

استخراج روابط معنایی میان فعل و وابسته‌های آن از متون زبان فارسی

مهرنوش شمس‌فرد^۱

فاطمه جعفری‌نژاد^۲

چکیده

به دست آوردن روابط معنایی میان افعال و دیگر اجزای سازنده جمله برای پردازش معنایی جمله کاربرد بسیاری دارد. به علاوه اطلاع از محدودیت‌های گزینشی که فعل به وابسته‌های خود اعمال می‌کند نیز در پردازش معنایی کاربرد دارد. علی‌رغم اینکه تلاش برای این کار در زبان‌های مختلف در حال پیگیری است، فراهم ساختن چنین اطلاعاتی برای افعال به صورت دستی مستلزم صرف هزینه‌های انسانی و زمانی است. در نتیجه خودکارسازی این روند بسیار با اهمیت و مورد توجه پژوهشگران است. در این مقاله سه روش برای استخراج این روابط معنایی ارائه می‌شود. روش مبتنی بر ریخت‌شناسی و تحلیل‌های لغوی به صورت ساده شده‌ای به حل مسئله می‌پردازد. روش مبتنی بر تعمیم، با بررسی آماری وابسته‌های افعال به محدودیت‌های گزینشی دست می‌یابد؛ و در روش مبتنی بر قاعده و تعمیم، برچسب‌زنی نقش‌های معنایی و یافتن محدودیت‌های گزینشی افعال به یافتن وابسته‌ها منجر می‌شود. در انتها این روش‌ها با هم مقایسه و مزایا و معایب هر یک بررسی می‌شود.

واژه‌های کلیدی: پردازش زبان طبیعی، تحلیل معنایی سطحی، استخراج نقش‌های موضوعی، استخراج محدودیت‌های گزینشی افعال و وابسته‌ها

۱- مقدمه

در توصیف رخدادهای جهان واقع با استفاده از زبان طبیعی و در قالب جمله، عموماً افعال، نقش اساسی دارند و سایر ارکان جمله به عنوان وابسته‌های معنایی فعل در توصیف بیشتر ابعاد رخداد استفاده می‌شوند. از این رو شناسایی روابط میان افعال و وابسته‌های آن در درک معنایی متون بسیار کارآمد است. چالش اساسی در درک متن، کشف زیرساخت واحد از روی روساخت‌های مختلف توصیف

^۱ - استادیار، دانشکده مهندسی برق و کامپیوتر، دانشگاه شهیدبهشتی m-shams@sbu.ac.ir

^۲ - دانشجوی دکتری، دانشکده مهندسی کامپیوتر و فن‌آوری اطلاعات، دانشگاه صنعتی شاهرود

jafarinejad@shahroodut.ac.ir

³ - Argument

یک رخداد واحد است. شناسایی وابسته‌های فعل و نوع ارتباط معنایی آنها در جمله راهکار بسیار پرکاربردی در این حیطه است که با تحلیل معنایی سطحی^۱ شناخته می‌شود.

روش‌های برچسب‌زنی نقش معنایی در یافتن روابط معنایی بین فعل و وابسته‌های آن بسیار مناسب است. در برچسب زنی نقش معنایی با داشتن قاب معنایی و مجموعه ملزومات معنایی یک معنای فعل، نقش هر یک از گروه‌های غیر فعلی موجود در جمله تعیین می‌شود (مارکز^۲ و کاریراس^۳، ۲۰۰۸). رویکردهای مطرح شده در این حیطه را می‌توان در دو دسته کلی تقسیم کرد. در روش‌های مبتنی بر قاعده مانند آنچه در لواین^۴ و مویروس (۲۰۰۶) معرفی شده است ابتدا ویژگی‌های مربوط به هر نقش معنایی تحت قوانینی به سیستم معرفی می‌شوند، سپس با توجه به قواعد کلی بیان شده، صدق کردن و صلاحیت داشتن هر یک از گروه‌های ورودی در داشتن هر یک از نقش‌های معنایی بررسی و در صورت دارا بودن، نقش معنایی خاص به آن گروه نگاشته می‌شود. این قواعد کلی یا به صورت صریح در سیستم معرفی می‌شوند و هدایت جریان تشخیص نقش‌ها با قوانین پیش می‌رود و یا به صورت غیر صریح و با استفاده از دانشی است که با روشی مانند جستجوی درخواست‌هایی در اینترنت و منابع گسترده دیگر یا تجزیه با استفاده از گرامر به دست می‌آید.

رویکرد دیگری که در برچسب‌زنی مشاهده می‌شود رویکرد آماری است. بر اساس نظریه اتصال^۵ لوین^۶ و هوواو^۷ (۱۹۹۶) که بیان می‌کند «وابسته‌های نحوی یک جمله، با داشتن وابسته‌های معنایی معنایی آن قابل پیش‌بینی‌اند» و نیز با لحاظ کردن این مسئله که نحو و معنا در جملات به گونه‌ای به هم مرتبط‌اند می‌توان با داشتن نقش‌های معنایی، مجموعه بزرگی از افعال موجود در جملات پیکره معنایی، روشی برای استخراج خودکار نقش‌های معنایی بر اساس روش‌های یادگیری و اطلاعات آماری داشت.

همانطور که بیان شد داشتن مجموعه بزرگ آموزشی اولیه از ملزومات روش‌های یادگیری در این کار است. روند رو به رشد ساخت پیکره‌های معنایی در زبان‌های مختلف، موجب افزایش رغبت به رویکرد یادگیری در برچسب‌زنی نقش‌های معنایی شده است. در این روش‌ها مسئله برچسب‌زنی نقش‌های معنایی به صورت مسئله‌ای دسته‌بندی‌شده در نظر گرفته می‌شود. به عنوان نمونه‌هایی از کارهای برچسب‌زنی انجام شده با این رویکرد می‌توان به لی^۸ و سانگ^۹ (۲۰۰۷) اشاره نمود.

^۱- Shallow semantic parsing

^۲- Marquez, L.

^۳- Carreras, X.

^۴- Levine, R.D.

^۵- Linking theory

^۶- Levin, B.

^۷- Hovav, M.R.

^۸- Lee, J.

^۹- Song, Y.

مسئله مهم دیگری که پس از برچسب‌زنی نقش‌های معنایی یک فعل مطرح است استفاده از مدل مناسبی جهت بیان اطلاعات به دست آمده در قالب روشی ساختارمند است. یکی از نمونه‌های متداول این کار استفاده از هستان‌شناسی^۱ های زبانی است. وارد کردن اطلاعات معنایی میان افعال و وابسته‌های آن در هستان‌شناسی‌ای مانند روئیز کاسادو^۲ و همکاران (۲۰۰۷) منجر به غنای آن هستان‌شناسی خواهد شد. این مسئله یعنی بهبود هستان‌شناسی با افزودن روابط معنایی میان فعل و وابسته‌ها گاه به صورت مسئله یافتن محدودیت‌های گزینشی یا ترجیحات گزینشی نیز مورد بحث قرار می‌گیرد (مانند اسپور^۳ (۲۰۰۸) و وونش^۴ و همکاران (۲۰۱۰)).

