

## نخستین پیکره نقش‌های معنایی زبان فارسی

آزاده میرزائی\*<sup>۱</sup> و امیرسعید مولودی<sup>۲</sup>

۱- استادیار گروه زبان‌شناسی دانشگاه علامه طباطبائی<sup>(۵)</sup>، تهران

۲- استادیار بخش زبان‌های خارجی و زبان‌شناسی دانشگاه شیراز

(تاریخ دریافت: ۱۳۹۴/۰۶/۳۱؛ تاریخ پذیرش: ۱۳۹۴/۰۸/۲۵)

### چکیده

نخستین پیکره نقش‌های معنایی زبان فارسی که حدود ۳۰.۰۰۰ جمله از زبان فارسی معاصر را شامل می‌شود، به صورت دستی برچسب‌گذاری شده‌است. این پیکره بر اساس مفهوم نقش‌های معنایی فیلمور، لایه‌ای از اطلاعات مربوط به رابطهٔ محمول - موضوع را به ساخت نحوی پیکره وابستگی اضافه می‌کند. در این مجموعه، افعال، اسم‌های گزاره‌ای و صفت‌های گزاره‌ای به‌عنوان محمول‌های جمله در نظر گرفته شده و بنا بر نوع رویدادشان، در جمله تعیین ظرفیت شده‌اند. خروجی این پیکره بر اساس الگوی همایش زبان‌شناسی رایانه‌ای و پردازش زبان طبیعی (CoNLL) آماده شده‌است. نقش‌های معنایی مورد استفاده در برچسب‌زنی معنایی شامل دو گروه برچسب‌های موضوعی و برچسب‌های نقشی هستند. برچسب‌های موضوعی ساخت ظرفیتی محمول‌های جمله را به دست می‌دهند و برچسب‌های نقشی، افزوده‌های توصیف‌گر فعل یا کل جمله را شامل می‌شوند. تعداد نقش‌های معنایی ۲۷ مورد و تعداد موضوع‌های نقشی ۱۵ مورد است. دو برچسب وجه‌نما و نفی هم به‌عنوان برچسب نقشی مورد استفاده قرار گرفته است.

**کلیدواژه‌ها:** نقش‌های معنایی، پیکره گزاره‌های معنایی، اسم گزاره‌ای، صفت گزاره‌ای،

الگوی وابستگی CoNLL.

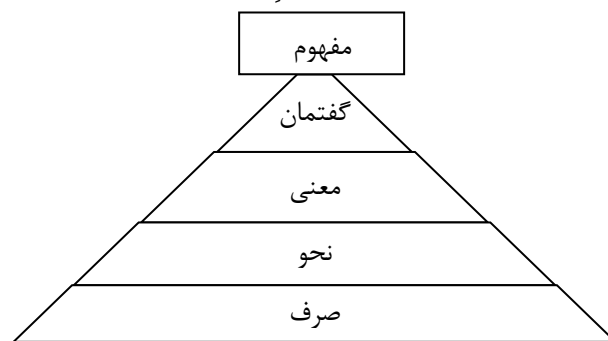
---

\* E-mail: azadeh.mirzaei@atu.ac.ir

۱. مقدمه<sup>۱</sup>

تولید پیکره‌های برچسب‌خورده زبانی و بهره‌گیری از آن‌ها در امر یادگیری ماشینی از جمله اقداماتی است که دست‌یابی به حداکثر ظرفیت‌ها را هدف قرار داده است. انسان از قوه اندیشه برخوردار است و دست‌یابی به لایه‌های مختلف زبان برایش امری طبیعی و ناخودآگاه محسوب می‌شود. از سوی دیگر، ماشین از ویژگی سرعت در پردازش اطلاعات بهره می‌برد. ترکیب این دو ویژگی موجب دست‌یابی به اطلاعات لایه‌های مختلف زبان با امکان دسته‌بندی، استخراج و پردازش اطلاعات در سریع‌ترین زمان ممکن خواهد شد.

بر این اساس، در بررسی ماشینی زبان آدمی که با هدف استخراج، دسته‌بندی و پردازش اطلاعات برای کاربردهایی چون ترجمه ماشینی، سامانه پرسش و پاسخ، مشابهت‌یابی، ویرایش متن و مانند آن انجام می‌شود، هدف نهایی، دست‌یابی به مفهوم است. برای این منظور، دسترسی به اطلاعات موجود در لایه‌های مختلف متن (نوشتاری یا گفتاری) ضروری است، اما از آنجاکه بهترین نوع اطلاعات، برای ماشین، صورتی‌ترین آن‌هاست، تهیه پیکره‌های زبانی از صورتی‌ترین سطوح زبان آغاز می‌گردد و همچنان که شکل (۱) نشان می‌دهد، دست‌یابی ماشین به سطوح مختلف زبان، مرحله‌به‌مرحله انجام می‌شود و برای کسب تجربه ماشینی، گذر از صرف، نحو، معنی و گفتمان به ترتیب صورت می‌گیرد. به بیان دیگر، ماشین از اطلاعات هر سطح بهره می‌برد تا سطح بعدی را بفهمد و این روند ادامه می‌یابد تا نهایتاً به مفاهیم پنهان در پس متون دسترسی کامل پیدا کند.



شکل (۱) سطوح تحلیل زبان

در پیکره‌سازی برای مقاصد یادگیری ماشینی، در نخستین سطح از سطوح زبان برچسب اجزاء سخن<sup>۲</sup> و اطلاعات صرفی مربوط به واژه‌های زبان تعیین می‌شوند. نمونه‌ای از برچسب‌گذاری اطلاعات صرفی در جمله<sup>۱</sup> ارائه شده است. از جمله این پیکره‌ها در زبان فارسی می‌توان به پیکره بی‌جن‌خان<sup>۳</sup> (۲۰۱۱) اشاره کرد.

