

درآمدی بر زبان‌شناسی پیکره‌ای

سندرین زوفری

ترجمه

دکتر محمدحسن ترابی

(دکتری زبان‌شناسی از پژوهشگاه علوم انسانی و مطالعات فرهنگی)

دکتر شهره صادقی

(دکتری زبان‌شناسی از دانشگاه علامه طباطبائی)



سازمان اسناد و کتابخانه ملی
جمهوری اسلامی ایران

تهران، ۱۴۰۳

درآمدی بر زبان‌شناسی پیکره‌ای

نویسنده: سندرین زوفری

ترجمه: دکتر محمدحسن ترابی، دکتر شهره صادقی

ویراستار: دکتر پریسا بخشنده

مدیریت هنری: آتلیه نویسه پارسی ■ چاپ و صحافی: روز

چاپ اول، تهران، بهار ۱۴۰۳ ■ ۳۰۰ نسخه

شابک: ۹۷۸-۶۲۲-۷۳۸۶-۷۵-۲

این کتاب ترجمه‌ای است از:

Sandrine Zufferey. (2020) *Introduction to Corpus Linguistics*. ISTE Ltd and John Wiley & Sons, Inc.

مرکز پخش:

پخش گسترش، تهران، خیابان دماوند، خیابان شهید باباییان، خیابان خورشید، پلاک ۳۱
تلفن: ۹۱۰۰۶۰۱۰

نماینده‌گی فروش:

کتابفروشی توس، تهران، خیابان انقلاب، نبش خیابان دانشگاه، پلاک ۱۷۸
تلفن: ۶۶۴۶۱۰۰۷



تلفن دفتر انتشارات: ۷۷۰۵۳۲۴۶ ایمیل: info@neveeseh.com

ارتباط با نشر نویسه پارسی: ۰۹۰۲۱۰۰۵۱۶۱

www.neveeseh.com

© تمامی حقوق این اثر برای نشر نویسه پارسی محفوظ است.

هرگونه استفاده تجاری از این اثر یا تکثیر آن، کلاً و جزئاً، به هر صورت (چاپ، فتوکپی، صوت، تصویر و انتشار الکترونیکی) بدون اجازه مکتوب ناشر ممنوع است.

زبان‌شناسی پیکره‌ای: ۲ ■ مسلسل انتشارات: ۱۵۸

فهرست مطالب

پیشگفتار.....	۱۱
۱. زبان‌شناسی پیکره‌ای را چگونه تعریف کنیم؟.....	۱۵
۱-۱. عناصر معرف.....	۱۵
۲-۱. تجربه‌گرایی در مقابل خردگرایی در زبان‌شناسی.....	۱۹
۳-۱. استدلال‌های چامسکی علیه تجربه‌گرایی در زبان‌شناسی.....	۲۰
۴-۱. زبان‌شناسی پیکره‌ای و ابزارهای رایانه‌ای.....	۲۶
۵-۱. روش‌های کمی در مقابل کیفی.....	۲۹
۶-۱. تفاوت‌های زبان‌شناسی پیکره‌ای و زبان‌شناسی آزمایشگاهی.....	۳۳
۷-۱. انواع پیکره.....	۳۷
۸-۱. نتیجه‌گیری.....	۴۱
۹-۱. پرسش‌ها و پاسخ‌ها.....	۴۲
۱-۹-۱. پرسش‌ها.....	۴۲
۲-۹-۱. پاسخ‌ها.....	۴۳
۱۰-۱. برای مطالعه بیشتر.....	۴۷
۲. چگونه پیکره‌ها را در زبان‌شناسی نظری به کار بگیریم.....	۴۹
۱-۲. آواشناسی و واج‌شناسی.....	۴۹
۲-۲. ساخت‌واژه.....	۵۴
۳-۲. نحو.....	۵۹

- ۴-۲. واژگان ۶۳
- ۵-۲. تحلیل گفتمان ۶۸
- ۶-۲. کاربردشناسی ۷۲
- ۷-۲. جامعه‌شناسی زبان ۷۶
- ۸-۲. زبان‌شناسی در زمانی ۷۹
- ۹-۲. نتیجه‌گیری ۸۳
- ۱۰-۲. پرسش‌ها و پاسخ‌ها ۸۳
- ۱-۱۰-۲. پرسش‌ها ۸۳
- ۲-۱۰-۲. پاسخ‌ها ۸۴
- ۱۱-۲. برای مطالعه بیشتر ۸۸
۳. چگونه از پیکره‌ها در زبان‌شناسی کاربردی استفاده کنیم ۸۹
- ۱-۳. فراگیری زبان ۸۹
- ۲-۳. اختلالات زبانی ۹۵
- ۳-۳. فراگیری زبان دوم ۱۰۰
- ۴-۳. آموزش زبان ۱۰۵
- ۵-۳. فرهنگ‌نگاری ۱۱۱
- ۶-۳. سبک‌شناسی ۱۱۷
- ۷-۳. زبان‌شناسی قضایی ۱۲۰
- ۸-۳. نتیجه‌گیری ۱۲۴
- ۹-۳. یادداشت‌ها ۱۲۵
- ۱۰-۳. پرسش‌ها و پاسخ‌ها ۱۲۵
- ۱-۱۰-۳. پرسش‌ها ۱۲۵
- ۲-۱۰-۳. پاسخ‌ها ۱۲۶
- ۱۱-۳. برای مطالعه بیشتر ۱۲۹
۴. چگونه از پیکره‌های چندزبانه استفاده کنیم؟ ۱۳۱
- ۱-۴. پیکره‌های مقایسه‌پذیر و پیکره‌های موازی ۱۳۲
- ۲-۴. جستجو برای مبنای مقایسه ۱۳۶