در این مقاله روش‌هایی برای برچسب‌زنی نقش‌های موضوعی در جملات فارسی، استخراج مفاهیم متناظر با وابسته‌های موضوعی افعال و همچنین محدودیت‌های گزینشی آنها و غنی‌سازی وردنت فارسی با این اطلاعات پیشنهاد شده است. عمده تفاوت روش‌های ارائه شده در این مقاله با پژوهش‌های انجام شده در زمینه برچسب‌زنی نقش معنایی در زبان فارسی مانند صدر موسوی (۱۳۸۶)، کامل قالیباف و همکاران (۲۰۰۹) و سعیدی و فیلی (۲۰۱۲) و حتی در زبان‌های دیگر، در این است که در پژوهش‌های دیگر برای برچسب‌زنی وجود یک پایگاه داده با ساختار وابستگی^۵ افعال، نقشی بسیار محوری و کلیدی داشته است به گونه‌ای که با نبودن چنین پایگاه داده‌ای کار برچسب‌زنی معوق می‌ماند و روش‌های دیگر روی تعداد محدودی از افعال که پایگاه داده برای آنها ساخته شده است نتایج را اعلام کرده‌اند. اما در این پژوهش سعی بر آن بود تا با حذف محوریت این پایگاه و استفاده از آن فقط در قسمت‌هایی که اطلاعات آن در دسترس است سیر جدیدی ارائه شود. همچنین در پژوهش‌های دیگر در زبان فارسی عمدتاً شرط ساده بودن افعال مطرح است اما در این روش با ارائه راه حلی برای مسئله یافتن مرز فعل، پوشش افعال مرکب نیز میسر شده است.

همچنین در روش‌های دیگر مانند کامل قالیباف و همکاران (۲۰۰۹) و سعیدی و فیلی (۲۰۱۲) برخی از مقادیر بردار ویژگی‌های لغوی و نحوی و معنایی برای فعلی که نقش‌های آن در جمله مورد بررسی است به صورت دستی مقداردهی شده‌اند اما در روش معرفی شده در این مقاله به دست آوردن ویژگی‌های لازم برای تشخیص نقش به صورت خودکار انجام می‌شود که این کار نشان دهنده افزایش سطح خودکارسازی این روش نسبت به روش‌های قبلی است.

^۱- ontology

^۲- Ruiz-Casado, M.

^۳- Spohr, D.

^۴- Wunsch, H.

^۵- Argument Structure

۲- مجموعه نقش‌های معنایی

نکته مهم در مسائل برچسب‌زنی انتخاب مجموعه نقش‌ها است. در این مقاله از میان مجموعه نقش‌های پذیرفته شده در منابع مختلف، مجموعه نقش‌های معرفی شده در صدر موسوی (۱۳۸۶): (۲۲-۱۵) ملاک عمل است:

- عامل: انجام دهنده کار و یا عامل مختار وقوع رویدادی که توسط گزاره مشخص می‌شود. مانند «علی» در جمله «علی به خیابان رفت.» یا «پرنده» در جمله «پرنده پرواز کرد».
- نیرو: موجودیتی که از جنس نیرو یا انرژی است و باعث وقوع یک اتفاق یا تغییر حالت گردد. مانند «سیل» در جمله «سیل خانه‌ها را ویران کرد».
- تجربه‌گر: موجودیتی که از جنس نیرو نیست و اختیاری برای انجام فعل نداشته باشد اما فعل از جانب او صادر شده باشد. مانند «دوست بابک» در جمله «دوست بابک در تصادف مرد» یا «برگ درخت» در جمله «برگ درخت به زمین افتاد».
- همراه: موجود زنده‌ای که در انجام فعل یا وقوع اتفاق، موجود زنده دیگری را همراهی کند. مانند «مادرش» در جمله «زهرا با مادرش کیف را به خانه آورد».
- پذیرا: موجودیتی که از اتفاق یا عمل مشخص شده توسط گزاره تأثیر پذیرفته است. مانند «حسن» در جمله «دیروز حسن را دیدم».
- موضوع: موجودیتی که به وسیله یک اتفاق یا عمل مشخص شده توسط گزاره، دچار تغییرات فیزیکی شده است. مانند «گوشت» در جمله «شیر گوشت را خورد».
- ذینفع: موجود زنده‌ای که از انجام عمل بیان شده توسط فعل به نحوی سود برده است. مانند «من» در جمله «مادرم برای من کیک پخت».
- سبب: اتفاق یا حالتی که خود باعث وقوع اتفاق یا تغییر یک حالت شود. مانند «بازگشت پدر» در جمله «بازگشت پدر مرا هیجان‌زده کرد».
- زمان: زمان وقوع اتفاق یا عمل مشخص شده توسط یک گزاره. مانند «دیروز» در جمله «دوستم دیروز به خانه ما آمد».
- حالت: چگونگی وقوع یک اتفاق یا انجام یک عمل. مانند «با خوشحالی» در جمله «علی با خوشحالی به خانه برگشت».
- مکان: مکان وقوع اتفاق یا عمل مشخص شده توسط یک گزاره. مانند «آسمان» در جمله «هوایما در آسمان منفجر شد».
- مبدأ: محل یا موجودیتی که در طی جریان وقوع فعل، عمل انتقال از سمت آن صورت گرفته است. مانند «مدرسه» در جمله «علی از مدرسه خارج شد». با توجه به تعریف فوق منظور از مکان تنها مکان فیزیکی نیست و در نتیجه منظور از انتقال هم تنها انتقال فیزیکی نیست. مانند «جمع کشورهای متحدالمنافع» در جمله «آذربایجان از جمع کشورهای متحدالمنافع خارج خواهد شد».

– مقصد: محل یا موجودیتی که طی جریان وقوع فعل، عمل انتقال به سمت آن صورت گرفته است. در این تعریف علاوه بر مکان، موجودیت‌های غیر مکانی نیز می‌توانند دارای نقش مقصد باشند. مانند «علی» در جمله «معلم نزدیک علی رفت».

– ابزار: ابزار، وسیله‌ای که در انجام یک اتفاق یا عمل مشخص شده توسط یک گزاره مورد استفاده بوده است. مانند «شیشه» در جمله «او دستش را با شیشه برید».

در یک جمله هر یک از گروه‌های غیرفعلی می‌توانند در تناظر با یک نقش معنایی قرار گرفته باشند. به عنوان مثال در جمله «علی با حسن با خوشحالی به خانه خود بازگشت» «علی» در نقش عامل، «با حسن» در نقش همراه، «با خوشحالی» نقش حالت، «به خانه خود» در نقش مقصد، برای فعل «بازگشت» بودند.

گاهی نقش‌ها را صرف نظر از جزئیات به دسته‌های کمتری تقسیم می‌کنند. مختصرترین این تقسیم‌بندی‌ها در نظر گرفتن نقش‌های معنایی تعمیم‌یافته شبه‌عامل و شبه‌پذیرا است. شبه‌عامل شخص یا نیرو یا ابزاری است که رخداد توصیف شده توسط فعل به وسیله آن انجام گرفته است و شبه‌پذیرا چیزی است که تأثیر رخداد را دریافت نموده است. شبه‌عامل بر طبق معنای فعل می‌تواند به نقش‌های موضوعی مانند عامل، تجربه‌گر، ابزار، ذینفع و نیرو نگاشته شود. شبه‌پذیرا نیز در تناظر با نقش‌های موضوعی مانند پذیرا و موضوع قرار می‌گیرد.

۳- منابع ورودی

در این مقاله سه روش جهت استخراج روابط میان فعل و وابسته‌های آن ارائه شده است. در هر یک از این روش‌ها منبع یا منابعی استفاده شده است که تأمین کننده اطلاعات لغوی، نحوی و یا معنایی مورد نیاز آن روش بودند. شرح مختصری از این منابع ورودی و ویژگی‌های زبانی آنها در زیر آمده است.

– پیکره بیجن‌خان (۲۰۰۴): این پیکره که در آزمایشگاه زبان‌شناسی دانشگاه تهران ساخته شده است، شامل برخی اخبار روزنامه‌ها و متون معمولی است. این پیکره حاوی حدود ۷,۵ میلیون واژه و ۵۵۰ برچسب است. هر برچسب در این مجموعه از یک ساختار سلسله‌مراتبی پیروی می‌کند. بخش‌هایی از نام برچسب که در ابتدای نام آن قرار دارند بیان‌کننده توصیف کلی‌تری از آن برچسب و بخش‌هایی که در انتهای نام برچسب قرار دارند بیان‌کننده توصیف جزئی‌تر در مورد آن برچسب است. در این مقاله از زیرمجموعه‌ای تصادفی شامل ۲ میلیون کلمه از این پیکره استفاده شده است.