(۱) پسر	کتاب‌هایی	خرید.
اسم، جاندار، مفرد	اسم، بی‌جان، جمع	فعل، گذشته، ساده، سوم شخص مفرد

در سطح نحو، نقش هر واژه در ارتباط با واژه‌های دیگر جمله مطرح است. در این سطح، اطلاعات نحوی جملات و عبارت‌های زبانی برحسب رویکرد زبان‌شناختی هر پیکره نحوی و غالباً در ارتباط مستقیم با فعل جمله (به‌عنوان مرکز ثقل ساختاری جمله) مشخص می‌شود. برای مثال، تحلیل نحوی جمله زیر در ارتباط با فعل «زدن» صورت پذیرفته است.

(۲) علی	تابلو را	به دیوار	زد.
فاعل	مفعول	مفعول حرف اضافه	

تظاهرات نحوی متفاوت افعال، عموماً حکایت از معناهای متفاوت دارند. برای مثال، همان‌طور که در نمونه‌های ۳ و ۴ دیده می‌شود، ظرفیت‌های نحوی متفاوت فعل «زدن» معنی‌های متفاوت آن را به دست داده است و این یعنی اینکه تعیین و یادگیری صورت‌بندی‌های نحوی متفاوت، راهی به سوی تشخیص معنی است.

(۳) علی دوستش را زد. <فاعل، (مفعول) [را+]>

(۴) علی پوسترها را به دیوار زد. <فاعل، (مفعول) [را+/-]، (مفعول حرف‌افزای) [به|بر|روی|زیر] >

پیکره نحوی وابستگی زبان فارسی (رسولی و همکاران، ۲۰۱۳) این ارتباطات نحوی را طبقه‌بندی و ارائه کرده است، اما نکته در این است که تنوع‌های زبانی به همین جا ختم نمی‌شوند و بلافاصله صورت‌هایی پدیدار می‌شوند که علی‌رغم یکسان بودن ساخت نحوی‌شان، معانی متفاوت دارند. برای مثال، اگر دو جمله «آن‌ها داور را زدند» و «آن‌ها موهایشان را زدند» را با یکدیگر مقایسه کنیم، درمی‌یابیم که فعل «زدن» با دو معنای متفاوت (یکی کتک زدن و دیگری کوتاه

کردن) در هر دو جمله دارای ساخت نحوی واحد < فاعل، مفعول > است. همین نکته، یعنی یکسان بودن ساخت نحوی با وجود تفاوت‌های معنایی، یادگیری ماشینی را با مشکل مواجه کرده و این امر نشانگر اهمیت تولید پیکره نقش‌های معنایی برای پردازش ماشینی است.

(۵) آن‌ها داور را زدند. فاعل، (مفعول) [+/-] کنشگر، کنش‌پذیر  
 (۶) آن‌ها موهایشان را زدند. فاعل، (مفعول) [+/-] سبب، کنش‌گذار

در این موارد دسترسی به لایه معنایی می‌تواند در ابهام‌زدایی از گزاره‌ها مفید واقع شود.

البته گاهی عکس نکته فوق نیز در زبان رخ می‌دهد؛ به این معنی که صورت‌هایی ارائه می‌شوند که علی‌رغم یکسان بودن معنی و اشاره به یک گزاره واحد معنایی، ساخت نحوی متفاوت دارند. در این موارد نیز اگر ماشین فقط به اطلاعات نحوی، تکیه کند، در تحلیل زبان با مشکل مواجه می‌شود.

(۷) علی غذا را پخت.  
 فاعل مفعول  
 (۸) غذا پخت.  
 فاعل

فعل «پختن» در دو مثال (۷) و (۸) به‌لحاظ معنایی یک گزاره واحد محسوب می‌شود، اما بدون آنکه مجهول شده باشد و ساخت‌واژه آن تغییر کرده باشد، دو ساخت نحوی متفاوت را به نمایش گذاشته است. بنابراین، اتکا به سطح نحو سبب می‌شود ماشین برای فعل «پخت» به‌اشتباه، دو معنی متفاوت را دریافت کند.

کارآمدی اطلاعات معنایی در امور پردازشی ارزشمندی همچون ترجمه ماشینی، خلاصه‌سازی، سامانه‌های پرسش‌وپاسخ و غیره، سبب شده است تا به شکل فزاینده‌ای، حتی با صرف هزینه‌های بالا، برای زبان‌های مختلف چنین پیکره‌هایی ساخته شود.

## ۲. پیشینه مطالعات

شمار پیکره‌های معنایی نسبت به پیکره‌های نحوی کمتر است. در زبان انگلیسی پروپبنک<sup>۴</sup> (پالمر<sup>۵</sup> و همکاران، ۲۰۰۵)، فریمنت<sup>۶</sup> (بیکر<sup>۷</sup> و همکاران، ۱۹۹۸) و وربنت<sup>۸</sup> (دانگ<sup>۹</sup> و همکاران، ۱۹۹۸؛ کیپر<sup>۱۰</sup> و همکاران، ۲۰۰۰) از مهم‌ترین طرح‌های معنایی هستند که هدفشان دستیابی به اطلاعات گزاره‌های معنایی است. پروپبنک، پیکره دادگان معنایی زبان انگلیسی، با روش‌شناسی مشخص در تعریف نقش‌های معنایی، لایه‌ای از اطلاعات مربوط به نقش‌های معنایی را به پیکره دادگان نحوی پرتربنک<sup>۱۱</sup> (مارکوس<sup>۱۲</sup>، ۱۹۹۳) افزوده است. در این پیکره برای پرهیز از مناقشات موجود در تعریف نقش‌های معنایی، برحسب شباهت میان نقش‌ها، از اختصارات شماره‌ای از صفر تا ۵ استفاده شده است. تمایز میان دو نقش Arg0 و Arg1 بر اساس اثر دوتی<sup>۱۳</sup> (۱۹۹۱) انجام شده است. لازم به توضیح است که پیکره بزرگ نقش‌های معنایی زبان فارسی بر اساس این رویکرد نیز برچسب‌گذاری شده که بحث در خصوص شیوه تولید و آمار مربوط به آن در اثر دیگری به‌طور مبسوط ارائه می‌شود.

