



مرکز منطقه‌ای اطلاع‌رسانی علوم و فناوری

گزارش نهایی طرح پژوهشی:

تهیه‌ی سامانه و پیکره‌ی معنایی افعال زبان فارسی بر پایه‌ی مؤلفه‌های معنایی افعال

محمدهادی فلاحی

استادیار گروه پژوهشی زبان‌شناسی رایانه‌ای

تاریخ آغاز: ۱۳۹۸/۲/۱۵

تاریخ پایان: ۱۳۹۹/۶/۳

فهرست مطالب

صفحه	عنوان
۳	چکیده
۴	۱- فصل اول: مقدمه
۶	۲- فصل دوم: مبانی نظری و پیشینه‌ی پژوهش
۱۲	۳- فصل سوم: روش‌شناسی پژوهش
۱۸	۴- فصل چهارم: معرفی سامانه‌ی معنایی
۲۲	فهرست منابع

چکیده

زمینه: بررسی افعال و طبقه‌بندی آنها یکی از مهم‌ترین بخش‌های دستور زبان محسوب می‌شود. تاکنون افعال بر اساس ساختمان، معنی، وجه و ... طبقه‌بندی شده‌اند. این پژوهش افعال را بر مبنای یکی از معیارهای معنایی، مؤلفه‌های معنایی، با استفاده از پیکره‌های زبانی همشهری، پرسیکا و بیجن‌خان بررسی و طبقه‌بندی کرده و فهرست و سامانه‌ی معنایی افعال زبان فارسی را ارائه داده است.

روش‌شناسی: در تهیه‌ی فهرست معنایی افعال و پیکره‌ی جملات، روند ذیل طی گردید:

در گام نخست کلیه‌ی افعال متداول زبان فارسی استخراج شد و مؤلفه‌های معنایی آنها تعیین گردید. چگونگی انتخاب این افعال بدین صورت بود که در گام نخست این افعال در سه پیکره‌ی همشهری، پرسیکا و بیجن‌خان جست‌وجو شدند. قاعدتا بخشی از این افعال در این سه پیکره وجود نداشتند. با توجه به عدم وجود این افعال در این پیکره‌ها و همچنین نظر مجری و همکاران، بخشی از این افعال به عنوان افعال کاملاً مهجور، حذف شدند. در گام دوم جمع‌بندی و بررسی نهایی افعال و مؤلفه‌های معنایی موضوعات صورت گرفت و فهرست افعال به همراه مؤلفه‌های آنها در فایل اکسل قرار گرفت. در گام سوم استخراج جملات از پیکره‌های همشهری، پرسیکا و بیجن‌خان صورت گرفت. بدین منظور نخست جملاتی که حاوی افعال ساده، پیشوندی، مرکب و پیشوندی مرکب بودند (افعال موجود در سامانه‌ی ظرفیت نحوی افعال) استخراج شدند. قاعدتا تعدادی از افعال در این پیکره‌ها وجود نداشتند ولی افعال متداول فارسی هستند. به منظور دستیابی به مؤلفه‌های معنایی موضوعات این افعال و محدودیت‌های گزینشی از پنج سخنگوی بومی فارسی زبان به منظور ارائه‌ی این جملات کمک گرفته شده است. در پایان نیز پالایش جملات صورت گرفت. بدین منظور نخست کلیه‌ی جملات استخراج شده، بررسی شد و مثال‌های ناقص و نامرتب حذف گردید و برای هر فعل در حدود ۲۰ مثال باقی ماند. سپس با بررسی دوباره‌ی مثال‌ها برای هر فعل حداکثر ۵ مثال باقی ماند و در نهایت این ۵ جمله پالایش شدند و سازه‌هایی که نقش موضوعی نداشتند مانند قیدها، صفات، مضاف‌الیه‌ها، متمم‌ها و جملات درونه‌ای، حذف شدند. در مجموع ۴۱۹۹ فعل در سامانه فهرست شده است که ۲۴۵ مورد، فعل ساده و ۳۹۵۴ فعل، مرکب، پیشوندی و پیشوندی-مرکب هستند. ۱۵۸۴ فعل دارای یک موضوع، ۲۴۴۳ فعل، دو موضوعی و ۱۷۳ فعل، سه موضوع دارند. ۳۵۴۲ فعل در پیکره‌ها موجود بودند و ۶۵۷ فعل در پیکره‌ها وجود نداشتند. افعال حساس به جنسیت نیز ۲۵ فعل شناخته شد. پیکره‌ی به دست آمده نیز شامل ۱۳۷۵۰ جمله است که برای هر فعل، ۳/۳ جمله وجود دارد.

کلیدواژه‌ها: فعل، مؤلفه‌های معنایی، محدودیت‌های گزینشی، نقش‌های موضوعی

فصل اول: مقدمه

۱-۱ مقدمه

مقوله‌ی دستوری فعل در دستور زبان به دلیل باز بودن فهرست واژه‌های متعلق به آن، گوناگونی ساختاری اعضای این مقوله دستوری و ویژگی‌های معنایی که هر ساختار با خود به همراه دارد، از بحث برانگیزترین مباحث دستوری بوده است (کرد زعفرانلو، عبدالکریمی و گلفام، ۱۳۹۰: ۲). افعال از منظرهای مختلف طبقه‌بندی شده‌اند: نحوی، صرفی، معنایی، میزان کاربرد و انواع فعل از لحاظ نحوی عبارتند از: فعل واژگانی، فعل رابط (ربطی)، فعل کمکی (معین)، فعل وجهی و عنصر فعلی (در فعل مرکب). در پژوهش حاضر که ادامه‌ی دو پژوهش پیشین است، کلیه‌ی افعال متداول زبان فارسی بر مبنای مؤلفه‌های معنایی موضوعاتشان فهرست شده‌اند و برای هر فعل تا پنج جمله برای نشان دادن موضوعات از پیکره‌های همشهری، پرسیکا و بیجن‌خان استخراج شده و با انجام اصلاحاتی، فهرست شده‌اند. دست‌آورد این پژوهش نیز سامانه و پیکره‌ای نرم‌افزاری است که کلیه‌ی افعال به همراه مؤلفه‌های آنها و همچنین بین یک تا پنج جمله برای هر فعل در آن گردآوری شده است که با درج هر فعل می‌توان به مؤلفه‌های آن و همچنین نمونه جملاتی برای آن دست یافت.

۱-۲ بیان مساله

به‌طور کلی افعال بر اساس ساختمان، معنی، وجه و ... طبقه‌بندی شده‌اند. یکی از دغدغه‌های همیشگی زبان‌شناسان و دستورنویسان این بوده است که افعال را بر حسب ویژگی‌های نحوی طبقه‌بندی کنند. اما این که کدام ویژگی نحوی معیار طبقه‌بندی قرار گیرد، محل اختلاف بوده است. به نظر می‌رسد تقسیم‌بندی افعال به دو دسته‌ی لازم و متعدی قدیمی‌ترین رویکرد در این زمینه باشد. این طبقه‌بندی هرچند سودمند است اما در تبیین بسیاری از پیچیدگی‌های نحوی نا کاراست. از جمله شیوه‌های نوین و کارا در بررسی افعال و ساختمان جمله، توجه به ظرفیت افعال و به عبارتی تعداد متمم‌های اجباری و اختیاری افعال است. نظریه‌ی دستور وابستگی که یکی از نظریه‌های صورتگراست به مساله‌ی ظرفیت فعل، اسم و صفت می‌پردازد. پژوهش‌های متعددی به بحث ظرفیت افعال فارسی پرداخته است که مهم‌ترین آنها طبیب زاده (۱۳۸۵) است. همچنین نرم‌افزار فرهنگ ظرفیت نحوی افعال فارسی نیز توسط رسولی، مولودی، کوهستانی و مینایی بیدگلی (۱۳۹۱، ۲۰۱۱) تهیه شده است که منبع استخراج افعال این پژوهش بوده است. افعال برای انتخاب موضوع‌های خود محدودیت‌هایی را اعمال می‌کنند و به عبارتی هر کلمه‌ای با هر ویژگی معنایی را به عنوان فاعل یا مفعول نمی‌پذیرند (کروگر، ۲۰۰۵: ۱۷۳). در واقع محدودیت‌های گزینشی محدودیت‌هایی معنایی هستند که بر مبنای آنها هر واژه‌ای نمی‌تواند با واژه‌ای دیگر همراه شود و واحد زبانی بزرگ‌تر را تشکیل دهد هر چند تشکیل این واحد زبانی بزرگ‌تر از نظر نحوی بدون اشکال باشد.

در این پژوهش افعال بر اساس مؤلفه‌های معنایی و محدودیت‌های گزینشی طبقه‌بندی شده‌اند. با توجه به مشخص نبودن تعداد و چگونگی مؤلفه‌های معنایی موضوع‌ها، در این پژوهش افعال با توجه به پنج مؤلفه‌ی [+ و - عینی]، [+ و - جاندار]، [+ و - جانور]، [+ و - انسان] و [+ و - مذکر] مورد بررسی قرار گرفته و طبقه‌بندی شده‌اند و در سامانه‌ی معنایی و پیکره‌ی معنایی لحاظ خواهند شد.