۳-۴	ترجمه به‌عنوان یک ژانر گفتمانی	۱۴۱
۴-۴	پیکره‌های چندزبانه و زبان‌شناسی مقابله‌ای	۱۴۴
۵-۴	پیکره‌های موازی و مطالعات ترجمه	۱۵۰
۶-۴	پیکره‌های موازی و فرهنگ‌های لغت دوزبانه	۱۵۵
۷-۴	نتیجه‌گیری	۱۵۷
۴-۸	یادداشت‌ها	۱۵۸
۹-۴	پرسش‌ها و پاسخ‌ها	۱۵۸
۱-۴-۹	پرسش‌ها	۱۵۸
۲-۴-۹	پاسخ‌ها	۱۵۹
۱۰-۴	برای مطالعه بیشتر	۱۶۲
۵	چگونه پیکره‌های زبان فرانسه را بیابیم و تحلیل کنیم	۱۶۳
۱-۵	انواع پیکره و دسترسی به آن‌ها	۱۶۳
۲-۵	پیکره‌های مرجع	۱۶۷
۳-۵	پیکره‌های نوشتاری زبان فرانسه	۱۶۹
۴-۵	پیکره‌های زبان فرانسه گفتاری	۱۷۳
۵-۵	پیکره‌های کودکان و زبان‌آموزان	۱۷۶
۶-۵	پیکره‌های چندزبانه شامل زبان فرانسه	۱۸۱
۷-۵	ابزارهای کاربرد پیکره	۱۸۷
۱-۷-۵	بسامدگیرها	۱۸۷
۲-۷-۵	بسامدگیر AntConc	۱۹۲
۳-۷-۵	بسامدگیر CLAN	۱۹۴
۸-۵	نتیجه‌گیری	۱۹۶
۹-۵	پرسش‌ها و پاسخ‌ها	۱۹۷
۱-۹-۵	پرسش‌ها	۱۹۷
۲-۹-۵	پاسخ‌ها	۱۹۸
۱۰-۵	برای مطالعه بیشتر	۲۰۵
۶	چگونه پیکره بسازیم	۲۰۷
۱-۶	پیش از تصمیم‌گیری برای ساخت پیکره	۲۰۷

- ۶-۲. تعیین اندازه و معرف‌بودگی داده‌ها ۲۱۰
- ۶-۳. انتخاب نمونه‌های زبانی ۲۱۶
- ۶-۴. آماه‌سازی و کدگذاری فایل‌های پیکره ۲۲۱
- ۶-۵. ضبط و ترانویسی داده‌های گفتاری ۲۲۸
- ۶-۶. مسائل اخلاقی و حقوقی ۲۳۱
- ۶-۷. نتیجه‌گیری ۲۳۵
- ۶-۸. پرسش‌ها و پاسخ‌ها ۲۳۵
- ۶-۸-۱. پرسش‌ها ۲۳۵
- ۶-۸-۲. پاسخ‌ها ۲۳۶
- ۶-۹. برای مطالعه بیشتر ۲۴۱
۷. چگونه پیکره‌ای را نشانه‌گذاری کنیم ۲۴۳
- ۷-۱. نشانه‌گذاری‌های پیکره ۲۴۳
- ۷-۲. انواع مختلف نشانه‌گذاری ۲۴۸
- ۷-۳. استانداردهای طراحی‌شده‌های نشانه‌گذاری ۲۵۴
- ۷-۴. مراحل فرایند نشانه‌گذاری ۲۵۵
- ۷-۵. ابزارهای نشانه‌گذاری ۲۶۳
- ۷-۶. ارزیابی کیفیت و پایایی نشانه‌گذاری ۲۶۸
- ۷-۷. اشتراک‌گذاری نشانه‌گذاری‌ها ۲۷۴
- ۷-۸. نتیجه‌گیری ۲۷۶
- ۷-۹. پرسش‌ها و پاسخ‌ها ۲۷۷
- ۷-۹-۱. پرسش‌ها ۲۷۷
- ۷-۹-۲. پاسخ‌ها ۲۷۷
- ۷-۱۰. منابع بیشتر ۲۹۳
۸. چگونه داده‌های پیکره‌ای را تحلیل کنیم ۲۹۵
- ۸-۱. آمارهای توصیفی برای داده‌های پیکره‌ای ۲۹۶
- ۸-۲. اندازه‌گیری غنای واژگانی پیکره ۳۰۲
- ۸-۳. اندازه‌گیری پراکندگی واژگانی در پیکره ۳۰۴