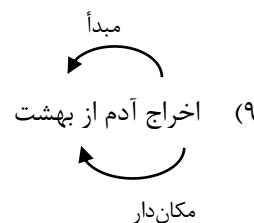
زبان آلمانی نیز با افزودن لایه‌ای از اطلاعات معنایی به پیکره دادگان نحوی خود، یعنی پیکره نحوی تایگر<sup>۱۴</sup> (براننتس<sup>۱۵</sup> و دیگران، ۲۰۰۲)، پیکره ۴۰ هزار جمله‌ای گزاره‌های معنایی زالساک<sup>۱۶</sup> (بورخارت<sup>۱۷</sup>، ۲۰۰۶) را به دست داده است.

در زمینه شیوه ارائه پیکره‌های نحوی و معنایی، در بین زبان‌ها تنوع به چشم می‌خورد؛ به این صورت که در برخی از زبان‌ها همچون دو مورد اخیر، پروژه‌های معنایی و نحوی در دو مرحله جدا و مستقل ارائه شده‌اند؛ زبان چینی نیز جزو این گروه از زبان‌هاست که پیکره گزاره‌های معنایی (خو<sup>۱۸</sup> و پالمر، ۲۰۰۹) و پیکره نحوی (خو، ۲۰۰۵) خود را در دو مرحله مجزا تهیه و ارائه کرده است. در برخی دیگر از زبان‌ها، اطلاعات هر دو سطح نحو و معنا در یک پیکره واحد گردآوری شده است؛ پیکره اسپانیایی و کاتالان (توله<sup>۱۹</sup> و همکاران، ۲۰۰۸) متعلق به این گروه از زبان‌هاست. نهایتاً در گروهی دیگر از زبان‌ها، همچون زبان ژاپنی (کاواهارا<sup>۲۰</sup>، ۲۰۰۲) بخشی از پیکره نحوی موجود برچسب‌گذاری معنایی شده است. تولید پیکره‌های گزاره‌های معنایی برچسب‌خورده در زبان‌های مختلف به سرعت رو به افزایش است.

### ۳. پیکره بزرگ نقش‌های معنایی زبان فارسی

پیکره بزرگ نقش‌های معنایی زبان فارسی، لایه معنایی را بر روی پیکره نحوی وابستگی زبان فارسی (رسولی و همکاران، ۲۰۱۳) قرار داده است. مجموعه جملات برچسب‌خورده حدود ۳۰۰۰۰ جمله است. در هر یک از جملات پیکره، بر مبنای دستور حالت فیلمور<sup>۲۱</sup> (۱۹۷۸، ۱۹۹۸، ۲۰۰۱) کلیه نقش‌های معنایی مربوط به افعال، اسم‌های گزاره‌ای و صفات گزاره‌ای تعیین شده‌اند. علاوه بر این، چون پیکره حاضر مجموعه‌ای از جملات را شامل می‌شود، در کنار سازه‌های موضوع، سازه‌های غیرموضوع شامل افزوده‌ها و برخی اطلاعات کلامی نیز برچسب‌گذاری شده‌اند.

منظور از اسم گزاره‌یدر گزاره‌ای، اسم‌هایی هستند که همانند فعل‌ها، ساخت موضوعی<sup>۲۲</sup> دارند، اما برخلاف فعل، به متمم‌های خود حالت نمی‌دهند و قادر به حمل شناسه‌های فعلی هم نیستند (کریمی‌دوستان، ۱۳۸۶). به‌عنوان نمونه‌ای از اسامی گزاره‌ای می‌توان به واژه «اخراج» در عبارت «اخراج آدم از بهشت» اشاره کرد. چون اسم «اخراج» ساخت موضوعی دارد، آن را متعلق به طبقه اسامی گزاره‌ای می‌دانیم. موضوع‌های هسته اسمی مذکور در عبارت فوق عبارت‌اند از: «آدم» (مکان‌دار<sup>۲۳</sup>) و «از» (مکان). باید به این نکته اشاره کرد که در پیکره حاضر، برچسب‌زنی معنایی اسم‌های گزاره‌ای بر اساس برچسب‌زنی معنایی فعل صورت گرفته است؛ به این معنی که اسم‌های مشتق از فعل‌ها را همانند فعل‌ها برچسب می‌زنیم.



همچنین، صفات گزاره‌ای آن گروه از صفاتی هستند که می‌توانند همانند محمول‌های فعلی، ساخت موضوعی داشته باشند. برای مثال، در گروه صفتی «خوشحال از قبولی»، گروه حرف‌افزافه‌ای /از قبولی/، یکی از موضوع‌های صفت خوشحال است و بنابراین، خوشحال یک صفت گزاره‌ای است. لازم به ذکر است که چون تعداد زیادی از افعال مرکب زبان فارسی در پیکره حاضر در بخش فعل‌یاد خود صفت داشتند، تعداد صفات گزاره‌ای در این بخش کاهش یافته است.

خروجی داده‌های برچسب‌خورده بر اساس الگوی همایش زبان‌شناسی رایانه‌ای و پردازش زبان طبیعی (CoNLL) <sup>۲۴</sup> فراهم آمده است. در جدول زیر برخی از اطلاعات آماری پیکره ارائه شده است.