پیکره‌ی زبانی عبارت است از مجموعه‌ای از متن‌های نوشتاری یا گفتاری آوانویسی شده که می‌توان در توصیف و تحلیل زبان از آن بهره گرفت (کندی، ۱۹۹۸). منظور از پیکره، مجموعه‌ای نسبتاً بزرگ از متون الکترونیک است که به صورت حساب شده برچسب‌گذاری و دسته‌بندی شده‌اند و امکان بررسی‌های مختلف را به کاربر می‌دهند (مکنزی و ویلسون، ۲۰۰۱). زبان‌شناسی پیکره‌ای شامل سه حوزه‌ی پژوهشی می‌شود که عبارتند از:

۱- استخراج خودکار داده‌های زبانی از پیکره

۲- پردازش آماری داده‌ها

۳- ارزیابی و تفسیر داده‌ها (تئوری، ۱۹۹۱)

پیکره‌ای که در این پژوهش تهیه خواهد شد، حوزه‌ی نخست را پوشش می‌دهد و در ادامه‌ی طبقه‌بندی افعال زبان فارسی بر پایه‌ی مؤلفه‌های معنایی در دو فاز گذشته‌ی طرح، در صدد تهیه‌ی سامانه‌ی معنایی این افعال است.

۳-۱ اهمیت پژوهش

بدون شک پرداختن به مسایل و مشکلات زبان فارسی به طور اعم و تسهیل حضور آن در فضای مجازی به طور اخص اهمیتی انکارناپذیر دارد. زبان فارسی از جمله زبان‌های زنده و مطرح دنیاست که پرداختن به ظرایف و جوانب آن برای هر پژوهشگر زبان فارسی، لازم و ضروری است. تاکنون سامانه‌ی ظرفیت نحوی افعال زبان فارسی (رسولی، مولودی، کوهستانی و مینایی بیدگلی، ۱۳۹۱) و سامانه‌ی معنایی افعال زبان فارسی بر مبنای نقش‌های تتایی (میرزایی، مولودی، ۱۳۹۳) تهیه شده است ولی سامانه معنایی افعال زبان فارسی بر مبنای مؤلفه‌های معنایی تاکنون تهیه نشده و این پژوهش نخستین کار در این زمینه است. همچنین چندین پیکره‌ی نحوی زبان فارسی (از جمله بیجن‌خان، شیخ زاده، بحرانی و قیومی، ۲۰۱۱) نیز ارائه شده است. در حال حاضر پیکره‌ها امکان بهره‌گیری سریع از داده‌های متنی را با کم‌ترین هزینه ممکن می‌سازند و وجود آن‌ها در امر پردازش زبان بسیار ضروری است. پژوهش حاضر تاجایی که مجری طرح آگاه است نخستین سامانه‌ی معنایی زبان فارسی بر مبنای مؤلفه‌های معنایی است.

۴-۱ اهداف پژوهش

هدف اصلی این طرح تهیه‌ی سامانه‌ی معنایی افعال زبان فارسی به همراه فهرستی از جملات مرتبط با این افعال بر پایه‌ی مؤلفه‌های معنایی است که منجر به سهولت کاربرد زبان فارسی در فضای مجازی می‌شود. اهداف کاربردی این پژوهش عبارتند از: تهیه‌ی غلطیاب‌های رایانه‌ای بر مبنای مؤلفه‌های معنایی برای زبان فارسی، کمک به شیوه‌های آموزش زبان فارسی به غیرفارسی‌زبانان و کمک به بازیابی اطلاعات با توجه به مؤلفه‌های معنایی. به طور کلی کاربرد پیکره‌ها عبارتند از پردازش زبان طبیعی، بررسی تحولات تاریخی زبان، استخراج اطلاعات زبانی خاص و کاربرد در فرهنگ‌نگاری است.

۵-۱ ساختار پژوهش

پژوهش حاضر شامل پنج فصل است. فصل اول شامل بیان مساله، اهداف، اهمیت پژوهش است. فصل دوم مبانی نظری پژوهش و پیشینه معرفی می‌شود. در فصل سوم جامعه‌ی آماری، نمونه‌ها و روش پژوهش معرفی می‌شود. در فصل چهارم نیز سامانه معنایی معرفی می‌شود.

فصل دوم: مبانی نظری و پیشینه‌ی پژوهش

۲-۱ مقدمه

در این فصل اشاره‌ی مختصری به مؤلفه‌های معنایی و محدودیت‌های گزینشی می‌شود و سپس پیشینه‌ی پژوهش در دو بخش معرفی پژوهش‌های صورت گرفته درباره‌ی افعال زبان فارسی و پژوهش‌های صورت گرفته با موضوع سامانه‌ها و پیکره‌های زبانی ارائه می‌گردد.

۲-۲ مؤلفه‌های معنایی و محدودیت‌های گزینشی

پیش از پرداختن به چیستی محدودیت‌های گزینشی لازم است به چگونگی شکل‌گیری آن اشاره شود. سرچشمه‌ی شکل‌گیری بحث محدودیت‌های گزینشی و مؤلفه‌های معنایی، بحث مختصات واجی در واج‌شناسی بود. در واقع پس از طرح مختصات واجی از سوی واج‌شناسان و به ویژه مؤلفه‌هایی نظیر [+واک] / [-واک]، [+خیشومی] / [-خیشومی] و جز آن که بر حسب آرای واج‌شناسان مکتب پراگ و به طور اخص در دیدگاه تروبتسکوی مؤلفه‌های جهانی به حساب آمدند، معنی‌شناسان نیز به این وسوسه افتادند تا با الگوگیری از واج‌شناسی، در مطالعه‌ی معنی نیز به مؤلفه‌هایی دست یابند که جهانی نماید (صفوی، ۱۳۷۹).

افعال برای انتخاب موضوع‌های خود محدودیت‌هایی را اعمال می‌کنند و به عبارتی هر کلمه‌ای با هر ویژگی معنایی را به عنوان فاعل یا مفعول نمی‌پذیرند (کروگر، ۲۰۰۵: ۱۷۳). محدودیت‌های گزینشی محدودیت‌هایی هستند که مانع از شکل‌گیری بسیاری از ترکیبات زبانی می‌شود به گونه‌ای که مثلاً افعال برای انتخاب موضوع‌های خود محدودیت‌هایی اعمال می‌کنند و به عبارتی هر کلمه‌ای با هر ویژگی معنایی را به عنوان فاعل و یا مفعول خود نمی‌پذیرند. به عنوان مثال فعل خوردن الزاماً نیاز به کنشگری جاندار و کنش‌پذیری جاندار و یا بی‌جان دارد. بر این اساس، مؤلفه‌های معنایی انتزاعی / عینی، جاندار / غیرجاندار، جانور / غیرجانور، انسان / غیر انسان، مذکر / مؤنث، ملاک تعیین محدودیت‌های گزینشی قرار می‌گیرد. در این پژوهش با بررسی موضوع‌های افعال با توجه به نوع موضوعاتی که به دلیل محدودیت‌های گزینشی می‌پذیرند، طبقه‌بندی نهایی صورت می‌گیرد.

۲-۳ پژوهش‌های داخلی و خارجی مرتبط

- پژوهش‌های صورت گرفته درباره‌ی افعال زبان فارسی

جامع‌ترین پژوهش در زمینه‌ی ظرفیت افعال فارسی را طبیب زاده (۱۳۸۵) انجام داده است. نویسنده در این کتاب فعل‌های زبان فارسی را با عنوان بسیط، مرکب، پیشوندی و پی‌بستی بررسی کرده و متمم‌های آنها را بازشناخته است و بر همین اساس افعال را طبقه‌بندی کرده است. نویسنده با بررسی ۲۰۰۰ فعل آنها را بر حسب ظرفیتشان به چهار طبقه‌ی تک ظرفیتی، دو ظرفیتی، سه ظرفیتی و چهار ظرفیتی تقسیم کرده است. او همچنین هر طبقه از افعال را نیز بر حسب متمم‌هایی که می‌گیرند، بازشناخته است مثلاً فعل‌های سه ظرفیتی را به پنج گروه تقسیم کرده است. نویسنده با تعیین ظرفیت افعال و تعیین متمم‌های آنها، ۲۳ ساخت بنیادین را در زبان فارسی باز می‌شناسد. مهاجر قمی (۱۹۷۸)، به نقل از طبیب زاده، (۱۳۸۵) نخستین اثری است که ساخت‌های ظرفیتی افعال فارسی را از دیدگاه نظریه‌ی ظرفیت واژگانی بررسی کرده است. وی در این پژوهش ۱۰ نوع متمم برای افعال فارسی برشمرده است. سمایی (۱۳۷۹) در مقاله‌ی خود معیارهای تمیز فعل مرکب را در نظام‌های پردازش زبان و به طور

ویژه در نرم‌افزاری به نام دستورسنج ارایه کرده است. این قواعد عبارت بودند از قواعد زیرمقوله‌ای و زایایی روابط مضمونی یا نقش‌های تتایی. محمد ابراهیمی و فرقانی‌فرد (۱۳۸۹) تلاش کرده‌اند تا با بهره‌گیری از تحلیل ساختار فعل مرکب فارسی گفتاری معیار در چارچوب نظریه‌ی حاکمیت و مرجع‌گزینی و نیز با استفاده از شیوه‌های نوین و علمی آموزش زبان، راهکار علمی و کارآمدی برای آموزش فعل مرکب فارسی به غیرفارسی‌زبانان ارایه دهند. آنان در این پژوهش از پیکره‌ی زبانی گفتار معیار به مدت ۳۰۰ دقیقه استفاده کردند. فرخ (۱۳۸۱) دسته‌بندی مبسوطی از افعال زبان فارسی برای تشخیص خودکار رایانه ارائه داده است. دانشکار آراسته (۱۳۸۱) نیز نرم‌افزاری برای تشخیص فعل طراحی کرده است.

