

# شناسایی رابطه تقابل در گفتمان فارسی به کمک روش‌های یادگیری باسرپرستی

حبيب خدادادی<sup>۱</sup>، سعيد راحتی قوچانی<sup>۲</sup> و اعظم استاجی<sup>۳</sup>

<sup>۱</sup> گروه مهندسی کامپیوتر، واحد میناب، دانشگاه آزاد اسلامی، میناب، ایران

<sup>۲</sup> گروه مهندسی برق، واحد مشهد، دانشگاه آزاد اسلامی، مشهد، ایران

<sup>۳</sup> گروه زبان‌شناسی همگانی، دانشکده ادبیات و علوم انسانی، دانشگاه فردوسی مشهد، مشهد، ایران

## چکیده

گفتمان به هر قطعه یا پاره‌ای از زبان گفته می‌شود که به قصد برقراری ارتباط به کار برده شده باشد. سامانه شناسایی روابط موجود در گفتمان، روابط موجود بین واحدهای متنی یک گفتمان را شناسایی می‌کند. یکی از روابط موجود در گفتمان‌های زبان فارسی رابطه تقابل است که شناسایی آن به توانایی تولید و درک گفتمان کمک می‌کند و در سامانه‌های متعددی مانند خلاصه‌ساز، تفسیر و... کاربرد دارد. رابطه تقابل می‌تواند به کمک نقش‌نماهای خاص رابطه تقابل مانند «اما» و «ولی» شناسایی شود؛ اما در بعضی موارد این نقش‌نماها حذف می‌شوند و شناسایی رابطه را با مشکل مواجه می‌کنند و باید از ویژگی‌هایی مانند زمان فعل‌ها، جفت کلمات و... برای شناسایی استفاده کرد. برای انجام آزمایش‌ها، ۱۰۰۰۰ نمونه از رابطه تقابل و سایر روابط از پیکره متنی پژوهشکده هوشمند علائم به‌عنوان مجموعه داده گردآوری شد. از این نمونه‌ها، بردار ویژگی استخراج شد و در نهایت برای دسته‌بندی و تشخیص رابطه تقابل از چند روش یادگیری باسرپرستی شامل ماشین بردار پشتیبان، K نزدیک‌ترین همسایه، پنجره پارزن و همچنین ترکیب این دسته‌بندها استفاده شد. بهترین میزان صحت ۸۷/۱۳ و مربوط به ترکیب دسته‌بندها در بهترین حالت است.

واژگان کلیدی: پردازش زبان طبیعی، گفتمان، نقش‌نمای گفتمان، شناسایی رابطه تقابل، یادگیری باسرپرستی.

## ۱- مقدمه

گفتمان<sup>۱</sup> به هر قطعه یا پاره‌ای از زبان گفته می‌شود که به قصد برقراری ارتباط به کار برده شده باشد (مقدم کیا، ۱۳۸۴). گفتمان می‌تواند هر قطعه‌ای از کاربرد زبان را از اصوات گرفته تا مکالمه یا یادداشت کوتاه تا یک سخنرانی کامل یا نقد طولانی بر یک کتاب، شامل شود. تحلیل گفتمان به بررسی تمامی شیوه‌ها، راه‌بردها و عناصری اطلاق می‌شود که به گفتمان پیوستگی می‌بخشد؛ یعنی شناسایی و توصیف قواعد و سازوکارهایی که از ره‌گذر آنها کاربرد پاره‌ای از زبان دارای پیوستگی می‌شود و گویندگان زبان می‌توانند آن را به‌درستی تعبیر و درک کنند و نبودن آن سازوکارها موجب غیرقابل قبول یا قابل تعبیربودن آن پاره‌گفتار در یک موقعیت ارتباطی خاص می‌شود.

تفاوت اساسی بین مطالعه زبان در چارچوب زبان‌شناسی جمله‌بنیاد<sup>۲</sup> و تحلیل گفتمان، آن است که در اولی زبان به مثابه یک نظام انتزاعی و خودسازمان‌دهنده مطالعه می‌شود و چگونگی کارکرد قواعد آن بررسی و توصیف می‌شود و به‌طور معمول داده‌هایی که تحلیل‌گر برای شناسایی و توصیف واحدها و قواعد زبان به‌کار می‌برد یا گزینشی یا تحلیل‌گراند براساس شم زبانی خود تولید می‌کند (ذوقدار مقدم و دبیر مقدم، ۱۳۸۱)؛ اما در تحلیل گفتمان، زبان در کاربرد عملی و واقعی آن در بافت موقعیت یعنی با توجه به کارکردهایش در برقراری ارتباط و انتقال پیام مطالعه می‌شود؛ بنابراین داده‌های مورد استفاده نیز داده‌هایی طبیعی هستند که گویندگان زبان در موقعیت‌های مختلف تولید کرده‌اند (مقدم کیا، ۱۳۸۴).