فهرست مطالب ♦ ۹

۳۰۸	۴-۸. مقدمات آمار استنباطی
۳۱۲	۵-۸. متغیرهای معمول در مطالعات پیکره‌ای
۳۱۶	۸-۶. اندازه‌گیری تفاوت‌های بین گروه‌ها
۳۲۹	۷-۸. نتیجه‌گیری
۳۳۰	۸-۸. پرسش‌ها و پاسخ‌ها
۳۳۰	۸-۸-۱. پرسش‌ها
۳۳۳	۸-۸-۲. پاسخ‌ها
۳۳۸	۹-۸. منابع برای مطالعه بیشتر
۳۳۹	جمع‌بندی: مراحل انجام یک مطالعه پیکره‌ای
۳۴۷	منابع
۳۶۵	واژه‌نامه
۳۹۷	نمایه

به یاد و خاطره دوست سفر کرده‌امان

دکتر سلیمه زمانی

پیشگفتار

از دهه ۱۹۹۰، زبان‌شناسی نقطه‌عطفی پایه‌ای در روش‌شناسی را هر روز بیش از پیش تجربه می‌کند. بعد از کارهای زبان‌شناس آمریکایی، نوآم چامسکی^۱، زبان‌شناسی که از اواسط قرن بیستم به یک رشته ضرورتاً خردگرا^۲ تبدیل شده بود، به تدریج (دوباره) به رویکردهای تجربی^۳ روی آورد که زبان‌شناسی پیکره‌ای^۴ و زبان‌شناسی آزمایشگاهی^۵ نمونه‌هایی از آن هستند. طی دهه گذشته، سرعت این انتقال بیشتر هم شده است، به گونه‌ای که اغلب مقالات زبان‌شناسی منتشرشده در مجلات بین‌المللی در حال حاضر از داده‌های تجربی استفاده می‌کنند. بنابراین، پیکره‌های زبانی به تدریج جای خود را به‌عنوان ابزارهای در دسترس زبان‌شناسان باز کرده‌اند و کاربرد آن‌ها به دیگر حوزه‌های زبان‌شناسی، از جمله آن‌هایی که همانند نحو به‌طور سنتی رویکردی خردگرایانه داشته‌اند، تسری یافته است. توسعه زبان‌شناسی پیکره‌ای به معرفی روش‌های جدیدی نیز برای گردآوری و تحلیل داده‌های زبانی انجامیده است؛ روش‌هایی که به لطف توسعه رایانه‌ها و آمدن اینترنت ممکن شدند. این مسیر جدید در زبان‌شناسی پیشرفت‌های چشمگیری را در زمینه کار با ابعاد چندگانه زبان انسان با همه پیچیدگی آن از دیدگاه علمی رقم زده است. این کتاب قصد دارد این

¹ N. Chomsky

² rationalist

³ empirical

⁴ corpus linguistics

⁵ experimental linguistics

ثروت را به خوانندگانی معرفی کند که به خواندن نوشتگان^۱ با محوریت زبان‌شناسی عادت آن‌چنانی ندارند.

در دوران ما، توانایی تحلیل کتی داده‌های پیکره‌ای تبدیل به بخش جدایی‌ناپذیر جعبه ابزار زبان‌شناسان شده است. با وجود این، کاربرد این داده‌ها بر اصول دقیق نظری و روش‌شناختی مبتنی است که به شناخت کاملی نیاز دارند. این نقطه عطف در زبان‌شناسی بدین معناست که نسل جدید دانشجویان به استفاده از این روش‌ها نیاز دارند که به آنان کمک می‌کند مسائل زیربنای کاربرد آن‌ها در نوشتگان علمی را بهتر بشناسند، نتایج به‌دست‌آمده را با دیدگاه انتقادی ارزیابی کنند، و از آن‌ها در کارهای آکادمیکشان بهره ببرند. هدف از کتاب ما ارائه پشتیبانی آموزشی به این دانشجویان، و به‌طورکلی، به همه کسانی است که می‌خواهند استفاده از پیکره در زبان‌شناسی را یاد بگیرند.

برای خواندن این کتاب به داشتن پیشینه زبان‌شناسی خاصی نیاز ندارید، مگر اطلاعات پایه‌ای زبان‌شناختی، به‌علاوه تسلط حداقلی بر ابزارهای رایانه‌ای، مانند نرم‌افزارهایی شبیه به اکسل. این کتاب به گونه‌ای طراحی شده که منبعی مقدماتی باشد برای تدریس زبان‌شناسی پیکره‌ای در دانشگاه و همچنین ابزاری باشد برای دانشجویانی که دوست دارند استفاده از پیکره‌ها را یاد بگیرند. دانشجویان، به لطف پرسش‌های مروری ارائه‌شده در پایان فصل‌ها و پاسخ‌های مفصل آن‌ها می‌توانند مستقلاً کار کنند.