رفیعی (۱۳۸۶) به بررسی محدودیت‌های گزینشی پرداخته و آن‌ها را فقط در سطح واژه و ساخت درونی واژه می‌داند و این محدودیت‌ها را به چهار دسته‌ی انتزاعی/عینی، فعل فعال/فعل غیر فعال، صفت فعال/صفت غیرفعال و قید فعال/قید غیر فعال تقسیم کرده است و بر این مبنا چگونگی افزوده شدن وندها به واحدهای مختلف زبانی را بررسی کرده است. مولودی، کوهستانی و رسولی (۱۳۹۰) در پژوهش خود به اهمیت و چگونگی تدوین فرهنگ ظرفیتی افعال فارسی پرداخته‌اند. همچنین آنان نقش این فرهنگ را در آموزش زبان فارسی به غیرفارسی‌زبانان تبیین کرده‌اند. راسخ مهند (۱۳۹۱) با بررسی افعال ربطی با توجه به محدودیت‌های گزینشی این افعال، چنین نتیجه گرفته است که این افعال در محدودیت‌های گزینشی هیچ نقشی ندارند و اصولاً فاقد معنایی خاص هستند. در واقع در جملات دارای فعل ربطی، اعمال محدودیت گزینشی بر عهده‌ی عنصر غیر فعلی است. منصوری-زاده و نصیری (۱۳۹۱) در گزارشی روابط ساختاری میان واژگان را مورد بررسی قرار داده‌اند و بخش اصلی این گزارش به جایگاه فعل و نقش فعل مربوط می‌شود. در این پژوهش ۵۰۰ فعل کاربردی در حوزه‌ی فن‌آوری اطلاعات و ارتباطات با توجه به معیار ظرفیت افعال (تعداد موضوعات مورد نیاز برای هر فعل) بررسی و طبقه‌بندی شده است. رفیعی و اظهري (۲۰۱۳) به بررسی نقش محدودیت‌های دستوری، آوایی، ساخت‌واژی، منظورشناختی و معنایی در پیوستن وند (انه) به پایه‌های فارسی پرداخته‌اند. آنان واژه‌های لازم را از دو فرهنگ لغت استخراج کرده‌اند و به بررسی وضعیت این واژه‌ها با توجه به محدودیت‌های گزینشی البته در سطح واژه پرداخته‌اند.

کروگر (۲۰۰۵) در بخشی از کتاب خود به بحث محدودیت‌های گزینشی پرداخته است. وی معتقد است محدودیت‌های گزینشی بر پایه‌ی دانش مشترک سخنگویان یک زبان انتخاب می‌شود و می‌تواند از زبانی به زبان دیگر متفاوت باشد. همچنین وی اعلام می‌دارد که محدودیت‌های گزینشی باید بر مبنای نقش‌های معنایی (عامل، پذیرا و...) تعریف شوند و نه بر اساس نقش‌های نحوی (فاعل، مفعول و...). شوارتز و گومز (۲۰۰۸) در پژوهشی تلاش می‌کنند با بهره‌گیری از یک پیکره‌ی زبانی با برچسب‌دهی به نقش‌های معنایی افعال کاربرد آن‌ها را در ماشین‌های vector تسهیل کنند.

- پژوهش‌های صورت گرفته با موضوع سامانه‌ها و پیکره‌های زبانی

رویکرد پیکره محور در زبان‌شناسی کاربردی در چند دهه‌ی اخیر با توجه به امکانات تحلیلی بسیار دقیق و متنوعی که برای مطالعه‌ی زبان فراهم می‌کند، مورد استقبال زیادی قرار گرفته است به‌طوری‌که بسیاری از کشورهای دنیا اقدام به تهیه و تدوین پیکره‌ی ملی زبان‌های بومی خود کرده‌اند. در این میان، می‌توان به پیکره‌ی ملی بریتانیا، پیکره‌ی ملی زبان روسی، پیکره‌ی ملی لهستانی، و پیکره‌ی ملی انگلیسی استرالیا اشاره کرد. علاوه بر پیکره‌های عظیم ملی، بسیاری دیگر از پیکره‌های خردتر که اهداف پژوهشی خاصی را دنبال می‌کرده‌اند نیز در

سال‌های اخیر ایجاد شده‌اند. پیکره‌ی تاریخی زبان انگلیسی، پیکره‌های موازی دوزبانه، پیکره‌های متون ادبی، و پیکره‌های متون ترجمه شده نمونه‌هایی از این پیکره‌ها هستند. نوع دیگری از پیکره‌ها که از اهمیت بالایی در آموزش زبان خارجی برخوردارند، پیکره‌های متنی زبان آموزان با زبان‌های مادری و سطوح توانش زبانی مختلف است. همچنین با بهره گرفتن از چنین پیکره‌هایی می‌توان به مقایسه‌ی جنبه‌های مختلف زبانی از قبیل انواع اشتباهات، بسامد استفاده از انواع واژگان، عبارات، جمله‌واره‌ها و ساختارهای دستوری در بین گروه‌های مختلف زبان آموزان پرداخت. مقایسه‌ی الگوهای مختلف استفاده از زبان توسط زبان آموزان و بومی زبانان نیز می‌تواند شناخت بسیار خوبی از ماهیت پیشرفت توانش زبانی فراهم کند.

در سال‌های اخیر پژوهش‌های بسیاری در حوزه‌ی پیکره‌ها و تهیه‌ی سامانه‌ها و پیکره‌ها صورت گرفته است. در این جا به معرفی برخی از این سامانه‌ها و پیکره‌ها که در داخل و خارج از ایران از زبان فارسی جمع آوری و تهیه شده است، می‌پردازیم.

۱- پیکره‌ی داده‌ی کسره‌ی اضافه‌ی زبان فارسی که به بررسی نزدیک به ۳۰ هزار جمله‌ی زبان فارسی می‌پردازد (نوریان، رسولی، ایمانی و فیلی، ۲۰۱۵).

۲- واژگان نحوی و معنایی افعال مرکب فارسی: مجموعه‌ای است چندزبانه شامل اطلاعات نحوی و معنایی افعال مرکب زبان فارسی، ترجمه‌ی انگلیسی و فرانسوی افعال و حداقل یک جمله مثال برای هر فعل. این پیکره شامل اطلاعات مربوط به بیش از ۶۰۰ فعل مرکب شامل همکرد «زدن» است (سموالیان و فقیری، ۲۰۱۳).

۳- دادگان درختی فارسی در چارچوب دستور ساخت سازه‌ای هسته‌بنیان: مجموعه‌ای است شامل بیش از ۱۰۰۰ جمله‌ی برچسب‌خورده با اطلاعات نحوی (قیومی، ۲۰۱۲).

۴- دادگان صوتی هجاهای فارسی: دادگان گفتاری حاضر شامل حدود ۵۸۰۰ فایل صوتی هجاهای زبان فارسی است (اسلامی، شیخ زادگان، احمدی نیا و بهرامی راد، ۱۳۸۸).

۵- فاسیل متشکل از دو مجموعه داده برای ارزیابی روش‌های خطایابی املائی است. گروه اول شامل خطاهای معمولی است که از دانش‌آموزان مدارس و همچنین خطاهای املائی در هنگام تایپ فارسی جمع‌آوری شده‌اند (براری و قاسم زادگان، ۲۰۰۵).

۶- فارس‌نت (وردنت عمومی زبان فارسی)، پایگاه دانشی است که حاوی اطلاعات در مورد واژه‌ها و ترکیبات زبان (مفاهیم)، اطلاعات نحوی آن‌ها و روابط معنایی میان آن‌ها است (شمس‌فرد، حسابی، فدایی، منصوری، فامیان، باقریگی و فکری، ۲۰۱۰).

۷- فرهنگ ظرفیت نحوی افعال فارسی (نسخه ۳،۰): مجموعه‌ای است حاوی اطلاعات مربوط به ظرفیت نحوی بیش از ۴۵۰۰ فعل در زبان فارسی. در این فرهنگ، متمم‌های اجباری و اختیاری انواع فعل‌های ساده، مرکب، پیشوندی و عبارات فعلی مشخص شده است. فراوانی فعل‌های مرکب در زبان فارسی، نیاز به فرهنگ ظرفیت فعل را در این زبان دوچندان می‌نماید. چرا که شناخت فعل‌های مرکب چه از لحاظ انسانی

و چه از لحاظ پردازشی کاری دشوارتر از شناخت فعل‌های ساده است و به همین خاطر فراهم آوردن فهرستی از فعل‌های زبان (که شامل فعل‌های مرکب نیز می‌شود) به همراه ساخت‌های ظرفیتی افعال، کمکی شایان برای کارهای پردازشی است. از سوی دیگر، بر اساس نظریه‌ی وابستگی، ساخت بنیادین جمله را می‌توان از روی ساخت ظرفیتی فعل جمله به دست آورد و به همین دلیل بر اهمیت دانستن ساخت‌های ظرفیتی فعل در متن‌های زبانی افزوده می‌شود (رسولی، مولودی، کوهستانی و مینایی گل، ۲۰۱۱).