<sup>2</sup>Sentence-based

<sup>1</sup>Discourse

نقش‌نماهای گفتمان<sup>۱</sup> عناصری در سطح گفتمان هستند که کارکرد آن‌ها نشان‌دادن رابطه‌ای است که بین یک پاره‌گفتار و پاره‌گفتار قبلی وجود دارد (ذوقدار مقدم و دبیر مقدم، ۱۳۸۱). از جمله این عناصر در زبان انگلیسی می‌توان «And»، «Or»، «But»، «Well»، «Because» و ... و در زبان فارسی «اما»، «ولی»، «خب»، «حالا»، «پس»، «بنابراین»، و ... را نام برد. در صورت وجود نقش‌نماها، از آنها می‌توان به‌عنوان عامل شناسایی نوع رابطه استفاده کرد؛ ولی در مواردی این عناصر وجود ندارند. در پژوهش‌های گسترده‌ای که به‌خصوص در زبان انگلیسی صورت گرفته است، با استخراج ویژگی‌هایی سعی در شناسایی این روابط در سطح گفتمان شده است؛ در زبان فارسی تنها موارد (خدادادی و دیگران، ۱۳۸۹ الف؛ خدادادی و دیگران، ۱۳۸۹ ب) تاکنون ارائه شده است که در (خدادادی و دیگران، ۱۳۸۹ الف) نویسندگان ضمن بررسی متون فارسی چندین ویژگی را برای شناسایی رابطه تقابل در زبان فارسی ارائه داده‌اند که در این مقاله نیز از این ویژگی‌ها استفاده خواهیم کرد. در (خدادادی و دیگران، ۱۳۸۹ ب) نویسندگان سامانه‌ی هوشمند را جهت شناسایی رابطه تقابل در زبان فارسی طراحی کرده‌اند که بیش از ۶۰ درصد صحت وجود دارد.

تعیین رابطه‌های گفتمانی بین قسمت‌های دلخواه متن، کاربردهای گسترده‌ای دارد. از جمله اینکه شناسایی این روابط به توانایی تولید و درک گفتمان کمک می‌کند و در سامانه‌های پرسش و پاسخ و خلاصه‌ساز کاربرد دارند. همچنین با شناسایی روابط موجود در یک گفتمان، به ساخت یک تجزیه‌گر گفتمان و در نتیجه افزایش کارایی در خلاصه‌سازی متون کمک می‌کند و در سامانه‌های پرسش و پاسخ ما را قادر می‌سازد تا به پرسش‌های پیچیده‌ای مانند what contradicts Y? یا what were the causes of X? پاسخ دهیم.

در (مارکو و اجیهای، ۲۰۰۱) با جمع‌آوری نمونه‌های زیادی از چهار رابطه گفتمانی و سپس حذف نقش‌نماهای گفتمان و با فرض امکان تعیین نوع رابطه از روی کلمات موجود، یک مدل احتمالاتی از جفت کلمات دو طرف رابطه را به دست آورده (هر جفت واژه شامل یک واژه قبل و یک واژه بعد از نقش‌نماست) و سپس از این مدل و با داشتن کلمات دو طرف هر رابطه گفتمانی و بدون استفاده از نقش‌نما،

احتمال تعلق به هر کدام از رابطه‌ها را به دست آورده و بیشترین احتمال نوع رابطه را تعیین می‌کند.

در (سایتو و دیگران، ۲۰۰۶) که بر روی زبان ژاپنی کار شده است، علاوه بر استفاده از جفت کلمات از اطلاعات الگوهای عبارتی<sup>۲</sup> نیز استفاده شده است. این الگوها قسمت‌هایی از جمله هستند که در بعضی روابط ظاهر می‌شوند. به‌عنوان مثال رابطه تضاد در زبان انگلیسی در مواردی به صورت "... did ..." And "... should have done ..."