از آنجاکه این کتاب یک کتاب مقدماتی است، فقط به برخی مسائل پرداخته است و به همه مسائل مطرح در استفاده از پیکره‌ها در حوزه‌های زبان‌شناختی مختلف نمی‌پردازد. کتاب روش‌های تحلیل پیشرفته‌ای که به سطح بالایی از مهارت‌های رایانه‌ای و آماری برای تحلیل داده‌ها نیاز دارند را پوشش نمی‌دهد. با وجود این، منابع بیشتری برای مطالعه در پایان هر فصل آمده است که به افرادی که تمایل دارند در حوزه‌ای گامی فراتر بردارند کمک می‌کنند.

¹ literature

درنهایت، این کتاب تأکید ویژه‌ای بر زبان فرانسه به‌عنوان موضوع مطالعه دارد. زبان‌شناسی پیکره‌ای خود را به جهان انگلیسی‌زبان تحمیل کرده و رقیبی هم ندارد، و تعداد بسیار زیادی از پژوهشگران فرانسوی‌زبان هم از این روش‌ها استفاده می‌کنند، اما هنوز تدریس زبان‌شناسی پیکره‌ای در فرانسه امری حاشیه‌ایست. بنابراین، این کتاب سعی دارد بر قدرت و غنای مطالعات پیکره‌ای در زبان فرانسه تأکید کند و همچنین مهم‌ترین منابعی را معرفی کند که به این زبان اختصاص یافته‌اند، با امید به اینکه به ارتقای این رشته برای مطالعه زبان فرانسه کمک کند.

سندرین زوفری

ژوئن ۲۰۲۰

۱. زبان‌شناسی پیکره‌ای را چگونه تعریف کنیم؟

هدف این فصل این است که عناصر اصلی معرف زبان‌شناسی پیکره‌ای را مشخص کند تا بدانیم این رشته چه چیزهایی را شامل می‌شود. هدف دیگر این فصل این است که پایه‌های نظری و روش‌شناختی این رشته را بیان کند. ما به خصوص تفاوت بین روش‌های تجربی و خردگرا در زبان‌شناسی، نقش مهم علوم رایانه در زبان‌شناسی پیکره‌ای، و تفاوت بین پژوهش‌های کمی و کیفی، و همچنین تفاوت بین زبان‌شناسی پیکره‌ای و زبان‌شناسی آزمایشگاهی را مشخص کنیم. در پایان نیز بر انواع مختلف پیکره‌ها مروری کوتاه خواهیم داشت. در فصل‌های بعدی، این مقدمه به ما کمک می‌کند که به پرسش‌های پژوهشی بپردازیم که می‌توان با استفاده از زبان‌شناسی پیکره‌ای به آن‌ها پاسخ داد.

۱-۱. عناصر معرف

اصطلاح *corpus* (=پیکره) از ریشه‌ای لاتین به معنی «بدن» می‌آید. یک پیکره متنی^۱ به معنای واقعی بدنه‌ای از مجموعه‌ای از متون است. برای مثال، می‌توان مجموعه‌ای از مقالات روزنامه را گرد آورد و از آن‌ها پیکره‌ای ساخت و با استفاده از آن ویژگی‌های ژانر^۲ روزنامه‌نگاری^۳ را بررسی کرد. در حوزه آموزش زبان، می‌توان