جدول (۱) آمارهای مربوط به پیکره

۲۹,۹۸۲	تعداد کل جملات
۱۶/۶۱	میانگین طول هر جمله
۶۲۸۸۹	تعداد کل افعال
۹۲۰۰	تعداد فعل منحصربه‌فرد
۱۳۰۰	تعداد اسم‌های گزاره‌ای منحصربه‌فرد
۳۰۰	تعداد صفات گزاره‌ای منحصربه‌فرد

#### ۴. برچسب‌های معنایی

##### ۴-۱. نقش‌های معنایی

فیلمور در دستور حالت، فهرستی از نقش‌های معنایی را با عناوینی همچون حالت کنادی <sup>۲۵</sup>، حالت وسیله‌ای <sup>۲۶</sup>، حالت تأثیرپذیری <sup>۲۷</sup>، حالت تحقیقی <sup>۲۸</sup>، حالت مکانی <sup>۲۹</sup> و حالت مفعولی <sup>۳۰</sup> تعریف کرد (دبیرمقدم، ۱۳۸۹). از آن زمان تا به امروز، نوع و تعداد نقش‌های معنایی همیشه محل بحث و اختلاف نظر بوده است. وربنت با استفاده از فهرستی از نقش‌های معنایی و بر اساس طبقه‌بندی لوین <sup>۳۱</sup> (۱۹۹۳)، مجموعه‌ای از افعال زبان انگلیسی را دسته‌بندی و قاب معنایی هر یک از آن‌ها را مشخص کرده است. پیکره بزرگ نقش‌های معنایی زبان فارسی بر اساس نقش‌های مورد نظر وربنت که ریشه در نقش‌های معنایی فیلمور دارد، برچسب‌گذاری شده است. لازم به توضیح است که با توجه به برخی ویژگی‌های منحصربه‌فرد زبان فارسی، برخی از این نقش‌ها در هم تلفیق شده و درمقابل، برخی نقش‌های دیگر به این مجموعه اضافه شده‌اند.

فهرستی از نقش‌های معنایی به‌کاررفته در پیکره بزرگ نقش‌های معنایی در جدول شماره ۲ مشخص شده است. تعاریف مشخص کنشگر، کنش‌پذیر و مانند آن، به دلیل سابقه حضور در

پیشینه مطالعات زبان‌شناسی یا به دلیل شفافیت معنایی (همانند مفاهیم دارنده یا مقدار)، نیازی به بازتعریف ندارند. از این میان، نقش‌های هم‌کنش‌پذیر، هم‌کنشگر و هم‌مکان‌دار در تعیین افعال دوسویه و افعال ترکیبی مورد استفاده قرار گرفتند؛ به بیان دیگر، در افعال کنشی مانند گفتگو کردن، دو شرکت‌کننده فعال به‌عنوان کنشگر و هم‌کنشگر برچسب می‌گیرند و در صورت‌های فعلی همچون مخلوط کردن، ترکیب کردن، جوش دادن و مانند آن، دو شرکت‌کننده مخلوط‌شونده هم‌کنش‌پذیرند.

تمایز سه‌تایی کنش‌پذیر (شرکت‌کننده تحت تأثیر فعل)، مکان‌دار (شرکت‌کننده‌ای که در سکون یا حرکت است و فعل به وضعیت آن اشاره می‌کند) و کنش‌بر (شرکت‌کننده‌ای که نقطه مقابل کنش فعل است اما از آن تأثیر نمی‌گیرد) دسته‌بندی ویژه و منحصر به پیکره بزرگ دادگان معنایی زبان فارسی است.

مکان غیرفیزیکی یا وضعیت‌ها در شرایطی که رویداد فعلی به آن‌ها ختم می‌شوند، «نتیجه» اند و در شرایطی که فعل از سمت آن‌ها آغاز می‌شود، «منشأ» هستند. در همین خصوص، «پذیرنده» شرکت‌کننده جاندار است که نقطه پایان/انجام و یا هدف فعل قرار می‌گیرد.

(۱۰) او این مسئله را از من نمی‌پذیرد.

منشأ

(۱۱) او نوشته‌اش را به داستان تبدیل کرد.

نتیجه

(۱۲) علی مداد را به دوستش داد.

پذیرنده

«دربارگی» موضوع رخداد و رویداد جمله است و «آغاز و انجام» در بازه‌های زمانی، نقطه شروع و پایان رویداد محسوب می‌شوند.

(۱۳) علی با دوستش در مورد خاطراتشان صحبت می‌کردند و این صحبت از صبح تا ظهر طول کشید.

آغاز انجام

هم‌کنشگر      دربارگی



و در نهایت، «نسبت»<sup>۳۲</sup> عنوانی است برای ویژگی که به دیگر شرکت‌کننده‌ها نسبت و یا اسناد داده می‌شود.

جدول ۲) فهرست نقش‌های معنایی پیکره

کنشگر	کنش پذیر	آغاز	هم‌کنشگر	پذیرنده
کنش پذیر	کنش بر	انجام	هم‌کنش پذیر	نتیجه
تجربه‌گر	مکان‌دار	مقدار	هم‌مکان‌دار	منشأ
سبب	دارنده	مکان	نسبت	مبدأ
بهره‌ور	دارایی	محرك	دربارگی	مقصد

## ۲-۴. افزوده‌ها

افزوده‌ها خود به سه گروه افزوده‌های حاشیه‌ای<sup>۳۳</sup>، افزوده‌های وجهی<sup>۳۴</sup> و افزوده‌های متنی<sup>۳۵</sup> تقسیم می‌شوند. افزوده‌های حاشیه‌ای، اطلاعاتی در خصوص منظور، علت، زمان، مکان، شرایط، روش و مقدار محمول‌ها ارائه می‌دهند. افزوده‌های وجهی نظر نویسنده یا گوینده را در خصوص میزان قطعیت و امکان وقوع گزاره بیان می‌کنند و تمامی انواع آن‌ها اعم از وجه‌نماهای برداشتی<sup>۳۶</sup>، توانمندی<sup>۳۷</sup> و شواهدی<sup>۳۸</sup>، با هر تظاهر زبانی اعم از فعل، قید، گروه حرف‌افزای و مانند آن، برچسب وجه‌نما دریافت می‌کنند. افزوده‌های متنی نیز پیونددهنده‌هایی هستند که بین جمله‌ها ارتباط برقرار می‌کنند.