– فارس‌نت (شمس‌فرد و همکاران، ۲۰۱۰): فارس‌نت واژه‌ستان‌شناسی عمومی زبان فارسی است که در قالبی مشابه با «وردنت» زبان‌های دیگر توسعه یافته است. واژه‌ستان‌شناسی فارس‌نت به صورت پایگاهی از کلمات، معانی، مفاهیم، روابط لغوی میان کلمات و روابط معنایی میان مفاهیم است. هر کلمه می‌تواند یک یا چند معنی متفاوت داشته باشد. کلمات با معنای یکسان تحت یک مجموعه

هم‌معنا قرار می‌گیرند و یک مفهوم را می‌سازند. کلمات تحت مقوله‌های نحوی اسم، فعل، صفت و قید قرار دارند و روابط میان مفاهیم و کلمات می‌تواند روابطی میان مقوله‌ای یا درون‌مقوله‌ای باشد. در این مقاله از فارسن‌نت ۲ استفاده شده است.

– پایگاه داده افعال: این پایگاه داده مشتمل بر اطلاعات نحوی حدود ۳۷۰۰ فعل در زبان فارسی است که در آزمایشگاه پردازش زبان طبیعی دانشگاه شهید بهشتی تولید شده است. در این مقاله از اطلاعات نحوی موجود در این پایگاه داده شامل بن ماضی و مضارع هر یک از مصادر فعلی و لازم یا متعدی بودن آنها جهت تشخیص افعال مرکب موجود در پیکره و یافتن نقش‌های معنایی تعمیم یافته این افعال در سطح جمله استفاده شده است.

– پایگاه داده ساختار وابستگی افعال: این پایگاه، افعال پررخداد زبان فارسی را به همراه شماره مجموعه هم‌معنای آنها را در فارسن‌نت نگهداری می‌کند و برای هر معنای فعلی با آوردن جملات شاهد از پیکره بیجن‌خان و مشخص کردن وابسته‌های نحوی و معنایی مربوط به فعل در جمله، محدودیت‌های گزینشی این افعال را بیان می‌کند. این پایگاه داده در آزمایشگاه پردازش زبان طبیعی دانشگاه شهید بهشتی تولید و آماده بهره‌برداری شده است (شمس‌فرد و همکاران، ۱۳۹۱). در روند این مقاله از اطلاعات افعال موجود در این پایگاه داده جهت تفکیک و رسیدن به نقش‌های موضوعی از اطلاعات به دست آمده از روی نقش‌های معنایی تعمیم‌یافته استفاده شده است.

– واژگان زایا: نسخه الکترونیکی واژگان فارسی زایا (اسلامی و همکاران، ۱۳۸۳) به صورت پایگاه داده مشتمل بر حدود ۵۶۰۰۰ لغت زبان فارسی است. صورت واجی، صورت نوشتاری، کد مقوله نحوی، بسامد رخداد و الگوی تکیه اطلاعاتی‌اند که برای هر لغت در این پایگاه داده در نظر گرفته شده‌اند. در این مقاله از این فرهنگ جهت داشتن مجموعه تقریباً کاملی از واژگان زبان فارسی در تحلیل لغوی و استفاده از روابط اشتقاقی جهت یافتن روابط معنایی میان فعل و وابسته‌های آن استفاده شده است.

۴- پیمانه‌های کمکی

در قسمت‌هایی از این روش از پیمانه‌های کمکی موجود در زبان فارسی استفاده شده است. یکی از این پیمانه‌ها بخش‌بند معرفی شده در نوفرستی و شمس‌فرد (۲۰۱۱) است که مرز گروه‌های مختلف موجود در جمله را تعیین می‌کند. این پیمانه با استفاده از تلفیقی از روش‌های مبتنی بر قاعده و یادگیری ماشینی قادر به تعیین مرز گروه‌های اسمی، فعلی و حرف اضافه‌ای با دقت ۸۷٫۷ درصد است. در این روش همه کلمات ورودی برچسب‌های IOB می‌گیرند که نشان می‌دهند هر کلمه آغازگر گروه است یا درون آن قرار دارد و یا خارج از گروه‌های هدف است.

پیمانه‌های مورد استفاده دیگر ریشه‌یاب و قطعه‌بند استپ-وان است (شمس‌فرد و همکاران، ۲۰۱۰). قطعه‌بند برای تشخیص مرز کلمات و عبارات و ریشه‌یاب برای یافتن ریشه کلمات

موجود در پیکره استفاده شده‌اند. ریشه‌یاب با استفاده از قواعد و بهره‌مندی از واژگان زایا قادر به بازشناسی وندهای تصریفی و همچنین وندهای اشتقاقی موجود در یک کلمه با دقتی بیش از ۹۰ درصد است.

پیمانه بعدی مورد استفاده پیمانه رفع ابهام معنایی است (حق‌الهی و شمس‌فرد، ۲۰۱۱). این پیمانه برای تشخیص معنای فعل در جملات به منظور حرکت از سطح کلمه (در متن) به سطح مفهوم (در سلسله مراتب فارس‌نت) به کار می‌رود. تشخیص معنای فعل در جمله، در استخراج روابط فعل و وابسته‌ها با استفاده از رویکرد مبتنی بر قاعده و تعمیم از پیکره مورد استفاده این مقاله است. دقت این پیمانه در صورت نداشتن فاز آموزش قبلی ۴۰ درصد و در صورت داشتن فاز آموزش ۵۰ درصد گزارش شده است.

۵- روش‌های پیشنهادی

در این مقاله برای استخراج روابط معنایی بین فعل و وابسته‌ها سه روش مجزا پیشنهاد شده است. در هر یک از این روش‌ها با توجه به ورودی‌ها و منابع استفاده شده، رویکردی متفاوت در استخراج روابط اسم- فعل در بین مجموعه‌های هم‌معنای فارس‌نت اتخاذ شده است. در ادامه به ارائه توضیحی مختصر از هر یک از این روش‌ها خواهیم پرداخت.

۵-۱- روش اول: روش مبتنی بر ریخت‌شناسی با استفاده از فرهنگ واژگانی زبان فارسی

در روش اول بیشتر تمرکز بر یافتن واژه‌هایی است که ارتباطشان با فعل از نوع نقش‌های معنایی است. این روش برای تجهیز وردنت فارسی به روابط میان مقوله‌ای فعل و اسم به کار می‌رود و منبع ورودی آن واژگان فارسی است. در این روش به روابط اشتقاقی میان فعل و وابسته‌های آن توجه می‌شود. مشتقاتی مانند اسم فاعل و اسم مفعول نقش شبه عامل یا شبه پذیرا را بپذیرند. به عنوان مثال رابطه معنایی موضوع که بین مفهوم فعلی (خوردن، میل کردن، صرف کردن) و مفهوم اسمی (خوردنی، خوراکی، خوراک، ماده مغذی) وجود دارد، با استفاده از بررسی‌های ریخت‌شناسی^۱ قابل بازیابی است. در تحلیل الگوهای اشتقاقی در این بخش تمرکز بر ساخت صفت فاعلی و صفت مفعولی است. در آزمایش‌ها استفاده از ساخت صفت فاعلی و اسم فاعل برای استخراج روابط معنایی عامل و تجربه‌گر میان فعل و اسم به دقت مناسبی نرسید و از فهرست الگوهای کاندیدا حذف شد ولی مشاهده شد که استفاده از الگوهای ساخت، صفت مفعولی و اسم مفعول برای استخراج روابط معنایی موضوع و پذیرا از دقت نسبی مناسبی برخوردار است. این الگوها عبارتند از:

^۱- morphology

مصدر+ی| بن ماضی+ه| بن مضارع + اک

به این ترتیب در این روش با تحلیل‌های ریخت‌شناسی روابط بین یک فعل و وابسته‌های آن به دست می‌آیند. سپس کلمات مرتبط به مفاهیم متناظرشان در سلسله مراتب فارس نت نگاشته می‌شوند تا به این ترتیب رابطه فعل و وابسته از سطح کلمه به سطح مفهوم ارتقاء یابد. منابع ورودی این روش سلسله مراتب فارس‌نت، پایگاه داده افعال حاوی بن ماضی و مضارع و مصدر آنها و پایگاه داده مربوط به ساختار وابستگی افعال است.