¹ text corpus

² genre

³ journalist

متونی را گردآوری کرد که زبان‌آموزانی با سطوح مهارتی مختلف نوشته‌اند، و از این نوشته‌ها پیکره‌ای ساخت برای بررسی اشتباهات معمولی که زبان‌آموزان در سطوح مختلف یادگیری زبان تولید می‌کنند. روش‌شناسی‌ای که به جای دانش خود فرد از زبان، از داده‌های جهان خارج استفاده کند، روش‌شناسی تجربی^۱ نامیده می‌شود. زبان‌شناسی پیکره‌ای را می‌توان رشته‌ای تجربی به معنای واقعی دانست، زیرا هدف آن نتیجه‌گیری بر پایه تحلیل داده‌های برونی^۲ است، و نه بر پایه دانش زبانی محققان. بنابراین کار با زبان‌شناسی پیکره‌ای به معنای بودن در تماس با داده‌های زبانی به صورت متن و همچنین به صورت فایل صوتی، ویدیویی، یا هر نمونه دیگری است که محتوی زبان باشد. بیشتر مواقع این نمونه‌ها به فرمت رایانه‌ای گردآوری می‌شوند که این امکان را فراهم می‌کند که بتوانیم آن‌ها را به طور مؤثرتری نسبت به زمانی که روی کاغذ هستند بررسی کنیم. برای مثال، فکر کنید می‌خواهیم بدانیم که فلوربر^۳ در رمان مادام بواری^۴ چند بار و در چه متن‌هایی حس عشق را برمی‌انگیزد. اگر نسخه کاغذی کتاب را داشته باشیم، یافتن این متن‌ها کار طولانی‌مدت و خسته‌کننده‌ای خواهد بود که نیازمند این است که کل متن را مرور کنیم. اما اگر نسخه رایانه‌ای آن را در اختیار داشته باشیم، کار بسیار آسان‌تر خواهد شد. کافی است صرفاً صورت‌های مختلف واژه‌های *love*، *in love*، یا فعل *to love* را با تابع جستجوی واژه‌پرداز جستجو کنیم تا محل به کار رفتن آن‌ها را به ما نشان دهد و ما آن‌ها را بشماریم. برای بسیاری از پرسش‌هایی که زبان‌شناسی پیکره‌ای قرار است به آن‌ها پاسخ دهد، نمی‌توان یک دادگان^۵ کاغذی را کاملاً جستجو کرد، و به همین دلیل است که داشتن پیکره‌های رایانه‌ای ضروری است.

¹ empirical methodology

² external

³ G. Flaubert

⁴ *Madame Bovary*

⁵ database

هنگامی که می‌خواهیم چندین کاربرد یک پدیده زبانی خاص را مشاهده و ویژگی‌های خاص آن را بیابیم، مشکل ردیابی و شمارش دستی پیچیده‌تر هم می‌شود، چون زبان‌شناسی پیکره‌ای بر پایه میزان بسیار زیادی از داده‌هایی قرار دارد که فقط از یک کتاب گرفته نشده‌اند. برای مثال، فرض کنید می‌خواهیم بدانیم آیا فلور بر کارهایش درباره عشق صحبت می‌کند یا نه. در اینجا، صرف تمرکز بر مادام‌بواری باعث سوگیری می‌شود، زیرا این رمان معرف همه کارهای او نیست. بنابراین، برای اینکه بتوانیم به این پرسش پاسخ دهیم، باید همه رمان‌هایش را بررسی کنیم و این کار جستجوی دستی را بسیار سخت‌تر می‌کند. فرض کنیم حالا می‌خواهیم ببینیم آیا همه نویسندگان قرن نوزدهم فرانسوی به اندازه فلور به مسئله عشق پرداخته‌اند یا نه. در این صورت، غیرممکن خواهد بود که اصطلاحات مربوط به عشق را در همه رمان‌های نویسندگان فرانسوی قرن نوزدهم جستجو کنیم. برای اجتناب از این مشکل، باید نمونه‌ای^۱ از متون معرف آثار این دوره گردآوری کنیم. در فصل ۶ که به اصول روش‌شناختی زیربنای گردآوری پیکره می‌پردازد، درباره این موضوع صحبت می‌کنیم. اما در اینجا نکته مهم این است که به خاطر داشته باشیم که زبان‌شناسی پیکره‌ای معمولاً از روش‌شناسی کمی (بنگرید به بخش ۱-۲) استفاده می‌کند تا بتواند یافته‌هایش از یک نمونه زبانی را به کل زبان یا یک سیاق^۲ زبانی خاص تعمیم دهد.

همان‌گونه که در فصل‌های بعدی خواهیم دید، از زبان‌شناسی پیکره‌ای می‌توان در همه حوزه‌های زبان‌شناسی استفاده کرد. برای مثال، کاربرد آن در فرهنگ‌نگاری ضروری است، زیرا این امکان را به وجود می‌آورد تا سیاه کامل واژگان^۳ یک زبان را به دست آوریم. همچنین این امکان را به ما می‌دهد که از کاربرد در انواع منابع مختلف (ادبی، خبرنگاری، و غیره) نمونه‌هایی بیابیم و عباراتی را مشخص کنیم که

¹ sample

² register

³ lexicon

یک واژه خاص با بسامد بالایی در آن‌ها به کار می‌رود. به عبارت دیگر، این امکان را به ما می‌دهد که عناصر عبارت‌پردازی^۱ بسیار سودمندی را برای فرهنگ‌های لغت مشخص کنیم. برای مثال، خوب است که بدانیم واژه «knowledge» به چه معناست، اما به همان اندازه هم مهم است که این واژه اغلب در عباراتی به کار می‌رود مانند «acquire knowledge» یا «(having good knowledge of)» یا