۸- دادگان «مجموعه افعال تصریف‌شده‌ی فارسی» مجموعه‌ای است که به گردآوری مصادر مختلف فارسی در پیکره‌های بی‌جن‌خان، همشهری، تب‌کورپ و چند منبع دیگر پرداخته است (خلیفه سلطانی، گل مکانی، وحدانی و معلمی، ۲۰۱۰).

۹- مجموعه‌ی همشهری (نسخه‌ی ۲): پیکره‌ای است حاوی ۳۱۸ هزار سند مربوط به اخبار سال‌های ۱۳۷۵ تا ۱۳۸۶ که با خزش (Crawl) وبسایت همشهری و چندین مرحله پیش‌پردازش و برچسب‌زنی حاصل آمده است. همه‌ی اسناد مجموعه‌ی همشهری دارای برچسب «Cat» هستند که نشان می‌دهد هر سند در چه رده‌ای است (اقتصادی، سیاسی و...) (آل احمد، امیری، درودی، ره‌گذر و ارومچیان، ۲۰۰۹).

۱۰- پایگاه داده‌های زبان فارسی: مجموعه‌ای است از متون مختلف فارسی که بخشی از آن دارای نشانه‌گذاری‌هایی از جمله شناسنامه‌ی متن، برچسب‌های دستوری، آوایی، ریشه‌ای و معنایی است. این دادگان مجهز به نرم‌افزارهای اختصاصی جستجو، تقطیع و تحلیل متن است که می‌تواند انواع فهرست‌های واژگانی، بسامدی و آماری را ارائه کند (عاصی، ۱۹۹۷).

۱۱ - پرسیکا: پیکره‌ای است حاوی متون خبری برگرفته از خبرگزاری ایسنا. متون این پیکره در یازده طبقه‌ی موضوعی شامل ورزشی، اقتصادی، فرهنگی، مذهبی، تاریخی، سیاسی، علمی، اجتماعی، آموزشی، حقوق قضایی و بهداشت طبقه‌بندی شده‌اند (اقبال زاده، حسینی، خدیوی و خدابخش، ۲۰۱۲).

۱۲ - پیکره‌ی بی‌جن‌خان مجموعه‌ای است از متون فارسی شامل بیش از ۲ میلیون و ۶۰۰ هزار واژه که با ۵۵۰ نوع برچسب POS برچسب‌گذاری شده‌اند. این پیکره همچنین شامل بیش از ۴۳۰۰ برچسب موضوعی چون سیاسی، تاریخی و ... برای متون است (بیجن خان، ۱۳۸۳).

۱۳- پیکره‌ی فارسی روز پیکره‌ای خام، یک‌زبان و پیوسته، شامل ۱۲۷ متن فارسی است که تمام آن‌ها بین سال‌های ۱۳۸۰ تا ۱۳۹۲ نوشته شده‌اند و اغلب از میان داستان‌های کوتاه و رمان‌ها انتخاب شده‌اند. تعداد کل واژه‌های موجود در پیکره‌ی فارسی روز ۴,۲۷۴,۳۶۰ واژه است و ۱۴۴,۲۱۰ واژه‌ی بدون تکرار در آن وجود دارد (قربان زاده، ۱۳۹۴).

۱۴- پیکره‌ی نور مجموعه‌ای است عظیم از متون اسلامی که عمدتاً به زبان عربی و فارسی نوشته شده‌اند. متون مربوط به بیش از ۴ هزار کتاب عربی (بیش از یک میلیارد کلمه) و بیش از ۲ هزار کتاب فارسی (بیش از ۶۰۰ میلیون کلمه) و همچنین بالغ بر ۲۰۰ کتاب دوزبان (عربی-فارسی) است (الهی منش، مینایی گل، غلامی و جوزی، ۲۰۱۲).

۱۵- پیکره‌ی وابستگی نحوی زبان فارسی: نخستین پیکره‌ی نحوی زبان فارسی است که شامل حدود ۳۰ هزار جمله‌ی برچسب خورده که اطلاعات نحوی و ساخت‌واژی را بر مبنای دستور وابستگی تهیه و عرضه نموده است (رسولی، کوهستانی و مولودی، ۲۰۱۳).

۱۶- پیکره‌ی گفتار محاوره‌ای زبان فارسی که شامل ۳۵۰ ساعت داده‌ی گفتاری است که به‌صورت سیگنال اکوستیکی از فارسی‌زبانان داوطلب در موقعیت‌های مختلف ارتباطی با هدف تحقیقات کاربردی و آموزش و آزمون سامانه‌های محاوره‌ای ضبط شده است و در سطح نوبت و پاره‌گفتار نشانه‌گذاری می‌شود (بیجن خان، ۱۳۹۵).

اطلاعات کلیه‌ی پیکره‌های معرفی شده‌ی بالا از پایگاه www.peykaregan.ir قابل دست‌یابی است.

۱۷- نخستین پیکره‌ی نقش‌های معنایی توسط میرزایی و مولودی (۱۳۹۳) ارائه شده است. این پیکره نزدیک به ۳۰۰۰۰ جمله‌ی فارسی معاصر را شامل می‌شود که به صورت دستی برچسب‌گذاری شده است. این پیکره بر اساس مفهوم نقش‌های معنایی فیلمور، لایه‌ای از اطلاعات مربوط به رابطه‌ی محمول به موضوع را به ساخت نحوی پیکره می‌افزاید.

۱۸- پیکره‌ی زبان‌آموز «سلام فارسی» که به منظور جمع‌آوری تولیدات نوشتاری و گفتاری فارسی‌آموزان خارجی و نشانه‌گذاری خطاهای زبانی ایشان طراحی و تشکیل شده است (صفری، ۱۳۹۶).

۱۹- پیکره‌ی به‌روز شونده‌ی متون فارسی: این پیکره توسط پژوهشگران دانشگاه شیراز به سرپرستی دکتر امیرسعید مولودی در سال ۱۳۹۷ به بهره‌برداری رسید. در این طرح، متون خبری به صورت متوازن از ۶ حوزه‌ی موضوعی و از خبرگزاری‌های مختلف، به صورت روزانه به این پیکره افزوده می‌شوند. در این طرح کلیه‌ی متون به صورت خودکار نرمال‌سازی و ریشه‌یابی می‌شوند و برچسب اجزای کلام به واژه‌های آن اختصاص می‌یابد.

همچنین در ذیل به برخی از طرح‌ها و پیکره‌های تهیه شده در زبان‌های دیگر اشاره می‌شود.

طرح پروب بنک (پالمر و همکاران، ۲۰۰۵) پیکره‌ی دادگان معنایی زبان انگلیسی است که نقش‌های معنایی را به پیکره‌ی دادگان نحوی افزوده است. طرح فریم نت (بیکر و همکاران، ۱۹۹۸) از سال ۱۹۹۷ بر مبنای نظریات فیلمور (۱۹۷۶) درباره‌ی ظرفیت افعال زبان انگلیسی، آغاز شد. طرح ورب نت (کیپر و همکاران، ۲۰۰۶) بزرگ‌ترین فرهنگ رایانشی افعال زبان انگلیسی است. در این سامانه طبقه‌بندی افعال بر مبنای نقش‌های تتایی، محدودیت‌های گزینشی و چارچوب‌های نحوی صورت گرفته است. پیکره‌ی ملی زبان روسی پیکره‌ای است که جست‌وجو را بر اساس مشخصه‌های واژگانی و معنایی افعال ممکن می‌سازد. در این پیکره، طبقه‌بندی واژه‌ها بر اساس نقش‌های تتایی انجام شده است (کوستاوا و همکاران، ۲۰۰۹). همچنین فرهنگ معنایی افعال روسی نیز تدوین شده است که در این فرهنگ، طبقه‌بندی بر مبنای معانی مشترک انجام شده است (متویا، ۱۹۸۹). طبقه‌بندی پایه‌ای افعال ترکی در کارهای اوروساوا (۱۹۸۳) و گانیو (۱۹۸۴) انجام شده است. آنان افعال را به ۱۰ تا ۱۵ طبقه تقسیم کردند. اما

پژوهش‌های آن‌ها مبنای تهیه پیکره قرار نگرفت. در حال حاضر کار برای تهیه پیکره معنایی زبان ترکی و حتی گویش‌های آن در حال انجام است. پیکره معنایی زالسا (بوخارت، ۲۰۰۶) با ۴۰ هزار جمله برای زبان آلمانی تولید شده است. این پیکره با افزودن لایه‌ای از اطلاعات معنایی به پیکره دادگان نحوی زبان آلمانی ایجاد شده است. پیکره ملی زبان تاتار به دو صورت، برچسب‌گذاری معنایی شده است: برچسب‌های مقوله‌ای و برچسب‌های تتایی (۲۰۱۶).