۵-۲- روش دوم: روش مبتنی بر تعمیم با استفاده از پایگاه داده ساختار

وابستگی افعال

این روش از اطلاعات وارد شده توسط زبان‌شناس در مورد افعال موجود در جملات شاهد در پایگاه داده ساختار وابستگی افعال استفاده می‌کند و با تعمیم^۱ اطلاعات و بهره‌گیری از فارس‌نت به عنوان ساختار سلسله مراتبی، محدودیت گزینشی هر یک از وابسته‌های معنایی آن مفهوم فعلی را به صورت رابطه معنایی بین مفهوم فعلی و مفاهیم اسمی بیان می‌کند. به عنوان مثال با مشاهده جملات «سیب را خورد»، «او گلابی می‌خورد»، «شیر خورده شد»، «ماست خورده می‌شود» و با داشتن اینکه در هر یک از این جملات به ترتیب سیب، گلابی، شیر و ماست تحت رابطه معنایی موضوع برای معنای فعلی «خوردن، صرف کردن، میل کردن» قرار داشته‌اند، با بهره‌گیری از سلسله مراتب فارس‌نت و دانستن اینکه سیب، گلابی، شیر و ماست همگی زیرنوعی از (خوراک، خوراکی، خوردنی، ماده مغذی)‌اند، این مفهوم (خوراک، خوراکی، خوردنی، ماده مغذی) را به عنوان وابسته معنایی «موضوع» برای فعل خوردن به معنای «میل کردن، صرف کردن» تشخیص می‌دهد. در این روش مجموعه هم‌معنایی انتخاب خواهد شد که دارای دو خصوصیت زیر باشد:

- اول اینکه نماینده خوبی برای همه شاهد‌های موجود باشد؛ یعنی حاوی همه اطلاعاتی باشد که مجموعه گروه‌های اسمی موجود در شواهد در رابطه با مفهوم فعلی داشته‌اند. این شرط را «شرط جامعیت» می‌نامیم.

- دوم اینکه بهترین نماینده در سلسله مراتب باشد یعنی در سلسله مراتب فارس‌نت پایین‌تر از سایر کاندیدها باشد. به عنوان مثال هم «خوراک، خوراکی، خوردنی، ماده مغذی» و هم «وجود، هستینه» شامل همه شواهد سیب و گلابی، شیر و ماست در مثال بالا هستند، اما «خوراک، خوراکی، خوردنی، ماده مغذی» به دلیل اینکه در سطح پایین‌تری قرار دارد نماینده بهتری برای این چهار شاهد است. این شرط را «شرط مانعیت» می‌نامیم.

الگوریتمی که برای تعمیم استفاده شد به این ترتیب عمل می‌کند که وقتی کلمه‌ای به عنوان پرکننده یک نقش معنایی برای معنایی از فعلی انتخاب می‌شود. برای هر معنی آن کلمه عدد

¹- generalization

احتمالی برابر عکس تعداد معانی کلمه در نظر گرفته می‌شود. مثلاً اگر کلمه سه معنی مختلف داشته باشد (یعنی در سه مجموعه هم معنا عضو باشد) به هر معنی و در نتیجه به مجموعه هم معنای دربرگیرنده آن احتمال $\frac{1}{3}$ نسبت داده می‌شود. این احتمال در مراحل بعد، از هر یک از دسته‌های هم‌معنای متناسب با هر یک از معانی متعدد کلمه به سطوح بالاتر در سلسله مراتب انتشار می‌یابد و امتیاز سطوح بالایی را می‌سازد.

بدین ترتیب دسته‌های هم‌معنایی که در سطوح بالاتر سلسله مراتب قرار دارند از مسیرهای مختلف و از طریق کلمات زیرشمول خود اعدادی دریافت می‌کنند که ترکیبشان امتیاز آنها را می‌سازد. هرچه این امتیاز بیشتر باشد احتمال اینکه مفهوم جاری متناسب‌ترین مفهوم برای پرکردن نقش معنایی مورد نظر برای فعل باشد، بیشتر می‌شود. به عبارت دیگر این امتیاز معیاری برای تشخیص میزان جامعیت هر یک از دسته‌ها است.

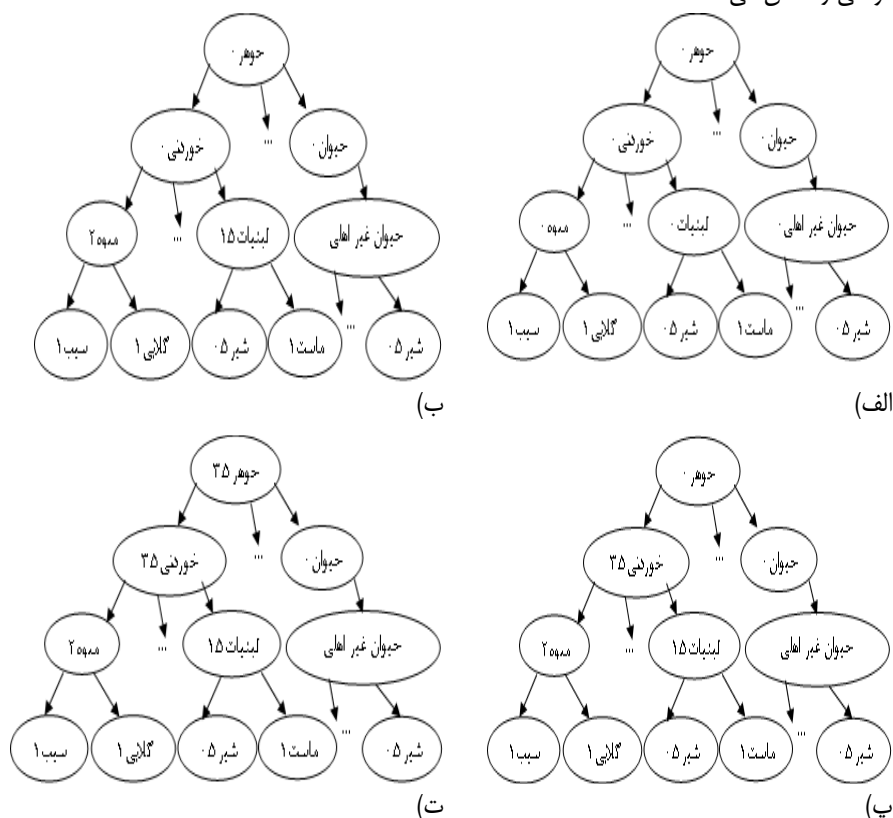
هر مفهوم سطح بالاتر نیز امتیاز خود را به سطوح بالاتر در سلسله مراتب انتقال می‌دهد تا به بالاترین سطح برسد. البته در مورد مفاهیم سطوح بالاتر شرط انتشار امتیاز به پدر خود در سلسله مراتب این است که امتیاز خود آنها از یک بیشتر باشد. این شرط جلوی انتشار امتیاز از مفاهیم نامرتب را می‌گیرد.