زبان‌شناسی پیکره‌ای به خصوص روش بسیار خوبی است برای مشخص کردن بافت‌های پربسامدی که یک واژه یا اصطلاح در آن‌ها به کار می‌رود. اما زبان‌شناسی پیکره‌ای برای انجام پژوهش در حوزه‌های پایه زبان‌شناسی مانند مطالعه نحو نیز کاربرد دارد، زیرا امکان شناسایی انواع ساخت‌های نحوی مورد استفاده در زبان‌های مختلف را به ما می‌دهد. برای مثال، با انجام یک پژوهش پیکره‌ای می‌توان تعیین کرد که جهت مجهول^۲ در کدام یک از ژانرهای متنی بیشترین کاربرد را دارد. درنهایت، به لطف وجود پیکره‌هایی از داده‌های شفاهی، زبان‌شناسی پیکره‌ای همچنین این امکان را به وجود می‌آورد تا به پرسش‌های مرتبط با واج‌شناسی و جامعه‌شناسی زبان پاسخ بدهیم. برای مثال، می‌توان با زبان‌شناسی پیکره‌ای پراکندگی جغرافیایی برخی از گرایش‌های تلفظی را مشخص کرد، مانند تمایز بین صورت /a/ کوتاه در واژه فرانسوی «patte» (پنجه) از صورت /a/ بلند در «pâte» (شیرینی). پاسخ به این گونه پرسش‌ها نیازمند استفاده از انواع مختلفی از پیکره‌ها و همچنین در دست داشتن داده از محتوای آن‌هاست. برای مثال، برای تعیین پراکندگی جغرافیایی یک گرایش تلفظی خاص، باید بدانیم هر گویشوری که داده از وی در پیکره داریم از کجا آمده است. به این نوع اطلاعات فراداده^۳ پیکره می‌گویند. ما انواع اصلی پیکره‌های موجود را در پایان این فصل می‌آوریم و درباره فراداده در فصل ۶ صحبت می‌کنیم.

¹ phraseology

² passive voice

³ metadata

در مجموع، در این بخش ما زبان‌شناسی پیکره‌ای را به‌عنوان یک رشته تجربی تعریف کردیم که نمونه‌های کمی زبان را که به فرمت رایانه‌ای گردآوری شده‌اند مشاهده و تحلیل می‌کند. در بخش‌های بعدی درباره نکات اصلی این تعریف که به‌صورت تیره مشخص شده است بحث می‌کنیم تا از پایگاه نظری و روش‌شناختی زبان‌شناسی پیکره‌ای شناخت بهتری پیدا کنیم.

۱-۲. تجربه‌گرایی در مقابل خردگرایی در زبان‌شناسی

زبان‌شناسی پیکره‌ای رشته‌ای تجربی است، بدین معنا که از داده‌های تولیدشده توسط گویشوران برای مطالعه زبان استفاده می‌کند. این روش‌شناسی در مقابل روش خردگرا قرار دارد که با تکیه بر دانش زبانی خود فرد و نه داده‌های برونی عمل می‌کند. برای مثال، برای تعیین اینکه عبارت «When do you think he will prepare which cake?» از نظر دستوری درست است یا نه، روش‌شناسی تجربی پیکره‌های بزرگ زبانی را بررسی خواهد کرد تا ببیند آیا انگلیسی‌زبانان این ساختار نحوی را به کار می‌برند یا نه؟

اگر هیچگاه، یا تقریباً هیچگاه، جملاتی با این ساختار نحوی در پیکره یافت نشوند، زبان‌شناسان نتیجه می‌گیرند که این جمله به‌ندرت در زبان انگلیسی به کار می‌رود. در مقابل، روش‌شناسی خردگرا به این پرسش با تکیه بر شم زبانی زبان‌شناسان پاسخ می‌دهد. در این مورد خاص، آن‌ها فکر می‌کنند که آیا می‌توانند چنین جمله‌ای بسازند یا نه، و آیا این جمله بر پایه دانش آنان از زبان، که بر اساس آن قضاوت دستوری^۲ می‌کنند، درست به نظر می‌رسد یا نادرست. قضاوت‌های دستوری معمولاً سه دسته می‌شوند: درست، نادرست، یا نشان‌دار، که یعنی آن جمله ممکن، اما غیرعادی به نظر می‌رسد.

¹ intuition

² grammaticality judgment

این مثال تفاوت اساسی بین روش‌شناسی تجربی و خردگرا را نشان می‌دهد. درحالی‌که روش‌شناسی خردگرا به صورت‌بندی قضاوت مقوله‌ای^۱ می‌انجامد، روش‌شناسی تجربی پاسخ ظریف‌تری به این پرسش می‌دهد، زیرا داده‌های پیکره معمولاً بسامدی دقیق به دست می‌دهند، و نه پاسخی بر مبنای وجود داشتن یا نداشتن. این یکی از دلایلی است که بسیاری از زبان‌شناسان امروزه اعتقاد دارند روش‌شناسی تجربی نسبت به روش کاملاً خردگرا به مطالعه زبان تناسب بهتری با رویکرد علمی دارد (به معنی مواجهه با واقعیت).