ظاهر می‌شود. با استفاده از این اطلاعات اقدام به شناسایی روابط شده است. در (پیتلر و دیگران، ۲۰۰۹) فقط به بررسی موارد بدون نقش‌نما در مجموعه برچسب‌خورده گفتمانی به زبان انگلیسی PDTB<sup>۳</sup>، که با کشف موارد حذف نقش‌نما، ویژگی‌هایی مانند وجود قطب‌های مختلف کلمات، طبقات فعلی و جفت کلمات استخراج شده و به کمک آن‌ها اقدام به دسته‌بندی شده است.

در (ژو و دیگران، ۲۰۱۰) که از مجموعه داده PDTB استفاده شده است، علاوه بر ویژگی‌های مقالات قبلی ویژگی جدیدی معرفی شده است که در مواردی که نقش‌نما وجود ندارد، با استفاده از یک مدل زبانی ابتدا نقش‌نما پیش‌گویی و سپس از این نقش‌نما به‌عنوان یک ویژگی اضافی برای شناسایی نوع رابطه استفاده می‌شود. در (پیتلر و ننکووا، ۲۰۰۹) برای حل مشکل نقش‌نماهای مشترک بین چند رابطه و کلماتی که در مواردی نقش‌نما هستند و در مواردی نیستند، استفاده از ویژگی‌های حاصل از درخت نحو پیشنهاد شده است. به این ترتیب که ابتدا درخت نحو رابطه را شکل داده و سپس با استخراج ویژگی‌هایی از این درخت، در مورد نقش‌نمای آن قضاوت می‌شود.

## ۲- رابطه‌های گفتمانی و بازشناسی آنها

### ۲-۱- روابط گفتمان و رابطه تقابل

در هر گفتمان، روابطی بین واحدهای متنی وجود دارد که در حالت عادی نوع این روابط به کمک نقش‌نماهای گفتمان تشخیص داده می‌شود. تشخیص ماهیت این روابط و ارائه تعریفی از آنها همواره مورد علاقه بوده است و پژوهش‌های متعددی برای شناسایی این روابط در سطح گفتمان انجام شده است؛ اما هنوز برنامه رایانه‌ای قوی برای شناسایی این روابط در سطح گفتمان وجود ندارد. در اینجا برای

<sup>2</sup> Phrasal Pattern

<sup>3</sup> Penn Discourse Treebank

<sup>1</sup> Discourse Marker

برای رابطه علیت<sup>۲</sup> به کار می‌رود (پیتلر و ننکووا، ۲۰۰۹).

- در خیلی از موارد نقش‌نماها حذف می‌شوند (در مثال ۴ نمونه‌ای نشان داده شده است). (لین و دیگران، ۲۰۰۹). بنابراین در حالتی که هیچ‌یک از سه مشکلات بالا برای یک نقش‌نما خاص اتفاق نیفتد، تنها با نقش‌نما می‌توان نوع رابطه را شناسایی کرد؛ اما در حالتی که یکی از مشکلات بالا برای یک نقش‌نمای خاص اتفاق بیفتد باید از ابزارهای واژگانی، نحوی و معنایی برای این منظور استفاده کرد. به‌طور مشابه، همین وضعیت در زبان فارسی نیز وجود دارد. از نظر مناسبات معنایی در رابطه تقابل در سطح گفتمان بین بخش نخست و بخش دوم به‌نوعی رابطه‌ای مغایر وجود دارد. این رابطه در سطح گفتمان به‌وسیله نقش‌نماهای گفتمان مخصوص رابطه تقابلی از قبیل اما، ولی، لکن، باین‌همه، باین‌حال، علی‌رغم، باوجوداین، از سوی دیگر، بالعکس، مع‌هذا، مع‌الوصف، لیکن شناخته می‌شوند. مثال‌هایی از بعضی از این نقش‌نماها را در جدول (۱-۲) مشاهده می‌کنید. این نقش‌نماها می‌توانند در مواردی که خواننده (شنونده) از روی معنا به نوع رابطه پی می‌برد، حذف شوند. در مثال‌های ۵ و ۶ دو نمونه رابطه تقابلی نشان داده شده است که اولی دارای نقش‌نمای تقابلی «ولی» است؛ اما دومی فاقد نقش‌نماست.

#### مثال ۵:

البته تمام اقلام غذایی ذکرشده برای همه مبتلایان به میگرن اثر نامطلوب ندارند؛ ولی در تعداد زیادی از بیماران مؤثر هستند.

#### مثال ۶:

این دوربین ابتدا برای تشخیص پرتوهای کیهانی طراحی شده بود، در حال حاضر با تغییراتی که در آن ایجاد شده است، مصارف پزشکی نیز پیدا کرده است.