از میان افزوده‌های حاشیه‌ای، سه افزوده همراهی، وصف و مرجع‌دار با سایر افزوده‌های این گروه تفاوت دارند. این سه افزوده برخلاف سایر اعضای این گروه که به‌لحاظ معنایی در ارتباط مستقیم با فعل جمله هستند، در ارتباط مستقیم با موضوع‌ها تعریف می‌شوند؛ به این ترتیب که افزوده همراهی، شخصی حقیقی یا حقوقی است که کنشگر جمله را در انجام کنش همراهی می‌کند، افزوده وصف افزوده‌ای است که با موضوع‌ها در یک رابطه اسنادی قرار دارد و به‌تنهایی می‌تواند در قالب یک وضعیت گزاره‌ای به‌کار رود و نهایتاً، افزوده مرجع‌دار به‌عنوان نظیر وجودی موضوع‌ها تعریف می‌شود. برای مثال، در جمله «علی با دوستش به سینما رفت»، گروه حرف‌افزای «با دوستش» که در انجام کنش رفتن به سینما، علی را همراهی کرده است، برچسب «همراهی»

دریافت می‌کند یا در جمله «علی خوشحال وارد کلاس شد»، «علی» و «خوشحال» در یک رابطه اسنادی با یکدیگر قرار دارند و «خوشحال» خود دارای معنای گزاره‌ای است؛ بنابراین، کلمه «خوشحال» برچسب «وصف» را دریافت می‌کند. در جمله‌ای نظیر «آن‌ها خودشان کارهایشان را انجام می‌دهند» نیز کلمه «خودشان» که نظیر وجودی «آن‌ها» محسوب می‌شود، برچسب «مرجع‌دار» را به خود اختصاص می‌دهد.

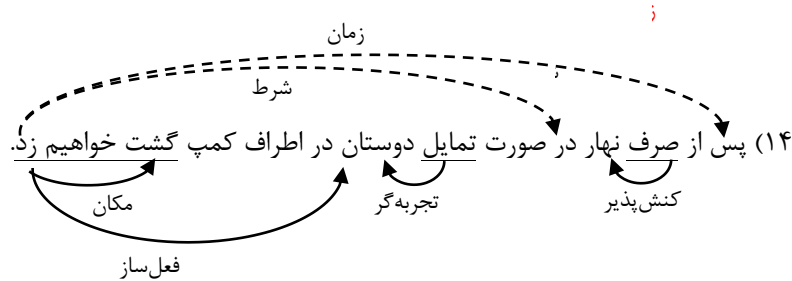
هدف، نقشی همانند بهره‌ور دارد، اما از آنجاکه در ظرفیت فعل حضور ندارد، در ردیف افزوده‌ها قرار می‌گیرد. برای مثال، در جمله «او به خاطر من برای این امتحان خیلی تلاش می‌کند»، «من» به‌عنوان هدف برچسب‌گذاری می‌شود.

قیدهای فراوانی همانند هرگز، گاهی، اغلب و مانند آن، برچسب تکرار دریافت کردند و صورت‌هایی چون نه... نه... به‌عنوان نفی برچسب‌گذاری شدند. لازم به توضیح است که برچسب نفی برای زبان‌هایی چون انگلیسی، آلمانی و زبان‌های دیگری که از سازه مشخص و جدا برای مفهوم نفی استفاده می‌کنند، معنادارتر است؛ در زبان فارسی عنصر نفی با وندافزایی مشخص می‌شود و در بخش صرف می‌تواند تعیین وضعیت شود.

فهرستی از افزوده‌های حاشیه‌ای و وجهی پیکره حاضر در جدول شماره (۳) ارائه شده است. لازم به توضیح است که در روند برچسب‌گذاری، اگر نوع افزوده در هیچ‌یک از طبقات مورد اشاره قرار نگرفته باشد، از عنوان پوششی قید استفاده شده است.

جدول (۳) فهرست افزوده‌های حاشیه‌ای و وجهی پیکره

جهت	روش	هدف
مکان	شرط	تکرار
زمان	ابزار	نفی
مقدار	همراهی	قید
منظور	وصف	وجه‌نما
علت	مرجع‌دار	



برای نمونه، نوع افزوده‌ها در جمله ۱۴ مشخص شده‌اند. همان طور که ملاحظه می‌شود، جمله فوق دو محمول اسمی و یک محمول فعلی دارد که زیرشان خط کشیده شده است. هر یک از وابسته‌های محمول‌ها نیز برحسب نقششان در جمله یا عبارت، یک برچسب معنایی دریافت کرده‌اند.

### ۴-۳. نقش‌های کلامی

نقش‌های کلامی در پیکره حاضر دربردارنده سه برچسب پیونددهنده گفتمانی، مبتدا و پیوند بند موصولی<sup>۳۹</sup> است. پیونددهنده‌های گفتمانی آن دسته از سازه‌های زبانی هستند که کارکرد اصلی‌شان برقراری ارتباط میان جملات است. برای مثال، زنجیره «به عبارت دیگر» در فاصله میان دو جمله، از مصادیق پیونددهنده‌های گفتمانی به شمار می‌رود. برچسب مبتدا برای تعیین سازه‌های مبتداده ضمی‌رگذار مورد استفاده قرار می‌گیرد و نهایتاً برای نشان دادن حرکت بندهای اسمی اعم از متممی یا موصولی (توضیحی و توضیحی) از برچسب پیوند بند موصولی استفاده می‌شود که در بیش از ۹۴۰۰ بند اسمی، نسبت جایگاهی هسته بند و بند را مشخص کرده است.