امتیاز مفهومی که دارای چند شامل (پدر) است نیز در هنگام انتشار به سطح بالاتر بین پدرهای مفهوم تقسیم می‌شوند. این کار باعث می‌شود که از افزایش بی‌رویه امتیاز در گره‌ای که اکثر فرزندان تنها پدر یک مفهوم بوده‌اند جلوگیری شود.

دلیل تقسیم کردن امتیاز یک کلمه مشاهده شده بین معانی متعدد آن نیز داشتن مکانیزمی جهت تشخیص معانی نامرتب است. مثلاً کلمه شیر چند معنی متفاوت دارد. برای سادگی فرض کنیم این کلمه تنها دو معنی متفاوت دارد که یکی از آنها گونه‌ای از لبنیات و دیگری گونه‌ای از حیوانات است. هر یک از این معانی به مفاهیم شامل خود امتیاز ۰.۵ را انتشار می‌دهد. معنای مرتبط، از هم‌شمول‌های دیگر معنای شیر هم شواهدی داشته و امتیازاتی دریافت نموده است اما معنای نامرتب این شرط را نداشته است. بدین ترتیب معنای نامرتب شرط حداقل امتیاز ۱ را برای انتشار امتیاز به مفهوم شامل خود نخواهد داشت و از انتشار باز می‌ماند و تنها معنای مرتبط در شناسایی درست مفهوم تعمیم‌یافته دخیل می‌شود.

در نهایت از بین مفاهیمی که بالاترین امتیاز را به دست آورده‌اند، دسته‌ای که در سطوح پایین‌تری از سلسله مراتب فارسی قرار دارد به عنوان بهترین دسته معرفی می‌شود. دلیل این امر نیز آن است که نسبت به دسته‌های هم‌معنای هم‌امتیازی که در سطوح بالاتری از سلسله مراتب قرار دارند شرط مانعیت را بهتر برآورده می‌کند؛ زیرا هر چه در سلسله مراتب فارسی بالاتر رویم از مفاهیم جزئی به سمت مفاهیم کلی‌تر پیش می‌رویم و علاوه بر گروه‌های اسمی شاهد، گروه‌های اسمی دیگری که ممکن است مرتبط با مفهوم فعلی نباشد اما ذیل دسته معنایی قرار دارند را نیز شامل

شود. به عنوان مثال فرض کنیم مفهوم فعلی(خوردن، صرف کردن، میل کردن) به عنوان وابسته موضوع این شواهد را دیده باشد: سیب، ماست، شیر، گلابی. نحوه به دست آمدن بهترین مفهوم متناظر با این نقش در شکل ۱ نمایش داده شده است. شکل ۱ نحوه انتشار امتیاز در سلسله مراتب فرضی را نشان می‌دهد.



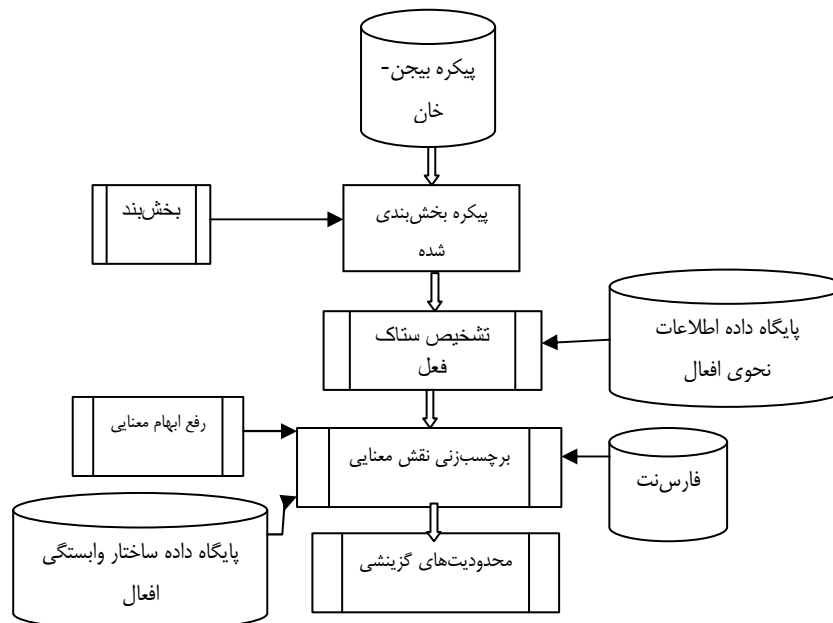
الف) امتیاز موجودیت‌های هستان‌شناسی با داشتن شواهد سیب، ماست، شیر و گلابی به عنوان وابسته موضوع مفهوم فعلی(خوردن، صرف کردن، میل کردن) در سطح اول. ب) انتشار امتیاز موجودیت‌های هستان‌شناسی به سطح دوم. پ) انتشار امتیاز موجودیت‌های هستان‌شناسی به سطح سوم. ت) انتشار امتیاز موجودیت‌های هستان‌شناسی به بالاترین سطح.

بدین ترتیب از بین جوهر و خوردنی که بالاترین امتیاز را دارد، خوردنی که در سطح پایین‌تری در سلسله مراتب هستان‌شناسی قرار دارد به عنوان وابسته موضوع مشخص می‌شود. این روش علاوه بر ساده بودن در پیاده‌سازی، این مزیت را داراست که اگر همه مصادیق مفهومی در شواهد دیده نشده باشد هنوز شانس برای تعمیم به سمت آن مفهوم وجود دارد. در تعداد زیاد شواهد برای

بهبود الگوریتم به جای در نظر گرفتن شواهد هر معنای فعلی به طور جداگانه، شواهد هر مفهوم فعلی (کل افعال یک مجموعه هم معنا) را با هم در نظر می‌گیریم. در این حالت فرض می‌کنیم که افعال هم معنی، وابسته‌ها و محدودیت‌های گزینشی مشابهی دارند.

۵-۳- روش سوم: روش مبتنی بر قاعده و تعمیم با استفاده از پیکره متون زبان فارسی

این روش با تحلیل معنایی سطحی جملات پیکره بیجن‌خان، روابط بین یک فعل و وابسته‌های آن را به دست می‌آورد. منابع ورودی این روش پیکره بیجن‌خان، سلسله مراتب فارس‌نت و مجموعه-های هم معنای آن، پایگاه داده مربوط به ساختار وابستگی افعال، پایگاه داده افعال فارسی که شامل بن ماضی و مضارع و گذرا یا ناگذرا بودن هر مصدر فعلی است. نمودار بلوکی این روش در شکل ۲ نمایش داده شده است. این نمودار به صورت مصور پیمانه-های موجود در این روش و عملکرد اصلی آنها، منابع ورودی آنها و اطلاعات خروجی از آنها، پیمانه-های کمکی که از آنها در رسیدن به نتایج مورد نیاز استفاده شده است را نشان می‌دهد.



شکل ۲ نمودار بلوکی روش استخراج روابط میان مقوله‌ای میان افعال و وابسته‌هایشان در پیکره

این روش از سه فاز اصلی تشکیل شده است:

فاز اول: تشخیص نقش‌های معنایی تعمیم‌یافته^۱ برای هر یک از افعال موجود در جملات پیکره بیجن‌خان، با استفاده از بررسی‌های نحوی مبتنی بر قاعده. فاز دوم: رفع ابهام معنایی فعل جمله با استفاده از معناهای برشمرده شده برای آن فعل در فارس‌نت. فاز سوم: جمع کردن اطلاعات شواهد دیده شده از قرارگیری گروه‌های اسمی متفاوت تحت رابطه معنایی مشخص با یک معنای فعلی مشخص و استفاده از آنها در جهت تعمیم. در نهایت در خروجی مجموعه هم‌معنای متناسب با هر یک از وابسته‌های معنایی افعال ارائه می‌شود.