فصل سوم: روش شناسی پژوهش

۱-۳ مقدمه

در این فصل به معرفی جامعه‌ی آماری پژوهش، نمونه‌های پژوهش و روش نمونه‌گیری، ابزارهای گردآوری داده‌ها و روش انجام پژوهش می‌پردازیم.

۲-۳ جامعه‌ی آماری و نمونه‌ها

جامعه‌ی آماری این پژوهش کلیه‌ی افعال ساده، مرکب، پیشوندی و پیشوندی-مرکب فهرست شده در سامانه‌ی ظرفیت نحوی افعال است. این سامانه توسط رسولی، مولودی، کوهستانی و مینایی بیدگلی (۲۰۱۱) تهیه شده است و رسولی (۱۳۹۱) آن را معرفی کرده است و این سامانه امروزه در دسترس همگان قرار دارد. نمونه‌های مورد بررسی کلیه‌ی افعال متداول زبان فارسی است که در پیکره‌ها نمونه‌هایی برای آنها به دست آمد و یا نمونه‌هایی برای آنها ارائه گردید.

۳-۳ روش انجام پژوهش

گام نخست: استخراج افعال و تعیین مؤلفه‌های معنایی

نمونه‌های مورد بررسی در این پژوهش کلیه‌ی افعال متداول در بین سخنگویان زبان فارسی در گفتار روزمره و متون نوشتاری غیر ادبی است. چگونگی انتخاب این افعال بدین صورت بود که در گام نخست این افعال در سه پیکره‌ی همشهری، پرسیکا و بیجن‌خان جست‌وجو شدند. قاعدتا بخشی از این افعال در این سه پیکره وجود نداشتند. با توجه به عدم وجود این افعال در این پیکره‌ها و همچنین نظر مجری و همکاران بخشی از این افعال به عنوان افعال کاملاً مهجور، حذف شدند افعال مزبور در فازهای اول و دوم این طرح استخراج شدند (البته شایان ذکر است که در مرحله‌ی استخراج افعال ساده از فرهنگ واژگان شش جلدی معین نیز استفاده شده است).

در استخراج افعال و تعیین مؤلفه‌های معنایی موارد ذیل مورد توجه بوده است:

۱- در این پژوهش محدودیت‌های گزینشی تنها بر مبنای کاربرد افعال در گونه‌ی رسمی نوشتاری بررسی شده‌اند و نه در گونه‌های غیر رسمی و ادبی. زیرا به ویژه در گونه‌ی ادبی به دلیل بهره‌گیری از صنایع ادبی، محدودیت‌های گزینشی افعال دستخوش تغییر می‌شود.

۲- در بررسی برخی از افعال نمی‌توان به محدودیت‌های گزینشی قطعی برای آنها دست یافت و چه بسا کاربر این سامانه ممکن است موارد دیگری را ذکر کند و یا برخی از این محدودیت‌های گزینشی را نپذیرد. تلاش این سامانه تعیین مؤلفه‌های حداقلی برای افعال است که در اغلب موارد این مؤلفه‌ها قطعی هستند و در مواردی نیز مرز شفافیت بین مؤلفه‌ها وجود ندارد. به دلیل این که این پژوهش، پژوهشی معنی‌شناختی است، دستیابی به قطعیت و مرزهای شفاف در مواردی مشکل و گاهی ناممکن است.

۳- بسیاری از افعال به ویژه افعال مرکب، هم کاربرد عینی و هم کاربردی انتزاعی و اصطلاحی دارند. در واقع می‌توان گفت که این افعال بر روی یک پیوستار قرار گرفته‌اند که در یک سوی آن کاربردی کاملاً

انتزاعی و اصطلاحی و در سوی دیگر کاربردی کاملاً عینی دارند. مثال:

<u>انتزاعی</u>	<u>عینی</u>
به پا کردن دعوا	به پا کردن کفش
سر رفتن حوصله	سر رفتن قابلمه
جا داشتن (امکان پذیر بودن)	جا داشتن (مکان)
ریشه دواندن مشکلات	ریشه دواندن گیاه
ریختن پول به حساب	ریختن آب
وارفتن آرش	وارفتن نان
به هم زدن قرار	به هم زدن مخلوط
افت کردن محبوبیت	افت کردن سطح آب
بازسازی اندیشه	بازسازی خانه
نفوذ کردن فکر	نفوذ کردن آب
آمیختن شعر و ادبیات	آمیختن آرد و آب
فرورفتن در غم	فرورفتن در آب
شکستن عهد	شکستن چوب
برق زدن چشم	برق زدن شیء
دست کشیدن (رها کردن)	دست کشیدن بر چیزی
پا پس کشیدن (خودداری کردن)	پس کشیدن دست
آرام گرفتن قیمت	آرام گرفتن انسان یا حیوان
دست و پا زدن (تقلا کردن)	دست و پا زدن در آب
به جا ماندن غم	به جا ماندن وسایل
برگرداندن (ترجمه کردن)	برگرداندن سر یا پس دادن چیزی
جا گذاشتن اثر	جا گذاشتن چیزی
جا گذاردن اثر	جا گذاردن چیزی
عقب نشینی کردن قیمت ها	عقب نشینی کردن افراد
خریدن وقت و اعتبار	خریدن کالا
دزدیدن ایمان	دزدیدن کالا
زدودن افکار	زدودن ماده
گسستن ریسمان الهی	گسستن بند
ماسیدن خنده	ماسیدن ماده
مردن امید و آرزو	مردن انسان یا حیوان
ربودن ایمان	ربودن کالا

در این پژوهش تا حد امکان کاربرد عینی افعال مدنظر قرار گرفته است و در مواردی نیز این افعال به صورت دو فعل

جداگانه بررسی شده‌اند.

۴- در گزینش جملات نمونه تلاش شده است تا حد امکان از جملات امری، سببی، درونه‌ای، بدون فاعل، به‌هم‌ریخته، پرسشی، مجهول، اصطلاحات و ضرب‌المثل‌ها، اشعار و همچنین جملاتی که بار معنایی منفی و یا جهت‌گیری خاصی دارند، پرهیز شود.

۵- تا حد امکان تلاش شده است تا افعال در حالتی که در معنای اولیه‌ی خود به‌کار رفته‌اند، گزینش شوند. به عنوان مثال:

آفریدن توسط خداوند صورت می‌گیرد	آفریدن اثر هنری توسط انسان
آمرزیدن توسط خداوند صورت می‌گیرد	آمرزیدن انسان توسط انسان
جنگیدن بین انسان‌ها و یا حیوانات	جنگ در ورزش یا جنگ با نفس
غریدن شیر	غریدن انسان
فهمیدن سخن و گفتار	فهمیدن درد و رنج
نگریستن به معنای نگاه کردن	نگریستن به معنای بررسی کردن
نگه داشتن به معنای حفظ کردن	نگه داشتن در راضی نگه داشتن یا سالم نگه داشتن
همراه داشتن چیزی یا کسی	همراه داشتن آلودگی یا نتیجه
تکیه زدن فرد بر چیزی	تکیه زدن بر قدرت
چنبره زدن مار	چنبره زدن انسان
دست زدن انسان یا حیوان به چیزی	دست زدن به عملیات
انعکاس یافتن نور	انعکاس یافتن اندیشه
لنگیدن انسان یا حیوان	لنگیدن کار

۶- افعال بسیاری وجود دارند که در همه‌ی متون کاربرد ندارند و به نظر می‌رسد کاربرد اصلی آنها در متون خاص (متون ادبی، تاریخی، حقوقی و ...) است بنابراین اغلب این افعال تعداد مثال‌های کافی در پیکره‌ها نداشتند. مثال:

خشکیدن، خفتن، زدودن، گستردن، نگاشتن، نهادن، نهفتن، پنجه افکندن، پنجه انداختن، گسستن، حسد بردن، سجده بردن، گمان بردن، طمع بستن، سبقت جستن، آواز دادن، اشارت داشتن، بردمیدن، بانگ زدن، بانگ کردن، بچه کردن، بیختن، تصدق کردن، عمارت کردن، سعایت کردن، قربان کردن، قضا کردن، هزیمت کردن، برنشستن، مسترد کردن، معروض داشتن، پیمانانه زدن، فرودوختن، بازفرستادن، درکشیدن، برگزشتن، فروگرفتن، مشاورت کردن، منازعت کردن

۷- تلاش شده است که جملات ارائه شده به عنوان نمونه، تا حد امکان موضوعاتی را که نمایانگر مؤلفه‌ها هستند، در خود داشته باشند. البته مواردی نیز وجود دارد که به دلیل نبودن مثال‌های مناسب، ممکن است مؤلفه‌ای نماینده نداشته باشد.

۸- کلیه‌ی جملات ارائه شده مورد بررسی قرار گرفته‌اند و واحدهایی که جزء موضوعات فعل به شمار نمی‌روند حذف شده‌اند، البته در بعضی موارد که ممکن بود معنای جمله به کلی تغییر کند و یا جمله غیرقابل فهم شود، برخی از این واحدها حذف نشده‌اند. شایان ذکر است که چون اغلب جمله‌ها از پیکره‌ها استخراج شده‌اند (به‌منظور ارائه‌ی مثال‌های طبیعی)، جملات ناقص در بین آنها وجود دارد.