### ۵. فرایند برچسب‌گذاری پیکره

برای تهیه پیکره مذکور دو درس‌نامه برچسب‌گذاری گزاره‌های فعلی و گزاره‌های اسمی تهیه شد و در اختیار برچسب‌زنان قرار گرفت. در میان راه نیز، بنا بر نیاز، یادداشت‌های تکمیلی به همراه مثال‌های کافی به مجموعه شیوه‌نامه‌ها اضافه شد. همچنین، اطلاعات نحوی جملات در تمام طول برچسب‌زنی در اختیار برچسب‌زنان قرار داشت.

فرایند برچسب‌گذاری پیکره حاضر در سه مرحله اصلی (البته به صورت هم‌زمان) انجام پذیرفت. در مرحله نخست، محمول‌های فعلی، اسمی و صفات جملات شناسایی شدند. در مرحله دوم، شرکت‌کنندگان محمول‌های مذکور برحسب نقششان در رویداد گزاره‌ای، برچسب معنایی مناسب را دریافت کردند. در مرحله سوم نیز افزوده‌ها و برخی از ویژگی‌های کلامی ذکرشده در قسمت پیشین، هدف برچسب‌گذاری قرار گرفتند؛ همچنین، در جریان برچسب‌گذاری نقش‌های معنایی، فعل‌هایی که در معانی متفاوت به کار می‌رفتند، مشخص شده، ضمن دریافت ساخت ظرفیتی مناسب، شماره مفهومی<sup>۴۰</sup> متفاوتی از همتای همنام خود دریافت کردند. برای مثال، فعل «زدن» در معناهای دزدیدن، کتک زدن، نصب کردن، مانستن، ضربان داشتن و مانند آن، شماره مفهومی متفاوتی دریافت کرده است.

## ۶. اصلاح پیکره

برای به دست آوردن آمار معناداری از درستی روند برچسب‌زنی، حدود ۶۰۰۰ جمله به صورت دوبرچسبی آماده شد؛ به این ترتیب که یک جمله واحد توسط دو برچسب‌زن متفاوت که از خروجی کار یکدیگر مطلع نبودند، برچسب‌گذاری شد. سپس، اختلافات موجود در داده‌های دوبرچسبی گزارش شدند و توسط کاربر ناظر (نویسندگان مقاله حاضر) مورد بازبینی و اصلاح قرار گرفتند. لازم به توضیح است که برای دسترسی به اختلاف‌های میان داده‌های دوبرچسبی به منظور آمارگیری، داده‌های اصلاح‌شده در بخشی مستقل از اختلافات ذخیره شدند. همچنین، در پایان کار، یک برنامه غلط‌یاب آماده شد که هدف آن، گزارش تنوع‌های موجود در ساخت‌های ظرفیت معنایی هر فعل در کل پیکره بود؛ به این ترتیب که برای هر فعل، تمامی تنوع‌های ساختی آن در پیکره توسط برنامه گزارش می‌شد و ناظران درباره این تنوع‌ها تصمیم‌گیری می‌کردند. اساساً تنوع‌های موجود در ساخت‌های ظرفیتی برنامه غلط‌یاب ممکن بود سه دلیل عمده داشته باشد. در حالت نخست، ممکن بود تنوع به دلیل چندمعنایی بودن فعل باشد؛ یعنی به‌طور مثال، برای فعل «زدن» در معناهای مختلف، ظرفیت‌های معنایی متفاوت در نظر گرفته شده باشد. در این حالت، امکان داشت برای معانی متفاوت فعل مذکور، به‌اشتباه شماره‌های مفهومی مستقلی در نظر گرفته نشده باشد که در نهایت این اشتباه توسط ناظران برطرف می‌شد. دلیل دوم تنوع‌ها، حذف بخشی از

ساخت موضوعی یک فعل در جملات مختلف بود. به‌طور مثال، برنامه غلط‌یاب به دلیل مشاهده تنوع موجود در ساخت‌های ظرفیتی فعل «گفت» در دو جمله زیر آن‌ها را گزارش می‌کرد. در این مورد هم اختلاف، طبیعی و مجاز بود و نیازی به اصلاح نداشت.

۱۵) علی ماجرا را برای دوستش گفت. ساخت ظرفیتی: کنشگر، دربارگی، پذیرنده

۱۶) علی ماجرا را گفت. ساخت ظرفیتی: کنشگر، دربارگی

دلیل سوم تنوع‌های ساخت‌های معنایی نیز خطاهای انسانی در برچسب‌زنی بود که توسط ناظران اصلاح شدند. البته باید خاطر نشان کرد که مرحله اصلاح، نگهداری و بهینه‌سازی پیکره همچنان باز خواهد بود.