۵-۳-۱- فاز اول: تشخیص نقش‌های معنایی تعمیم‌یافته

منابع ورودی این فاز پیکره بیجن‌خان با برچسب مقوله نحوی و پایگاه داده افعال است که اطلاعات بن ماضی و مضارع و مصدر و گذرا/ناگذر بودن افعال را دارد. پیمانه‌های بخش‌بند و ریشه‌یاب نیز پیمانه‌های کمکی‌اند که در این روش استفاده شده‌اند. خروجی این روش فایل متنی است که در آن برای هر جمله پیکره بیجن‌خان، فعل جمله و هسته گروه‌های غیر فعلی متناظر با نقش‌های معنایی تعمیم یافته متناسب فعل و نیز پیشنهادهایی برای مجموعه‌ای از نقش‌های عمومی موجود در جمله مشخص گشته‌اند.

در این مرحله، ابتدا گروه‌های نحوی مجزای موجود در جملات با برچسب مقوله نحوی پیکره بیجن‌خان توسط پیمانه کمکی بخش‌بند تشخیص داده می‌شوند. مرز فعلی و مصدر فعل یا افعال هر جمله به دست می‌آید. در مرحله بعد برای هر گروه اسمی موجود در جمله، هسته گروه تشخیص داده می‌شود و ریشه هسته توسط پیمانه کمکی ریشه‌یاب تشخیص داده می‌شود. سپس با استفاده از قواعدی که در نظر گرفته می‌شود نقش گروه‌های اسمی که متناسب با شبه‌پذیرا و شبه‌عامل در جمله باشند تشخیص داده می‌شود. با در نظر گرفتن لازم یا متعدی بودن فعل مورد بررسی، مجهول یا معلوم بودن جمله و ساده یا مرکب بودن جمله این قواعد می‌توان فاعل و مفعول جمله را مشخص کرد. در جملات مرکب هم پایه ممکن است فاعل به قرینه جمله پایه حذف شده باشد و تشخیص آن توسط جمله پایه صورت گیرد. در جملات مرکب تودرتو نیز در صورت حذف شدن فاعل یا مفعول، می‌توان در جملات پایه دنبال آنها گشت. با داشتن این قوانین، در جملات معلوم معمولاً فاعل متناظر با شبه‌عامل و مفعول متناظر با شبه‌پذیرا خواهد بود. البته برخی از افعال برخلاف این قاعده‌اند به عنوان مثال، فاعل فعل افتاد -که فعلی لازم است و تنها فاعل می‌پذیرد- نقش معنایی شبه‌پذیرا دارد نه شبه‌عامل. در مورد جملات مجهول نیز نائب فاعل موجود در جمله عموماً نقش شبه‌پذیرا خواهد داشت.

¹- generalized Semantic Roles

گروه‌های غیر اسمی موجود در جمله مثلاً گروه‌های متممی نیز با توجه به نوع حرف اضافه آنها می‌توانند کاندیدی برای برخی از نقش‌های معنایی مانند ذینفع، زمان، مکان، دلیل باشند. گروه‌های قیدی نیز معمولاً کاندیدی برای نقش‌های معنایی زمان، مکان، حالت‌اند.

نقش تعمیم یافته شبه‌عامل، در بررسی‌های بعدی که بررسی‌هایی عموماً معنایی خواهد بود و در فاز بعد انجام می‌شود به یکی از نقش‌های موضوعی نیرو، ذینفع، عامل، تجربه‌گر و یا ابزار نگاشته خواهد شد. نقش معنایی شبه‌پذیرا نیز بسته به اینکه فعل مورد نظر قابلیت ایجاد تغییر فیزیکی داشته یا نه، به یکی از نقش‌های موضوعی پذیرا و موضوع نگاشته می‌شود.

۵-۳-۲- فاز دوم: رفع ابهام معنایی فعل و تشخیص نقش‌های موضوعی

منابع ورودی این فاز واژه‌شناسی فارسی‌ت، پایگاه داده ساختار وابستگی افعال و نتایج حاصل از خروجی فاز قبل‌اند(فایل متنی که در آن برای هر جمله پیکره بیجن‌خان، فعل جمله و گروه‌های غیرفعلی متناظر با نقش‌های معنایی تعمیم یافته متناسب فعل مشخص گشته‌اند). خروجی این فاز نیز به صورت فایل متنی است که در آن برای هر فعل در هر جمله، نقش موضوعی و هسته گروه غیرفعلی متناظر با آن نقش فهرست شده است.

در این فاز ابتدا معنای افعال تشخیص داده شده در جمله، توسط پیمانه کمکی رفع ابهام معنایی مشخص می‌گردد. به عنوان مثال معنای فعل خوردن در دو جمله «او سیب را خورد» و «سر او به دیوار خورد» متفاوت است. سپس برای هر یک از گروه‌های غیر فعلی که در فاز قبل به عنوان کاندیدی برای نقش‌های متفاوت معنایی فعل در نظر گرفته شده بودند، نقش متناسب به دست خواهد آمد. به عنوان مثال در مورد معنای فعلی(خوردن، صرف کردن) شبه‌پذیرا که کاندید نقش‌های موضوعی پذیرا و موضوع بود، موضوع تشخیص داده می‌شود. این کار با استفاده از اطلاعاتی که در پایگاه داده ساختار وابستگی افعال وجود دارد و اینکه معنای فعلی(خوردن، صرف کردن) دارای موضوع و نه پذیرا است تشخیص داده می‌شود.

برای بازشناسی نیرو، ذینفع، عامل، تجربه‌گر و یا ابزار در مورد گروه‌هایی که نقش معنایی تعمیم یافته شبه‌عامل را داشته‌اند، از اطلاعات سلسله مراتب فارسی‌ت و پایگاه داده ساختار وابستگی افعال استفاده می‌شود. به این صورت که هسته گروه اسمی‌ای که نقش «نیرو» را دارد، در سلسله مراتب فارسی‌ت زیرشمولی از «نیرو، انرژی» است. هسته گروه اسمی‌ای که نقش «ذینفع» را دارد، در سلسله مراتب فارسی‌ت زیرشمول «ارگان، سازمان» و یا «شخص، انسان» است و فعل آن نیز باید در ساختار وابستگی خود، «ذینفع» داشته باشد. در حالی که هسته گروه اسمی عامل، «جاندار» است. هسته گروه اسمی‌ای که نقش «ابزار» را دارد، در سلسله مراتب فارسی‌ت زیرشمولی از شیء فیزیکی است. بازشناسی تجربه‌گر به این است که فعل مربوطه در سلسله مراتب فارسی‌ت از افعال حسی «درک کردن، دیدن، دریافتن، فهم کردن، دانستن، پی بردن، متوجه شدن، خواندن، فهمیدن» باشد.

۵-۳-۳- فاز سوم: تعمیم

این قسمت از پژوهش که آخرین و مهم‌ترین قسمت است به یافتن محدودیت‌های گزینشی هر یک از نقش‌های معنایی فعل می‌پردازد. اساس کار این فاز مشابه تعمیمی است که در روش تعمیم مورد استفاده قرار گرفته است ولی به دلیل متفاوت بودن منبع ورودی، نتایج متفاوتی به دست خواهد داد. از آنجا که نقش‌های عمومی مانند مکان، زمان، حالت، مبدأ و مکان به عنوان وابسته‌های خاص فعلی محسوب نمی‌شوند از فهرست نقش‌ها در این مرحله حذف شده‌اند. همچنین با توجه به سطح تمایز موجود در فارس نت برخی نقش‌های دیگر نیز در این مرحله با هم ترکیب شده و نقش‌های عمومی‌تری را ایجاد کرده‌اند. برای مثال نقش‌های عامل، تجربه‌گر، ذینفع به یک نقش کلی عامل نگاشت شده‌اند و نقش‌های پذیرا و موضوع نیز تحت مجموعه یک نقش معنایی موضوع قرار گرفته‌اند. نقش‌های معنایی ابزار و دلیل نیز از جمله نقش‌های انتخابی برای تعمیم بوده‌اند.