۹- با توجه به این که ضمائر زبان فارسی نسبت به جاننداری و بی‌جانی و همچنین نسبت به جنسیت اصولاً بی‌نشان هستند، در برخی از مثال‌ها ممکن است مؤلفه‌ها هم‌پوشانی داشته و یا متناقض به نظر برسند.
۱۰- افعال هم‌نویسه نیز مورد توجه قرار گرفته‌اند:

فوت کردن [fut]/ فوت کردن [fowt]، سیر کردن [sir]/ سیر کردن [seir]، گل کردن [gel]/ گل کردن [goll]، عرضه داشتن [arze]/ عرضه داشتن [orze]، قل خوردن [qel]/ قل خوردن [qol]، کشتن [koftæn]/ کشتن [keftæn]، گزیدن [gæzidæn]/ گزیدن [gozidæn]

۱۱- نمونه‌هایی از افعال هم‌آوا - هم‌نویسه نیز وجود داشت:

میل کردن به معنای خوردن و تمایل یافتن
گذشتن به معنای عبور مکانی، عبور زمانی و بخشیدن
گردیدن و گشتن به معنای چرخیدن و شدن
ساختن به معنای بنا کردن و کنار آمدن
بردن به معنای حمل کردن و پیروز شدن
پرداختن به معنای توجه کردن و پرداختن پول
پیچیدن به معنای دور زدن و پیچیدن چیزی
کشیدن به معنای تحمل کردن و کشیدن چیزی
گذراندن به معنای گذراندن وقت و عبور دادن چیزی

۱۲- در حالتی که فاعل جمله یکی از واژه‌های زیر بود و این واژه‌ها عملی انسانی را انجام می‌دادند، آن موضوع با مؤلفه‌ی [+ انسان] مشخص شده است. پیش‌فرض چنین است که انسان‌های شاغل یا حاضر در این مکان‌ها و یا اجتماعات انسانی، کاری را انجام می‌دهند.

دسته، هیأت، گروه، شرکت، مؤسسه، دولت، حکومت، حاکمیت، قوه، باشگاه، هیأت مدیره، هیأت مؤسس، هیأت رئیسه، خانواده، فامیل، تیم، کارخانه، کارگاه، مکتب، فرقه، قوم، قبیله، طایفه، ایل، کمیسیون، کمیته، فدراسیون، نمایندگی، جنبش، جریان، جبهه، شبکه، باند، اتحادیه، آژانس، کنگره، سنا، مجلس، شورا، حزب، نهاد، مجمع، جمعیت، جامعه، انجمن، رادیو، تلویزیون، روزنامه، مجله، دانشگاه، دانشکده، گروه، کانون، اداره، حوزه، نام کشور، نام شهر، وطن، میهن، وزارت، بنگاه، سازمان، انستیتو، دستگاه، بانک، ستاد، دفتر

۱۳- درباره‌ی موضوع اول برخی افعال با یک مشکل مواجه شدیم که تعیین مؤلفه‌ها را دشوار می‌نمود و آن هم افعالی که به دلیل این که عملی را انسان با واسطه‌ی چیزی انجام می‌دهد، کنشگر آنها می‌تواند انسان و غیرانسان باشد. در چنین مواردی مؤلفه‌های [+ جاندار] [+ انسان] در نظر گرفته شد، مانند: ترمز کردن (انسان، خودرو)، گاز دادن (انسان، خودرو)، سرعت گرفتن (انسان، خودرو)، اوج گرفتن (انسان، هواپیما)، شخم زدن (انسان، تراکتور)، شکستن (انسان، شیء)، بریدن (انسان، اره، چاقو)، درو کردن (انسان، کمباین)، صاف کردن (انسان، وسیله)، آمیختن (انسان، وسیله)، کشیدن (انسان، وسیله)، سبقت گرفتن (انسان، خودرو)، گشت‌زنی کردن (انسان، خودرو)، گران‌فروشی کردن (انسان، فروشگاه)،

آب‌پاشی کردن (انسان، وسیله)، عکس‌برداری کردن (انسان، دوربین)، دوشیدن (انسان، دستگاه)، دوختن (انسان، چرخ خیاطی)، روبیدن (انسان، دستگاه)، شستن (انسان، دستگاه)، شکافتن (انسان، کلنگ)، نشستن (انسان، هواپیما)

۱۴- توجه به این مسأله که وجود فاعل در جملات فارسی اختیاری است و بجز در موارد تأکیدی، در اغلب جملات فاعل حذف می‌شود. این مسأله بر گزینش نمونه‌ها و همچنین تعیین مؤلفه‌های معنایی تأثیرگذار بود بنابراین تلاش شد تا حد امکان جملاتی گزینش شوند که فاعل آشکار دارند. در موارد مورد لزوم نیز با توجه به فعل جمله، فاعلی/کنشگری به صورت ضمیر اضافه گردید.

۱۵- تا حد امکان از گزینش جملاتی که ساختار غیرشخصی داشتند، جملاتی با ساختار «توان، باید و شاید»، به دلیل نداشتن فاعل، پرهیز شد.

۱۶- برخی افعال مرکب به دلیل ساختار خاصشان امکان اشتباه در تشخیص فعل مرکب را افزایش می‌دهند به این صورت که سازه‌ای دیگر می‌تواند بین دو بخش آنها قرار گیرد و در ظاهر فعل مرکب جدیدی را تشکیل دهد. این مسأله در یافتن جمله‌های مناسب مورد توجه قرار گرفت. برخی از این افعال عبارتند از:

نیاز داشتن، قصد داشتن، امکان داشتن، اعتراف کردن، وارد کردن، چاره یافتن، اعلام کردن، تقاضا کردن، دست زدن، صرف کردن، ابراز کردن، متهم کردن، اقدام کردن، ایجاد کردن، احساس کردن، دچار کردن، شروع کردن، تبدیل کردن، تهدید کردن، وادار کردن، وقف کردن، دعوت کردن، پیشنهاد دادن، ملزم کردن

۱۷- در جست‌وجوی افعال، حالات مختلف مورد توجه قرار گرفت: توجه به فاصله، نیم‌فاصله و چسبیده بودن واحدهای سازنده‌ی افعال مرکب. در بسیاری از موارد وجود فاصله و یا چسبیده بودن واحدها منجر به بازیابی متفاوت افعال می‌شد ولی تنها یک حالت مورد گزینش قرار گرفت.

۱۸- برخی از افعال شکل‌های مختلف نوشتاری داشتند که بازیابی‌های متفاوتی را به همراه داشت. افعالی مانند: لشکرکشی کردن/لشگرکشی کردن و مسألت کردن/مسئلت کردن/مسالت کردن

۱۹- در موارد بسیار معدودی نیز غلط تاییی منجر به بازیابی‌های نادرست شد. مانند: عهده کردن به جای عهد کردن، جلد کردن به جای جلب کردن

۲۰- در مواردی که کنشگر یا تجربه‌گر جمله یکی از اجزاء بدن بود، مؤلفه‌ی [+ جاندار] در نظر گرفته شد. مواردی مانند: دستم شکست، دلش درد گرفت، قلبم میتپد.

۲۱- در افعالی که یکی از موضوعات آن یک جمله بود، مانند افعال «توانستن» و «خواستن»، مؤلفه‌ی [- عینی] برای جمله لحاظ شد.

۲۲- در افعالی که یکی از موضوعات آن، یک صفت بود، مؤلفه‌ی [- عینی] برای صفت لحاظ شد.

گام دوم: تهیه‌ی فهرست نهایی افعال و سامانه‌ی معنایی

تهیه‌ی سامانه‌ی معنایی در سه مرحله صورت گرفت: جمع‌بندی و بررسی نهایی افعال و مؤلفه‌های معنایی موضوعات و تهیه‌ی سامانه و ارزیابی سامانه. برون‌داد این طرح یک سامانه در قالب فایل اکسل است که می‌توان با وارد کردن فعل، مؤلفه‌های معنایی موضوعات آن را مشاهده کرد.

گام سوم: تهیه‌ی پیکره‌ی جملات (فهرست جملات مرتبط)

استخراج جملات: در این مرحله از طرح، استخراج جملات از پیکره‌های همشهری، پرسیکا و بیجن‌خان صورت گرفت. بدین منظور نخست جملاتی که حاوی افعال ساده، پیشوندی، مرکب و پیشوندی مرکب بودند (افعال موجود در سامانه‌ی ظرفیت نحوی افعال) استخراج شدند. قاعدتا تعدادی از افعال در این پیکره‌ها وجود نداشتند ولی افعال متداول فارسی هستند. به منظور دستیابی به مؤلفه‌های معنایی موضوعات این افعال و محدودیت‌های گزینشی از پنج سخنگوی بومی فارسی زبان به منظور ارائه‌ی این جملات کمک گرفته شده است. جست و جو در این پیکره-ها بدین صورت بود که برای افعال ساده سقف ۲۵۰ جست‌وجو و برای افعال مرکب و پیشوندی سقف ۵۰ جست‌وجو تعیین گردید. جست‌وجوها به ترتیب در پیکره‌های همشهری سپس پرسیکا و در نهایت در بیجن‌خان انجام گرفت. در طول روند جست‌وجو هر جا که جست‌وجوی فعلی به سقف تعیین شده می‌رسید، جست‌وجو متوقف می‌گردید. البته بسیاری از افعال مثال‌هایی کمتر از پنج مورد در هر سه پیکره داشتند.

