', *Applied Linguistics* 25/3: 371-405
- Biber, D., S. Johansson, G. Leech, S. Conrad & E. Finegan (1999) *Longman Grammar of Spoken and Written English*, New York: Longman
- Tognini-Bonelli, E. (2001) *Corpus Linguistics at Work*, Amsterdam: John Benjamins