اما انتخاب بین کاربرد روش‌های تجربی و خردگرا به حوزه زبان‌شناسی محدود نمی‌شود. برخی رشته‌های علمی، مانند فیزیک، شیمی، جامعه‌شناسی، و تاریخ نیز ضرورتاً رشته‌های تجربی هستند. در واقع، هم فیزیک‌دانان و هم تاریخ‌شناسان پایه نگرش‌های خود را بر داده‌های برونی می‌گذارند، که از جهان خارج گردآوری می‌کنند، تا نظریه^۲ بسازند، آن را بیازمایند، و از آن نتیجه‌گیری کنند. از طرف دیگر، رشته‌های دیگر، مانند ریاضیات یا فلسفه به طور سنتی بر پایه رویکرد خردگرا قرار دارند، زیرا ریاضی‌دانان و فلاسفه از استدلال‌های خودشان برای طرح نظریه و نتیجه‌گیری استفاده می‌کنند و نه از گردآوری و مشاهده داده‌های برونی. فیلسوف‌ها معمولاً از آزمایش تفکر^۳ کمک می‌گیرند، اما این‌ها به معنای تجربی آزمایش نیستند، زیرا بر پایه توانایی‌های تفکر پژوهشگر قرار دارند.

۱-۳. استدلال‌های چامسکی علیه تجربه‌گرایی در زبان‌شناسی

هرچند زبان‌شناسی پیکره‌ای در طول ۲۰ سال گذشته رشد چشمگیری داشته است، اما تجربه‌گرایی در زبان‌شناسی چیز جدیدی نیست. زبان‌شناسان مدت‌های مدیدی از داده‌های مشاهده‌ای^۴ استفاده می‌کردند. برای مثال، در قرن ۱۹، زبان‌های

¹ categorical judgment

² theory

³ thought experiment

⁴ observational data

هندواروپایی را با هم مقایسه می‌کردند تا منشأ مشترک آن‌ها را بازسازی کنند. این تحقیقات بر مبنای داده‌های موجود از زبان‌های گفتاری اروپایی مانند آلمانی و فرانسه و انگلیسی انجام می‌شد. همچنین، در نیمه اول قرن ۲۰ در ایالات متحده آمریکا، رویکرد موسوم به رویکرد توزیعی^۱ به نحو بر مطالعه جمله‌سازی در ساخت‌های نحوی آن گونه که در پیکره‌های متنی می‌آمدند متمرکز بود و این رویکرد سعی می‌کرد کارکرد کلی زبان را از آنجا استنباط کند. در اواخر دهه ۱۹۵۰، بر مبنای کارهای زبان‌شناس آمریکایی، نوآم چامسکی، کاربرد پیکره در زبان‌شناسی در حوزه‌هایی مانند نحو تقریباً متوقف شد. در واقع چامسکی مدافع رویکرد روش‌شناختی خردگرایی محض در زبان‌شناسی بود، و با هر گونه استفاده از داده‌های برونی به شدت مخالفت می‌کرد. چامسکی انتقادات پرشماری را بر کاربرد داده‌های برونی در زبان‌شناسی مطرح کرده است. ما در اینجا مروری کوتاه بر آن‌ها خواهیم داشت تا نشان دهیم که چگونه بیشتر آن‌ها علت وجودی خود را در بافت تحقیقات جدید از دست داده‌اند.

اولین انتقاد چامسکی به کاربرد پیکره، که البته اساسی‌ترین انتقاد او نیز هست، این است که پیکره‌ها نمونه‌های زبانی را در برمی‌گیرند که گویشوران تولید کرده‌اند. به اعتقاد او، زبان‌شناسی نباید بر کنش^۲ زبانی گویشوران متمرکز باشد، بلکه باید بر توانش^۳ آن‌ها در زبان مادری‌شان متمرکز باشد، چیزی که وی آن را زبان درونی^۴ می‌نامد. مسئله اینجاست که وقتی افراد صحبت می‌کنند، چیزی که تولید می‌کنند (یعنی کنش آن‌ها) لزوماً منعکس‌کننده آن چیزی نیست که آنان درباره زبانشان می‌دانند (یعنی توانش آن‌ها). برای مثال، گویشوران تحت تأثیر استرس یا خستگی گاه مرتکب خطا یا اشتباهات زبانی می‌شوند. گاه به‌گاه تقریباً برای هر کسی پیش می‌آید که فعلی بی‌قاعده را به‌درستی صرف نکند و مثلاً به جای «*he ate*» بگوید

¹ distributionist approach

² performance

³ competence

⁴ internal language

«*he eated*»). اما اگر صدای کسی که این خطا را مرتکب شده است ضبط کنیم و بعد پیرسیم که فکر می‌کند جمله‌ای که گفته است درست بوده است یا نه، تقریباً می‌توانیم مطمئن باشیم که اشتباهش را درمی‌یابد و می‌تواند صورت درست، یعنی «*he ate*» را تولید کند. در مقابل، یک گویشور می‌تواند واژه «*serendipity*» را بعد از شنیدن از زبان یک نفر تکرار کند، بدون اینکه معنی آن را بداند. این نمونه‌ها این واقعیت را نشان می‌دهند که واژه‌هایی که گویشوران «ادا می‌کنند» همیشه منعکس‌کننده درست توانش زبانی آنان نیستند. و به این صورت، به اعتقاد چامسکی، واقعیت مطالعه پیکره‌ها زبان‌شناسان را در مسیر اشتباه می‌اندازد، زیرا باعث می‌شود زبان‌شناس زبان را از منظر «تولید» بررسی کند، که انعکاس سوگیرانه‌ای از قواعد زبان است.