بنابراین و با توجه به مشکل حذف نقش‌نمای تقابلی، از روش‌های دیگری به‌جز استفاده از نقش‌نما برای شناسایی این رابطه استفاده می‌کنیم تا در موقع حذف نیز مشکلی برای شناسایی وجود نداشته باشد. سامانه پیشنهادی در شکل (۱-۲) نشان داده شده است.

در این سامانه دو واحد متنی به‌صورت بند، جمله یا پاراگراف به‌عنوان ورودی ارائه می‌شود و سامانه وجود رابطه

روشن‌شدن مطلب، مثال‌هایی از روابط گفتمانی در زبان انگلیسی و فارسی آورده شده است.

#### مثال ۱:

- (1a) Such standards would preclude arms sales to states like Libya, which is also currently subject to a U.N. embargo.
- (1b) But states like Rwanda before its present crisis would still be able to legally buy arms.
- (2a) South Africa can afford to forgo sales of guns and grenades.
- (2b) because it actually makes most of its profits from the sale of expensive, high-technology systems like laser-designated missiles, aircraft electronic warfare systems, tactical radios, anti-radiation bombs and battlefield mobility systems.

در مثال ۱ وجود نقش‌نمای گفتمان But در پاره‌گفتار (۱) بین قسمت a و b، نوع رابطه را CONTRAST و در پاره‌گفتار (۲) وجود نقش‌نمای because نوع رابطه را EXPLANATION-EVIDENCE تعیین می‌کند.

#### مثال ۲:

- (2a) Selling picked up as previous buyers bailed out of their positions and aggressive short sellers—anticipating further declines—moved in.
- (2b) my favorite colors are blue and green.

#### مثال ۳:

- (3a) there have been more than 100 mergers and acquisitions Within the European paper industry since the most recent wave of friendly takeovers was completed in the U.S. in 1986.
- (3b) It was a far safer deal for lenders since NWA had a healthier cash flow and more collateral on hand.

#### مثال ۴:

“A lot of investor confidence comes from the fact that they can speak to us,” he says.  
“To maintain that dialogue is absolutely crucial.”

تشخیص نوع رابطه به‌کمک نقش‌نما با سه مشکل مواجه است:

- بعضی از نقش‌نماها دارای کارکرد غیر نقش‌نمایی هم می‌باشند؛ مانند مثال ۲ قسمت 2b که واژه and نقش‌نمای گفتمان نیست؛ ولی در 2a نقش‌نمای گفتمان می‌باشد (پیتلر و ننکووا، ۲۰۰۹).
- بعضی از نقش‌نماها بین دو یا چند رابطه مشترک هستند و تشخیص رابطه با ابهام مواجه می‌شود؛ مانند مثال ۳ نقش‌نمای since هم برای رابطه زمانی<sup>۱</sup> و هم

<sup>2</sup>Causality

<sup>1</sup>Temporal

تقابل بین این دو ورودی را با استخراج ویژگی‌هایی (بخش ۴) و با توجه به آموزش‌هایی که در مرحله آموزش دیده است (سیستم از نوع آموزش با سرپرستی استفاده می‌کند)، بررسی می‌کند.

(جدول ۲-۱) مثال‌هایی برای بعضی نقش‌نماهای تقابل.

نقش‌نمای تقابل	مثال
ولی	برآورد مبالغ غیر قابل وصول و اندازه‌گیری درآمد احتمال برابری میان مبالغ مطالباتی که هرساله غیر قابل وصول می‌گردد و مبلغی که به‌عنوان ذخیره برای همین رویداد منظور می‌شود فوق‌العاده ضعیف است؛ ولی تفاوت‌های میان این دو رقم اگر به نسبت اندک باشد، قابل چشم‌پوشی است.
لیکن	یک بار دیگر به بهانه مذاکره با کاسترو درباره خلیج خوک‌ها لباس آغشته به سم را به او هدیه کردند؛ لیکن کاسترو موضوع را با دیده شک نگریست.
باین‌حال	در کانادا و آمریکا تنبیه بدنی کودکان توسط والدین آن‌ها تنها در شرایط خاصی جرم محسوب نمی‌شود. باین‌حال چنین استثنایی نیز به‌تدریج در حال رنگ باختن است و گروه‌های حمایت از کودکان درصدد هستند این موارد را نیز غیرقانونی اعلام کنند.