## ۷. آمار

برای مقایسه‌ی نمای نحوی و معنایی زبان فارسی، از پیکره‌ی نحوی زبان فارسی تعداد ۶ مورد از پربسامدترین نقش‌های نحوی مرتبط با فعل، شامل فاعل، فعل‌یار، مفعول، مسند، مفعول حرف‌اضافه‌ای و متمم بندی فعل، در جدول (۵) ارائه شده و سپس در مقابل هر یک از این ۶ نقش، از پیکره‌ی نقش‌های معنایی فارسی، تعداد ۵ مورد از معادل‌های معنایی آن از پربسامدترین نقش‌های معنایی ارائه شده است. برای مثال، فاعل که در پیکره‌ی دادگان نحوی زبان فارسی، فراوان‌ترین نقش نحوی مرتبط با فعل بوده است، در پیکره‌ی نقش‌های معنایی زبان فارسی در ۴۳ درصد از موارد به‌عنوان کنشگر، در ۱۶ درصد موارد به‌عنوان کنش‌بر، در ۹ درصد به‌عنوان تجربه‌گر، در ۸ درصد به‌عنوان کنش‌پذیر و در ۵ درصد از موارد به‌عنوان مکان‌دار برچسب خورده است. همان‌طور که ملاحظه می‌شود، مجموع درصدهای مذکور در مورد مابه‌ازاهای معنایی نقش فاعل به عدد ۸۱ می‌رسد و این نکته به این معناست که ۱۹ درصد باقی‌مانده به نقش‌های معنایی دیگری اختصاص یافته که در جدول زیر نیامده‌اند؛ بنابراین، در این جدول برای ۶ نقش نحوی مذکور، تنها ۵ مابه‌ازای معنایی پربسامد ذکر شده است.

جدول ۵) پربسامدترین نقش‌های نحوی فعل در پیکره نحوی به همراه درصدی از مابه‌ازاهای معنایی آنها

مفعول		فعل یار		فاعل	
۳۲	کنش بر	۹۹	فعل ساز	۴۳	کنشگر
۲۵	کنش پذیر	۰/۰۴۶	کنش بر	۱۶	کنش بر
۱۴	فعل ساز	۰/۰۲۱	کنش پذیر	۹	تجربه گر
۷	مکان دار	۰/۰۱۸	کنشگر	۸	کنش پذیر
۳	دربارگی	۰/۰۱۲	نسبت	۵	مکان دار
متمم بندی فعل		مفعول حرف اضافه ای		مسند	
۳۵	کنش بر	۲۰	مکان	۵۶	فعل ساز
۱/۴	وجه نما	۱۹	فعل ساز	۳۵	نسبت
۱/۳	محرك	۱۱/۸۲	پذیرنده	۱/۵	کنش بر
۱/۰۷	آغاز	۱۱/۴۵	کنش بر	۱/۰۸	کنشگر
۰/۹۱	نسبت	۷/۸۹	مقصد	۰/۵۸	مکان

فراوانی ۱۰ رابطه معنایی پربسامد پیکره را در سه جدول (۶)، (۷) و (۸) می‌توان مرور کرد. در جدول (۶)، پربسامدترین روابط معنایی در میان تمامی برچسب‌های معنایی اعم از برچسب‌های موضوعی، افزوده‌ها و نقش‌های کلامی ارائه شده است که برچسب قید با ۲۴/۷ درصد و فعل ساز با ۲۲/۲ درصد به ترتیب بیشترین فراوانی را به خود اختصاص داده‌اند. جدول (۷) به فراوانی ۱۰ برچسب پربسامد نقش‌های معنایی (موضوعی) اختصاص دارد. همان طور که ملاحظه می‌شود، برچسب کنش بر با ۲۵/۴ درصد و کنشگر با ۱۸ درصد بیشترین بسامد را دارند. در جدول (۸) نیز

فراوانی ۱۰ افزوده پرسامد پیکره نقش‌های معنایی ارائه شده است و همان طور که مشاهده می‌شود، افزوده مقدار با ۱۸ درصد و افزوده زمان با ۱۴/۹ درصد دارای بالاترین بسامدها هستند.

جدول ۸ فراوانی افزوده‌های معنایی		جدول ۷ فراوانی نقش‌های معنایی پیکره		جدول ۶ فراوانی برچسب‌های پیکره معنایی (اعم از برچسب‌های موضوعی و افزوده‌ها)	
افزوده‌ها		نقش‌های معنایی		کل برچسب‌ها	
۱۸	مقدار	۲۵/۴	کنش‌بر	۷/۲۴	قید
۱۴/۹	زمان	۱۸	کنشگر	۲/۲۲	فعلساز
۱۰/۲	قید	۱۱/۹	کنش‌پذیر	۱۲	کنش‌بر
۹/۸	روش	۷/۲	دربارگی	۵/۶	کنشگر
۷/۹	مکان	۵/۲	نسبت	۶/۵	کنش‌پذیر
۷/۵	وصف	۴/۶	نتیجه	۴/۳	دربارگی
۴/۸	تکرار	۴/۴	تجربه‌گر	۵/۲	نسبت
۴/۰۱	شرط	۴/۳	مکان	۲/۲	نتیجه
۴	منظور	۴	مکان‌دار	۱/۲	B0
۳/۳	علت	۳/۸	پذیرنده	۲	تجربه‌گر

## ۸. نتیجه‌گیری

اثر حاضر به معرفی نخستین پیکره نقش‌های معنایی برچسب‌خورده زبان فارسی، تحت عنوان پیکره بزرگ نقش‌های معنایی زبان فارسی می‌پردازد که اطلاعات سطح معنا را بر پیکره نحوی وابستگی زبان فارسی افزوده است. در پیکره حاضر تمامی وابسته‌های محمول‌های فعلی، اسمی و صفتی برحسب کارکردشان در جملات و عبارات زبانی، برچسب معنایی دریافت کردند و مجموعاً حدود ۳۰،۰۰۰ جمله در سطح معنا برچسب‌گذاری شدند. به منظور دستیابی به حداکثر دقت در برچسب‌زنی، درس‌نامه‌ها و دستورالعمل‌های مفصلی در اختیار برچسب‌زنان قرار گرفت و در طول

فرایند برچسب‌زنی، تمامی اطلاعات سطح نحو در اختیار برچسب‌زنان قرار داشت. شایان ذکر است که در کنار پیکره حاضر، دو محصول کارآمد دیگر تحت عنوان فرهنگ ظرفیت معنایی افعال فارسی و همچنین تجزیه‌گر خودکار معنایی نیز تهیه شد که جزئیات آن‌ها در آثار مستقل دیگری ارائه خواهد شد.