پالایش جملات: این مرحله طی سه گام صورت گرفت. در گام نخست کلیه‌ی جملات استخراج شده بررسی شد و مثال‌های ناقص و نامرتب حذف گردید و برای هر فعل در حدود ۲۰ مثال باقی ماند. در گام دوم با بررسی دوباره‌ی مثال‌ها برای هر فعل حداکثر ۵ مثال باقی ماند و در گام سوم، این ۵ جمله پالایش شدند و سازه‌هایی که نقش موضوعی ندارند مانند قیده‌ها، صفات، مضاف‌الیه‌ها، متمم‌ها و جملات درونه‌ای، حذف شدند. در مجموع ۴۱۹۹ فعل در سامانه فهرست شده است که ۲۴۵ مورد، فعل ساده و ۳۹۵۴ فعل، مرکب، پیشوندی و پیشوندی-مرکب هستند. ۱۵۸۴ فعل دارای یک موضوع، ۲۴۴۳ فعل، دو موضوع و ۱۷۳ فعل، سه موضوع دارند. ۳۵۴۲ فعل در پیکره‌ها موجود بودند و ۶۵۷ فعل در پیکره‌ها وجود نداشتند. افعال حساس به جنسیت نیز ۲۵ فعل شناخته شد. پیکره‌ی به دست آمده نیز شامل ۱۳۷۵۰ جمله است که برای هر فعل، ۳/۳ جمله وجود دارد.

برونداد این طرح عبارتست از فهرستی از افعال که شامل حداکثر ۵ جمله (مثال) و مؤلفه‌های معنایی موضوع‌های هر فعل متداول زبان فارسی اعم از افعال ساده، مرکب، پیشوندی و پیشوندی مرکب، است. این پیکره به صورت ساختارمند با استفاده از قالب اکسل تولید و ذخیره شده است؛ بنابراین از این فهرست می‌توان در مطالعات معنایی افعال و بسیاری از کاربردها در حوزه‌ی پردازش زبان طبیعی بهره برد. ساختار اکسل اجازه می‌دهد که پیکره به آسانی توسط رایانه قابل خواندن باشد و همچنین به صورت متنی نیز برای استفاده‌ی انسانی قابل بهره‌برداری باشد.

فصل چهارم: معرفی سامانه

۱-۴ مقدمه

در این فصل به معرفی سامانه‌ی معنایی و چگونگی کار با آن می‌پردازیم.

۲-۴ معرفی سامانه

در این سامانه یک جعبه برای جست‌وجوی فعل، یک جعبه برای تعیین نوع فعل (ساده، مرکب و پیشوندی)، پنج جعبه برای نمایش جمله‌های نمونه برای هر فعل و ۱۵ جعبه نیز برای نمایش مؤلفه‌های معنای افعال تعبیه شده است.



برای یافتن مؤلفه‌های معنایی هر فعل و نمایش جمله‌های نمونه، فعل مورد نظر در جعبه‌ی جست‌وجوی فعل تایپ می‌شود و دکمه‌ی جست‌وجو فشار داده می‌شود. در این حالت اطلاعات فعل مورد نظر نمایش داده می‌شود. افعال ستاره‌دار افعالی هستند که یا در پیکره‌ها موجود نبوده‌اند و یا این که مثال‌های قابل قبولی نداشته‌اند.

جستجوی فعل		
نوع فعل		
مؤلفه‌های معنایی		
موضوع اول	موضوع دوم	موضوع سوم

مرکز مخطی‌ای اطلاع‌رسانی علوم و فناوری

سامانه معنایی افعال زبان فارسی

جستجو

در ادامه نتایج جست‌وجوی چند فعل نشان داده می‌شود.

جستجوی فعل		
نوع فعل		
مؤلفه‌های معنایی		
موضوع اول	موضوع دوم	موضوع سوم
0	0	0
0	+/- جاندار	0
0	+ جانور	0
0	0	+ انسان
0	0	0

مثال‌ها

سپاهان ماهی بادکنک را داد کردند

نسا لاسیک را داد تکبید.

پزشکان یک بادکنک را داد کردند

0

0

مرکز مخطی‌ای اطلاع‌رسانی علوم و فناوری

سامانه معنایی افعال زبان فارسی

جستجو

بازیابی اطلاعات

انواع داشتن

جستجوی فعل

نوع فعل	مركب و پیشوندی
مولفه‌های معنایی	
موضوع اول	موضوع دوم
موضوع سوم	
0	- عینی
0	-
0	-
0	-
0	-
0	-
0	+
0	0
0	0

مثال‌ها

رئیس جمهور عرب تکران خود را ابلاغ دانست.

رئیس پیام خود را به وی ابلاغ داشت

0

0



مرکز مصلحت‌های اطلاع‌رسانی علوم و فناوری

سامانه معنایی افعال زبان فارسی

جستجو

بازیابی اطلاعات

بسته یا زده

جستجوی فعل

نوع فعل	مركب و پیشوندی
مولفه‌های معنایی	
موضوع اول	موضوع دوم
موضوع سوم	
0	- عینی
0	-
0	-
0	-
0	-
0	-
0	+
0	0
0	0

مثال‌ها

ما به آنها پشت پا زدیم

پارک به وعده هایش پشت پا زد.

دولت به موافقت نامه پشت پا زده است.

وی به دنیا پشت پا زده بود

دولت به آنها پشت پا زد.



مرکز مصلحت‌های اطلاع‌رسانی علوم و فناوری

سامانه معنایی افعال زبان فارسی

جستجو

بازیابی اطلاعات

جستجوی فعل

برده (موز) شدن

مثال‌ها	نوع فعل		
	ساده	مولفه‌های معنایی	
	موضوع سوم	موضوع دوم	موضوع اول
او سینه را برد	0	- عینی	0
ما رفقت را می‌بریم	0	-	-
-	0	-	+ جانور
-	0	-	-
-	0	0	0



مرکز مهندسی اطلاع‌رسانی علوم و فناوری

سامانه معنایی افعال زبان فارسی

جستجو



بازیابی اطلاعات

جستجوی فعل

گران آمدن

مثال‌ها	نوع فعل		
	مرکب و پیوندی	مولفه‌های معنایی	
	موضوع سوم	موضوع دوم	موضوع اول
این تصمیم به خاکشان، گران آمد	0	-	- عینی
دستگیری شیخ عمر بر جماعت املاک گران آمد	0	-	-
این سخنان، برای رئیس گران آمده بود	0	-	0
این سخن برای میزبان گران آمد	0	-	-
این کار برای اروپایی‌ها گران آمد	0	+ انسان	-
	0	0	0



مرکز مهندسی اطلاع‌رسانی علوم و فناوری

سامانه معنایی افعال زبان فارسی

جستجو



شایان ذکر است که برای نمایش بهتر مطالب، لازم است از قلم B Mitra استفاده شود.