به اعتقاد چامسکی، مشکل دیگر زبان‌شناسی پیکره‌ای از این واقعیت ناشی می‌شود که پیکره‌ها معرف زبان به صورت یک کل نیستند. وی این مشکل را به صورت افراطی و با در نظر گرفتن بیمار زبان‌پریشی که در پیکره ضبط شده باشد بیان می‌کند. زبان‌شناسانی که این پیکره را تحلیل می‌کنند درباره زبان مورد نظر نتایج یکسره غلط می‌گیرند، زیرا این فرد نماینده توانش زبانی یک گویشور معمول نیست. به علاوه، حتی اگر قرار نباشد هیچ گویشور غیرعادی‌ای را وارد پیکره کنیم، در مقایسه با همه تولیدات شفاهی و نوشتاری یک زبان، یک پیکره هیچگاه چیزی بیشتر از یک نمونه کوچک زبانی را نشان نمی‌دهد. به همین دلیل است که نمی‌توان نتیجه گرفت که واژه‌ای در زبانی وجود ندارد صرفاً به این دلیل که در پیکره دیده نمی‌شود. شاید اصلاً در آن بافت خاص تلفظ نمی‌شود، اما در دیگر سیاق‌های زبانی وجود داشته باشد یا گویشورانی آن را به کار برده باشند که در پیکره وارد نشده‌اند. این مشکل به خصوص برای حوزه‌های نادر زبانی، مانند واژه‌ها یا ساختارهای کم‌بسامد، وخیم‌تر هم می‌شود.

این محدودیت به سومین انتقاد چامسکی از پیکره منجر شده است، یعنی این واقعیت که یک پیکره هیچگاه نمی‌تواند کل یک زبان را دربرگیرد و بنابراین مشکلاتی که در بالا به آن‌ها اشاره شد هیچگاه حل نمی‌شوند. به اعتقاد او، این مشکل جدی‌تر هم می‌شود به این دلیل که حتی اگر پیکره‌ای بسیار بزرگ باشد و بخش معرفی از زبان را هم دربرگیرد، زبان‌شناس نمی‌تواند آن را به‌طور کامل تحلیل کند، با توجه به اینکه نمی‌توان محتوای میلیاردها جمله را به‌صورت دستی تحلیل کرد. دو انتقاد آخر چامسکی به دلیل پیشرفت‌های علوم رایانه دیگر منسوخ هستند. در واقع، در طی ۲۰ سال گذشته اندازه پیکره‌ها به‌صورت تساعدی افزایش یافته است، و ابزارهای تحلیل پیکره هم پیشرفت چشمگیری کرده‌اند. بنابراین این امکان به وجود آمده است که مقادیر بسیار بزرگ داده را تحلیل کنیم، که آینه بسیار دقیق‌تری از زبان به دست می‌دهد نسبت به زمانی که چامسکی انتقادات خود را مطرح کرد. ما در بخش ۱-۴ که به پیوند بین علوم رایانه و زبان‌شناسی پیکره‌ای اختصاص دارد، به این بحث باز می‌گردیم. در کنار این پیشرفت‌های فناوری، پیشرفت‌های نظری و روش‌شناختی نیز تا اندازه زیادی این امکان را فراهم کرده‌اند تا دیگر سوگیری‌هایی را که چامسکی مطرح کرده، حذف یا کنترل کنیم. برای مثال، روش درست برای ایجاد پیکره این است که نوع زبانی را که در آن آمده است مستند کنیم. این کار کمک می‌کند تا از تحلیل اشتباه یک تک‌بیمار زبان‌پریش، به قول چامسکی، اجتناب کنیم. اما این درست است که یک پیکره فقط می‌تواند چیزی را نشان دهد که حاوی آن است، و بنابراین نبود یک واژه یا یک ساختار در پیکره نمی‌تواند شاهد قطعی برای نبود آن در زبان باشد. بنابراین، برای برخی پرسش‌های پژوهش مرتبط با پدیده‌های نادر یا پدیده‌هایی که به سختی در پیکره قابل مشاهده‌اند، توصیه می‌شود که پژوهش را با یک روش تجربی دیگر تکمیل کنیم، یعنی با یک روش آزمایشگاهی. همان‌گونه که در ادامه این فصل خواهیم دید، این روش کاربرد روش‌شناسی کمی را با زبان‌شناسی پیکره‌ای ترکیب می‌کند.