### تقدیر و تشکر

طرح تولید پیکره بزرگ معنایی زبان فارسی که داده ارزشمندی را برای پژوهشگران زبان‌شناس، زبان‌شناسان رایانشی و مهندسی هوش مصنوعی فراهم آورده، با حمایت مرکز تحقیقات کامپیوتری علوم اسلامی نور انجام شده است. لازم است از گروه زبان‌شناس و رایانه‌ای مجموعه، ندا پورمرتضی خامنه، پریناز دادرس، مرتضی رضایی شریف‌آبادی، سلیمه زمانی، سارا شفیعی، فاطمه صدقی، سعیده قدردوست نخچی، منوچهر کوهستانی، مصطفی مهدوی و هومن مهیار که در مسیر تولید این پیکره همکاری داشتند، تشکر کنیم. همچنین، مراتب سپاس خود را از همکاری و حمایت ویژه دکتر مهدی بهنیافر ابراز می‌داریم.

### منابع

- دبیر مقدم، محمد. (۱۳۸۹). *زبان‌شناسی نظری: پیدایش و تکوین دستور زایشی*. ویراست دوم، چاپ چهارم. تهران: سمت.
- کریمی دوستان، غلامحسین. (۱۳۸۶). «اسامی و صفات گزاره‌ای در زبان فارسی». *ادب و زبان فارسی*، ۳، ۳: ۱۸۷-۲۰۲.
- 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.
- 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.
- Brants, S., Dipper, S., Hansen, S., Lezius, W., and Smith, G. (2002). "The TIGER treebank". *Proceedings of the workshop on treebanks and linguistic theories* (Vol. 168).



- 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)*.
- Dowty, David R. (1991). "Thematic proto-roles and argument selection". *Language*. 67(3):547–619.
- Dang, H. T., Kipper, K., Palmer, M., and Rosenzweig, J., (1998, August). "Investigating regular sense extensions based on intersective Levin classes". *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics-Volume 1* (pp. 293-299). Association for Computational Linguistics.
- Fillmore, C. J. (1976). "Frame semantics and the nature of language". *Annals of the New York Academy of Sciences*, Vol. 280, pp. 20-32.
- \_\_\_\_\_, and Atkins, B. S. (1998). "FrameNet and lexicographic relevance". *Proceedings of the First International Conference on Language Resources and Evaluation, Granada, Spain*. pp. 28-30
- \_\_\_\_\_, and Baker, C. F. (2001). "Frame semantics for text understanding". *Proceedings of WordNet and Other Lexical Resources Workshop*.
- Kipper, K., Dang, H. T., Schuler, W., & Palmer, M. (2000). "Building a class-based verb lexicon using TAGs". *TAG+5 Fifth International Workshop on Tree Adjoining Grammars and Related Formalisms*, Paris, France.
- Kawahara, D., Kurohashi, S., and Hasida, K. (2002). "Construction of a Japanese relevance-tagged corpus". *Proceedings of the 3rd International Conference on Language Resources and Evaluation (LREC-2002)*, pp. 2008–2013, Las Palmas, Canary Islands.
- Levin, B. (1993). *English Verb Classes and Alternations: A Preliminary Investigation*. University of Chicago Press, Chicago.
- Marcus, M. P., Marcinkiewicz, M. A., and Santorini, B. (1993). "Building a large annotated corpus of English: The Penn Treebank". *Computational Linguistics*. 19(2). pp. 313–330.
- Palmer, M., Kingsbury, P., and Gildea, D (2005). "The Proposition Bank: An Annotated Corpus of Semantic Roles". *Computational Linguistics*. 31 (1). pp. 71–106
- Rasooli, M. S., Kouhestani M., and Moloodi A. (2013). "Development of a Persian Syntactic Dependency Treebank". *Proceedings of the 2013 Conference of the North*

- American Chapter of the Association for Computational Linguistics: Human Language Technologies*. pp. 306–314.
- Taulé, M., Martí, M. A., & Recasens, M. (2008). “AnCora: Multilevel Annotated Corpora for Catalan and Spanish”. *Proceedings of the Sixth International Language Resources and Evaluation (LREC-08)*. pp. 96–101.
- Xue, N., Xia, F., Chiou, F. D., & Palmer, M. (2005). “The Penn Chinese TreeBank: Phrase structure annotation of a large corpus”. *Natural language engineering*. 11(02). pp. 207-238.
- , and Palmer, M., (2009). “Adding semantic roles to the Chinese Treebank”. *Natural Language Engineering*. 15(1). pp. 143-172.

#### پی‌نوشت:

۱. تولید نخستین پیکره نقش‌های معنایی زبان فارسی با حمایت مرکز تحقیقات کامپیوتری علوم اسلامی صورت گرفته است.

2. parts of speech
3. Bi Jen Khan
4. PropBank
5. Palmer
6. FrameNet
7. Baker
8. VerbNet
9. Dang
10. Kipper
11. Penn Treebank
12. Marcus
13. Dowty
14. TIGER
15. Brants
16. SALSA
17. Burchardt
18. Xue
19. Taulé
20. Kawahara
21. Fillmore
22. argument structure

۲۳. این برجسب معرف موضوع‌هایی است که درباره وضعیت موقعیتی آنها صحبت می‌شود؛ بنابراین، این‌ها موضوع‌های ساکن یا در حرکت هستند.

24. CoNLL= Conference on Natural Language Learning
25. agentive
26. instrumental
27. dative
28. factitive
29. locative
30. objective
31. Levin
32. attribute
33. circumstantial adjunct
34. mood adjunct
35. textual adjunct
36. epistemic
37. deontic
38. evidential
39. relative clause link
40. sense