۶- آزمون و ارزیابی

روش مبتنی بر ریخت‌شناسی روشی است که تنها با لحاظ کردن برخی الگوهای لغوی بر روی منابع لغوی و بدون در نظر گرفتن معنا به جستجوی روابط معنایی احتمالی می‌پردازد. به همین دلیل با وجودی که برای افعال تک معنایی بسیار مفید است ولی در کل دقت بالا از آن انتظار نمی‌رود. روش مبتنی بر تعمیم که از اطلاعات موجود در پایگاه داده ساختار وابستگی افعال، برای یافتن و شناسایی محدودیت گزینشی افعال استفاده می‌کند روش بسیار مفید و نسبتاً دقیق در یافتن ارتباطات معنایی فعل و وابسته تلقی می‌شود؛ زیرا هم منبع ورودی یک منبع معنایی است و هم الگوریتم استفاده شده در آن با تحلیل‌های معنایی به روابط نهایی دست می‌یابد. به لحاظ دقت، روش سوم یعنی روش مبتنی بر قاعده و تعمیم یک روش بینابینی دو روش پیشین است زیرا در این روش از منابع غیر معنایی استفاده شده است و با لحاظ کردن الگوریتمی مبتنی بر معنا سعی می‌شود به معنا از متن دسترسی یابد. این روش دارای سطح خودکارسازی بیشتر از روش دوم است.

۶-۱- ارزیابی روش مبتنی بر ریخت‌شناسی

دقت این روش برای بررسی لغوی مجموعه‌های پیشنهادی ۶۰٫۵ درصد و در بررسی معنایی با دسته‌های هم‌معنا تنها ۱۸٫۸۶ درصد برآورد شده است. یعنی عمده خطای این مرحله مربوط به ابهام در نگاشت کلمات به مفاهیم است.

از مهم‌ترین نکاتی که در قالب این آزمون مشهود بود، توانایی یافتن روابط بین کلماتی که به نظر نیاز به تحلیل‌های بسیار دارند، تنها با روش‌های ساده است. به عنوان مثال رابطه بین فعل

«صرف کردن» و اسم «ماده مغذی» به سادگی به دست آمده است. در واقع این روش نشان دهنده قابلیت‌های بالقوه‌ای است که به دلیل ایجاد شبکه‌ای معنایی از واژگان، در فارسی نت نهفته است. نمونه‌ای از روابط به دست آمده از این روش در جدول ۱ آمده است.

مفهوم فعلی	مفهوم اسمی	صحت
خوردن، صرف کردن، میل کردن	خوردنی، ماده مغذی، خوراکی، خوراک	ص
خوردن، کتک خوردن	خوردنی، ماده مغذی، خوراکی، خوراک	غ
خوردن، صرف کردن، خوردن، میل کردن	زادراه، قوت لایموت، خوراک، زاد، ره‌توشه	ص
پوشیدن	پوشاک، لباس، پوشیدنی، رخت، جامه	ص
نوشتن، رونویسی کردن، کپی کردن	نامه، دستخط، کاغذ، نوشته، رقع، خط	ص

جدول ۱- نمونه‌هایی از خروجی‌های روش مبتنی بر ریخت شناسی

۶-۲- ارزیابی روش مبتنی بر تعمیم

ارزیابی دستی، دقت این روش را ۵۴,۰۵ درصد با احتساب مجموعه‌های هم‌معنایی که شامل یا زیرشمول مجموعه مورد انتظارند به دست داده است. در مورد احتساب دقت تنها با مجموعه هم-معنای مورد انتظار، دقت ۲۲,۴۸ درصد بوده است. این میزان دقت با ارزیابی دستی خروجی حاصل از این روش بر روی داده‌های ساختار وابستگی افعال زبان‌شناس به دست آمده است که بر روی ۱۹۰ مفهوم فعلی آزمایش شده است

از جمله مهم‌ترین عوامل مؤثر در خطا در این روش می‌توان به لغوی و نه مفهومی بودن گروه-های اسمی تحت روابط معنایی با افعال در پایگاه داده ساختار وابستگی و ناکامل بودن یا کوتاه بودن سلسله مراتب فارسی نت در برخی دامنه‌ها یا برای بعضی مفاهیم، استعاری بودن برخی از جملات شاهد موجود در پایگاه داده ساختار وابستگی و تأثیر کم بودن جملات شاهد برای یک وابسته معنایی از یک معنای فعلی در ایجاد تعمیم ناکافی اشاره کرد. نمونه‌ای از روابط مستخرج از این روش در جدول ۲ آمده است.

مفهوم فعلی	مفهوم اسمی	نوع رابطه	صحت
پوشیدن	پوشش، ستر، ساتر	موضوع	ص
پوشیدن	انسان، بنی آدم، کس، بنی بشر	عامل	ص
تولید کردن، ساختن	چسباندن، متصل گرداندن	ابزار	غ
تولید کردن، ساختن	راهنما، رهبر، سرجنبان	عامل	زیرشمول
تولید کردن، ساختن	مصنوع، ساخته	پذیرا	ص

جدول ۲- نمونه‌هایی از خروجی‌های روش مبتنی بر تعمیم

۳-۶- ارزیابی روش مبتنی بر قاعده و تعمیم

برای ارزیابی این روش میزان دقت در فازهای مختلف بررسی می‌شود. برای به دست آوردن دقت و توجه قسمت‌های مختلف این فاز مجموعه‌ای متشکل از ۲۵۶ جمله از جملات پیکره بیجن‌خان به- صورت تصادفی مورد ارزیابی دستی قرار گرفت. این مجموعه جملات جمعاً شامل ۶۴۷ فعل بود که به برچسب‌زنی نقش‌های معنایی تعمیم‌یافته برای گروه‌های غیر فعلی موجود در جمله در ارتباط با فعل یا افعال موجود در جمله پرداخته شده است. دقت یافتن مرز فعلی ۸۵,۲ درصد و بازخوانی آن ۹۵,۴۹ درصد است. دقت این زیر بخش در تشخیص نقش‌های تعمیم یافته شبه‌عامل و شبه‌پذیرا ۷۱,۸۵ درصد است. بازخوانی نیز ۸۷,۳ درصد برآورد می‌شود. این دقت و بازخوانی با ارزیابی دستی به دست آمده است.

از مهم‌ترین عوامل مؤثر در خطا در این فاز می‌توان به پیچیده و غیر کلی بودن قواعد که گاه با هم تعارض دارند، وجود جملات استعاری در پیکره بیجن‌خان، وجود افعالی با ظاهر یکسان و ستاک فعلی متفاوت مانند مضارع التزامی افعال شدن و شستن (شو)، تشخیص نادرست مرز فعلی در برخی از افعال مرکب به دلیل ناکامل بودن پایگاه داده افعال در پوشش تمامی این افعال، خطای برچسب در پیکره، رعایت نشدن برخی از علائم نگارشی در پیکره اشاره نمود.

فاز دوم به طور متوسط دارای دقت ۶۹,۹ درصد و بازخوانی ۷۳,۷۱ درصد در ارزیابی دستی بود. در این ارزیابی از نتایج فاز اول در یافتن نقش‌های معنایی تعمیم‌یافته بر روی ۲۵۶ جمله تصادفی پیکره بیجن‌خان استفاده شده است. دقت و بازخوانی برای نقش‌های موضوعی مختلف نیز در جدول ۳ آمده است.