- اسلامی، محرم، شیخ‌زادگان، جواد، احمدی‌نیا، زهرا، و بهرامی‌راد، علی. (۱۳۸۸). مراحل و نحوه‌ی تهیه‌ی دادگان‌های صوتی هجایی و دایفونی برای سامانه‌ی تبدیل متن به گفتار فارسی. *دوفصل‌نامه‌ی علمی-پژوهشی پردازش علائم و داده‌ها*، (۱۲)، ۳-۱۲.
- بی‌جن‌خان، محمود. (۱۳۸۳). نقش پیکره‌های زبانی در نوشتن دستور زبان: معرفی یک نرم‌افزار رایانه‌ای. *مجله‌ی زبان‌شناسی*، ۱۹(۲)، ۴۸-۶۷.
- (۱۳۹۵). پیکره‌ی گفتار محاوره‌ای زبان فارسی امروز. *مجموعه مقالات دومین همایش ملی زبان‌شناسی پیکره‌ای*. تهران: نشر نویسه پارسی.
- دانشکار آراسته، پویان. (۱۳۸۱). نرم‌افزار تشخیص فعل در زبان فارسی. *پایان‌نامه‌ی کارشناسی ارشد*، تهران: دانشگاه علامه طباطبایی.
- راسخ مهند، محمد. (۱۳۹۱). «گزاره‌های غیرفعلی در زبان فارسی». *نشریه‌ی پژوهش‌های زبان‌شناسی تطبیقی*، ۱(۱)، ۲۵-۴۲.
- رسولی، محمد صادق. (۱۳۹۱). *فرهنگ ظرفیت نحوی افعال زبان فارسی*. دبیرخانه‌ی شورای عالی اطلاع‌رسانی. قابل دسترس در: www.dadegan.ir/catalog/pervallex
- رفیعی، عادل. (۱۳۸۶). نگاهی به عملکرد معنایی پسوندهای اشتقاقی زبان فارسی. *رساله‌ی دکتری در رشته‌ی زبان‌شناسی همگانی*. تهران: دانشگاه علامه طباطبایی.
- سمایی، سید مهدی. (۱۳۷۹). «پردازش فعل مرکب». *مجله‌ی زبان‌شناسی*. پاییز و زمستان ۱۳۷۹. سال ۱۵. شماره‌ی ۲.
- صفری، سعید. (۱۳۹۶). پیکره‌ی زبان‌آموز «سلام فارسی»: معرفی الگوی دسته‌بندی و تعیین خطاها، مجموعه‌ی برچسب و ابزار برچسب‌دهی، نخستین کنفرانس ملی پژوهش‌های کاربردی در زبان‌شناسی رایانشی (با محوریت خط و زبان فارسی)، شیراز: مرکز منطقه‌ای اطلاع‌رسانی علوم و فن‌آوری. صص ۲۰۸-۱۹۷.
- طیب زاده، امید. (۱۳۸۵). *ظرفیت فعل و ساخت‌های بنیادین جمله در فارسی امروز، پژوهشی بر اساس نظریه‌ی دستور وابستگی*. تهران: نشر مرکز.
- فرخ، ماندانا. (۱۳۸۱). بررسی ساختمان افعال ساده و مرکب فارسی و تدوین روش‌های سرواژه‌سازی به کمک رایانه. *پایان‌نامه‌ی کارشناسی ارشد*، تهران: دانشگاه تهران.
- قربان‌زاده، فرهاد. (۱۳۹۴). معرفی پیکره‌ی فارسی روز. *مجموعه مقالات نخستین همایش ملی زبان‌شناسی پیکره‌ای*. تهران: نشر نویسه پارسی.
- کرد زعفرانلو، عالیه و عبدالکریمی، سپیده و گلغام، ارسلان. (۱۳۹۰). «مجهول‌سازی افعال مرکب فارسی از منظر معنایی و نظریه‌ی معنی‌شناسی مفهومی». *فصل‌نامه‌ی زبان و ادبیات تطبیقی*. دوره‌ی ۲. شماره ۲. صص ۱۷-۱.
- منصوری‌زاده، محرم و نصیری، محمد. (۱۳۹۱). «روابط ساختاری در وردنت فاوا (فن‌آوری اطلاعات و ارتباطات)». *گزارش طرح پژوهشی در پژوهشکده‌ی فن‌آوری اطلاعات*. قابل دسترسی در: www.vlib.itrc.ac.ir
- مولودی، امیر سعید و کوهستانی، منوچهر و رسولی، محمد صادق. (۱۳۹۰). «فرهنگ ظرفیتی فعل: تلاشی در

جهت آموزش زبان فارسی به غیرفارسی زبانان». مجموعه مقالات نخستین همایش بین‌المللی چشم‌انداز، راهکارها و موانع گسترش زبان و ادب فارسی. تهران.

محمد ابراهیمی، زینب و فرقانی‌فرد، مرضیه. (۱۳۸۹). «آموزش فعل مرکب به غیرفارسی‌زبانان». فصل‌نامه‌ی پژوهش زبان و ادب فارسی. شماره‌ی ۱۶. بهار ۱۳۸۹. صص ۱۵۴-۱۳۵.

میرزایی، آزاده و مولودی، امیر سعید. (۱۳۹۳). «نخستین پیکره‌ی نقش‌های معنایی زبان فارسی». علم زبان. پاییز و زمستان ۱۳۹۳. ۲ (۳)، ۴۷-۲۹.

پیکره‌ی به‌روزشونده‌ی پارسی دسترسی از www.peikare.ir

AleAhmad, A., Amiri, H., Darrudi, E., Rahgozar, M., & Oroumchian, F. (2009) Hamshahri: A Standard Persian Text Collection. *Knowledge-Based Systems*, 22(5), 382–387.

Assi, S. M. (1997). Farsi linguistic database (FLDB). *International Journal of Lexicography*, 10(3), 5.

Baker, C. F., Fillmore, C. J., and Lowe, J. B. (1998). “The Berkeley FrameNet project”. *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics*. Volume 1 (pp. 86-90). Association for Computational Linguistics.

Barari, L., & QasemiZadeh, B. (2005). CloniZER spell checker adaptive language independent spell checker. In *AIML 2005 Conference CICC, Cairo, Egypt*, 65-71.

Bi Jen Khan, M., Sheykhzadegan, J., Bahrani, M., and Ghayoomi, M., (2011). “Lessons from building a Persian written corpus: Peykare”. *Language Resources and Evaluation*. 45, no. 2: 164-143.

Burchardt, A., Erk, K., Frank, A., Kowalski, A., Padó, S., and Pinkal, M. (2006). “The SALSA corpus: a German corpus resource for lexical semantics”. *Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC)*.

Eghbalzadeh, Hosseini, B., Khadivi, S., and Khodabakhsh, A. (2012). Persica: A Persian Corpus for Multipurpose Text Mining and Natural Language Processing. In *Sixth International Symposium on Telecommunications (IST)*. IEEE. Tehran.

Elahimanesh, M. H., Minaei-Bidgoli, B., Gholami, M. J., and Juzi, H. (2012). An Introduction to Noor Corpus and its Language Model. In *First International Conference on Persian Language Processing (ICPLP)*, Semnan, Iran.

- Ganiev, F.A. (1984). Semantic Classes of Tatar Verbs. In *Studies on Tatar linguistics*. Kazan, pp. 75-84. (In Russian).
- Ghayoomi, M., (2012). Bootstrapping the Development of an HPSG-based Treebank for Persian. In *Linguistic Issues in Language Technology*, 7 (1), 10-13.
- Granger, S., Gilquin, G., & Meunier, F. (2015). *The Cambridge Handbook of Learner Corpus Research*: Cambridge University Press.
- Khalifehsoltani, S. N., Cholmaghani, A., Vahdani, A., & Moallemi, R. (2010). Building a large Persian Verb Collection: A generative approach. In *The 2nd International Conference on Computer Engineering and Technology. (ICCET)*, IEEE.
- Kennedy, Graeme (1998). *An Introduction to CorPus Linguistics*, London: Longman.
- Kipper, K., Korhonen A., Ryant N. & Palmer M. (2006). Extending VerbNet with Novel Verb Classes. In *Proceedings of the Fifth International Conference on Language Resources and Evaluation -- LREC'06*. May, 2006, Genoa, Italy: 2006.
- Kroeger, P. R. (2005). *Analysing grammar, an introduction*. Cambridge: Cambridge University Press.
- Kustova G., Lashevskaja O., Paducheva E.& Rakhilina E. (2009), Verb Taxonomy: from Theoretical Lexical Semantics to Practice of Corpus Tagging. In Lewandowska B., K. Dziwirek(eds.), *Cognitive Corpus Linguistics Studies*. Frankfurt: Peter Lang.
- Matveeva T. (ed.) (1986). *Lexical Semantic Groups of Russian Verbs (an Academic Dictionary)*. Sverdlovsk. (In Russian).
- McEnery, T., & Wilson, A. (2001). *Corpus Linguistics: An Introduction*: Edinburgh University Press
- Nourian, A., Rasooli, M. S., Imany, M. and Faili, H. (2015). On the Importance of Ezafe Construction in Persian Parsing. Accepted in *the 53rd Annual Meeting of the Association for Computational Linguistics (ACL 2015)*, Beijing, China
- Orazov M. (1983). *Semantics of the Kazakh Verb (an Experience of Semantic Classification)*. Alma-Ata. (In Russian).

- Palmer, M., Kingsbury, P., and Gildea, D (2005). "The Proposition Bank: An Annotated Corpus of Semantic Roles". *Computational Linguistics*. 31 (1). pp. 71–106
- Rafiei, A. & Azhari, S. (2013). Selectional restriction on the derivational Persian suffix / -ane/. *International Journal of linguistics*. Vol. 15. No. 3.
- Rasooli, M. S., Moloodi, A. S., Kouhestani, M., Minaei Bidgoli, Behrouz. (2011). A Syntactic Valency Lexicon for Persian Verbs: The First Steps towards Persian Dependency Treebank. In *5th Language & Technology Conference (LTC): Human Language Technologies as a Challenge for Computer Science and Linguistics, Poznań, Poland*, 227–231.
- Rasooli, M. S. Kouhestani, M. and Moloodi, A. S. (2013). Development of a Persian Syntactic Dependency Treebank. In *The 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL HLT)*, Atlanta, USA.
- Samvelian, P. and Faghiri, P. (2013). Introducing PersPred, a syntactic and semantic database for Persian complex predicates. In *Proceedings of the 9th Workshop on Multiword Expressions, Atlanta, Georgia, USA. Association for Computational Linguistics*. 11-20.
- Shamsfard, M., Hesabi, A., Fadaei, H., Mansoory, N., Famian, A., Bagherbeigi, S., Fekri, E. and et al. (2010). Semiautomatic Development of Farsnet; the Persian Wordnet. *Proceedings of 5th Global WordNet Conference (GWA2010). Mumbai, India*.
- Schwartz, H. A. & Gomez, F. & Millward, C. (2008). A semantic feature for verbal predicate and semantic role labeling using SVMs. Association for advancement of artificial intelligence. Available at: www.Aaai.org.
- Teubert, W. (1999). Corpus Linguistics: A Partisan View, *International Journal of Corpus Linguistics*, 4(1), 1-10.