نقش	دقت	بازخوانی	نقش	دقت	بازخوانی
عامل	۶۰,۰۴٪	۷۰,۷۷٪	حالت	۶۶,۶٪	۶۹,۵۶٪
پذیرا	۶۶,۳۳٪	۶۵,۱۱٪	مسند	۶۳,۱۲٪	۸۷,۲۵٪
موضوع	۷۴,۳۴٪	۹۱,۵۸٪	تجربه‌گر	۵۲,۹۴٪	۴۲,۸۶٪
زمان	۹۱,۴۹٪	۷۰,۴۹٪	ابزار	۳۳,۳٪	۲۳,۷٪
مکان	۱۰۰٪	۵۴,۶۵٪	دلیل	۹۱,۶۶٪	۶۸,۷۵٪
مبدأ	۱۰۰٪	۱۰۰٪	ذینفع	۶۶,۶٪	۹۲,۳٪
مقصد	۷۶,۴۷٪	۶۵٪			

جدول ۳- دقت و بازخوانی به دست آمده برای نقش‌های موضوعی در فاز دوم روش قاعده و تعمیم

از عوامل مؤثر در خطا در این فاز می‌توان به خطا در پیمانه کمکی رفع ابهام معنایی در تشخیص معنای درست فعلی، خطا و یا ناکامل بودن سلسله مراتب فارس‌نت استفاده شده در مراحل کنونی، ناکامل بودن پایگاه داده ساختار وابستگی افعال و عدم پوشش تمامی افعال به دلیل همزمانی کار ساخت آن با کار ما، و تداخل میان قواعد مربوط به بازشناسی نقش‌های معنایی از یکدیگر اشاره نمود.

ارزیابی دقت فاز سوم مستقل از نتایج فازهای قبل نیست و خطای فازهای قبل در این فاز منتشر شده است. ارزیابی دستی، هنگامی که تنها روابط کاملاً درست را در نظر می‌گیریم دقت این روش را ۲۱٫۵۱ درصد برآورد می‌کند. دقت این روش با نادیده گرفتن اشتباهات شامل و زیرشمول ۴۳٫۱۷ درصد است. تفاوت بین میزان دقت کلی و دقت با نادیده گرفتن خطای سلسله مراتب نشان دهنده خطاهایی است که به دلیل کمبود اطلاعات و شواهد زیرشمولی از مفهوم اصلی به عنوان پیشنهاد ارائه شده است و یا به دلیل تشخیص نادرست معنای وابسته‌ها به مفاهیم سطح بالاتری از مفهوم اصلی نائل شده است.

از عوامل مؤثر در خطا در این فاز، می‌توان همه عواملی را که در خطای تعمیم روش قبل وجود داشت برشمرد. البته به دلیل استفاده شدن برچسب‌های پیکره بیجن‌خان در گروه‌های غیر فعلی، هسته گروه‌ها با خطای بسیار کمتری نسبت به روش تعمیم به دست می‌آید.

۷- نتیجه‌گیری

در این مقاله روشی جهت یادگیری روابط معنایی میان مقوله‌ای اسم و فعل در قالب استخراج روابط معنایی میان فعل و وابسته‌های آن ارائه شده است. برای این کار سه روش متفاوت مبتنی بر تحلیل-های لغوی و ریخت‌شناسی، مبتنی بر ویژگی‌های آماری و تعمیم و نیز روشی مبتنی بر قواعد بررسی شدند. دخیل شدن تحلیل‌های معنایی و بهره‌گیری از پایگاه داده معنایی موجب افزایش سطح خودکارسازی این روش نسبت به روش‌های مشابه شده است و امکان تشخیص وابسته‌های افعال و روابط میان فعل و وابسته‌ها را در شرایط نقص منابع زبانی اولیه نیز ایجاد می‌کند.

منابع

- اسلامی، محرم، مسعود شریفی آتشگاه، صدیقه علیزاده لمجیری، و طاهره زندی (۱۳۸۳). «واژگان زبانی زبان فارسی»، مجموعه مقالات اولین کارگاه پژوهشی زبان فارسی و رایانه، تهران.
- شمس‌فرد، مهرنوش، آرمان شریف‌زاده و آذین شهریاری‌فرد (۱۳۹۱). «توسعه نرم‌افزار مدیریت پیکره»، گزارش فنی، مرکز تحقیقات و مخابرات ایران.
- صدر موسوی، مریم (۱۳۸۶). *استخراج نقش‌های موضوعی جملات زبان فارسی*، پایان‌نامه کارشناسی ارشد، دانشگاه آزاد اسلامی واحد قزوین.

- Bijankhan, M. (2004). "Role of Language Corpora in Writing Grammar: Introducing Computer Software", *Iranian Journal of Linguistics*, No. 38.
- Gildea, D. and D. Jurafsky (2002). "Automatic Labeling of Semantic Roles", *Computational Linguistics*, Vol. 28, pp. 234-240.
- Hagholahi, M. and M. Shamsfard (2011). "A Semi-supervised Approach for Key-synset Extraction to be used in Word Sense Disambiguation", 7th Asia Information Retrieval Societies Conference (AIRS 2011), Dubai, United Arab Emirates.
- Kamel Ghalibaf, A., S. Rahati and A. Estaji (2009). "Shallow Semantic Parsing of Persian Sentences", 23rd Pacific Asia Conference on Language, Information and Computation.
- Lee, J. and Y. Song (2007). "Investigation of Weakly Supervised Learning for Semantic Role Labeling", IEEE. Sixth International Conference on Advanced Language Processing and Web Information Technology, pp. 165-170.
- Levin, B. and M.R. Hovav (1996). "From Lexical Semantics to Argument Realization" In *Handbook of Morphosyntax and Argument Structure*, ed. H. Borer, Dordrecht: Kluwer
- Levine, R.D. and W.D. Meurers (2006). "Head-Driven Phrase Structure Grammar Linguistic Approach, Formal Foundations, and Computational Realization", *Encyclopedia of Language and Linguistics*, Elsevier,
- Marquez, L. and X. Carreras (2008) "Semantic Role Labeling: An Introduction to the Special Issue", *Computational Linguistics*, Vol. 34, No 2, pp. 145-159.
- Noferesti, S. and M. Shamsfard (2011). "Developing a Persian Chunker", Technical Report, NLP Lab, Shahid Beheshti University.
- Ruiz-Casado, M., E. Alfonseca and P. Castells (2007). "Automatising the Learning of Lexical Patterns: An Application to the Enrichment of WordNet by Extracting Semantic Relationships from Wikipedia" *Elsevier, journal of Data & Knowledge Engineering*, vol. 61.
- Saeedi, P. and H. Faili (2012). "Feature Engineering Using Shallow Parsing in Argument Classification of Persian Verbs", AISP.

- Shamsfard, M. and others(2010). "Semi Automatic Development of FarsNet; The Persian WordNet", 5th Global WordNet Conference (GWA2010), Mumbai, India.
- Shamsfard, M. and H.S. Jafari, M. Ilbeigi(2010)."STeP-1: A Set of Fundamental Tools for Persian Text Processing", LREC 2010-8th Language Resources and Evaluation Conf., Malta.
- Spohr, D. (2008), "Extraction of Selectional Preferences for French Using a Mapping from EuroWordNet to the Suggested Upper Merge Ontology", 4th Global WordNet Conference.
- Wunsch, H. , P. Gupta and L. Lemnitzer(2010)."Enriching GermaNet with verb-noun Relations – a Case Study of Lexical Acquisition", 8th Language Resources and Evaluation Conf., Malta.

.....

ل س د ک ا ف