پیشوندهای یک واژه حتی‌الامکان به‌صورت جدا از هم و به‌صورت نیم‌فاصله نوشته می‌شود؛ ولی عدم رعایت این رسم‌الخط باعث می‌شود که حدومرز یک واژه را نتوان به درستی تشخیص داد و همچنین تشخیص کلمات مرکب نیز به سختی صورت می‌گیرد. بسیاری از واقع فاصله‌گذاری‌ها در متون فارسی رعایت نمی‌شوند. گاهی اوقات وسط واژه‌ای یک فاصله قرار می‌گیرد این در حالی است که حتی ممکن است بین دو واژه مجزا فاصله قرار نگیرد. این مسائل علاوه بر اینکه شناسایی حدود واژه را مشکل می‌کند در مواردی مانند استفاده از جفت‌کلمات و ویژگی‌های لغوی مشکل‌ساز خواهند بود.

از دیگر مسائل، وجود حروف عربی هستند که وارد زبان فارسی شده‌اند مثل همزه یا تنوین که موجب شده‌اند برخی کلمات به چند شکل متفاوت نوشته شوند. نمونه این مشکل در کلماتی مثل «مسائل» یا «مسایل» دیده می‌شود. همچنین حروف «ک» و «ی» دارای دو یونیکد فارسی و عربی هستند که بعضاً به‌جای هم استفاده می‌شوند و برخی از کلمات فارسی دارای دو املا متفاوت مانند اتاق و اطاق هستند.

از دیگر مشکلات مهم و اساسی در این تحقیق، نبود ابزارهای مناسب و با دقت بالاست. برای کار بر روی بسیاری از سامانه‌های حوزه پردازش زبان‌های طبیعی، نیاز به ابزارهای از قبل آماده داریم تا در گام‌های استخراج ویژگی مورد استفاده قرار گیرند که در زبان فارسی محققان با مشکلات زیادی با آن مواجه هستند. از جمله این ابزارها می‌توان به سامانه‌های ایجاد کننده درخت نحو و درخت وابستگی، برچسب‌زن مقوله نحوی، تجزیه‌گر سطحی، شبکه معنایی واژگان (مانند wordnet) فارسی و ... اشاره کرد.



(شکل ۲-۱): سامانه پیشنهادی.

## ۲-۲- مشکلات بازشناسی روابط گفتمان در زبان فارسی

یکی از بزرگ‌ترین موانع پیش روی پردازش رایانه‌ای گفتمان، نبود داده برچسب‌خورده مخصوص گفتمان جهت آموزش سامانه است که ایجاد چنین داده‌ای خود امری مهم و وقت‌گیر است. در زبان انگلیسی چندین مجموعه داده مخصوص روابط گفتمان و پردازش‌های گفتمانی وجود دارد که بزرگ‌ترین آن داده (پرساد و دیگران، ۲۰۰۸) PDTB است.

مشکل عمده دیگری که در شناسایی روابط گفتمان وجود دارد، مربوط به رسم‌الخط زبان فارسی است. در رسم‌الخطی که توسط فرهنگستان ادب و زبان فارسی مورد تأیید و تصویب قرار گرفته است؛ اغلب پسوندها و

## ۳- معرفی مجموعه داده

اگر بخواهیم سامانه شناسایی و طبقه‌بندی روابط گفتمان را به روش با سرپرستی آموزش دهیم نیاز به یک مجموعه داده برچسب‌خورده داریم. در اینجا داده‌های آموزشی از متون برچسب‌خورده پیکره متنی پژوهشکده پردازش هوشمند علائم استخراج شده است. (بیجن‌خان و دیگران، ۲۰۱۱).

این پیکره حجم انبوهی از متون الکترونیکی فارسی است که برحسب نوع و موضوع از منابع معتبر استخراج شده است. در این پیکره تنوع داده‌های زبانی و موضوعات مختلف وجود دارد و نماینده کاملی از جامعه آماری متون فارسی به

وجود یا عدم وجود فعل‌ها و واژه‌های منفی: از آنجاکه در عبارات تقابلی ممکن است کلمات و فعل‌های مثبت و منفی در دو طرف نقش‌نما ظاهر شوند، این ویژگی عبارت است از وجود فعل‌ها و واژه‌های منفی در یک طرف نقش‌نما و عدم وجود آن در طرف دیگر. در مثال ۷ نمونه‌ای آورده شده است.

#### مثال ۷:

یکی از بچه‌ها به نام طبری بسیار خوش‌رو، مخلص و پاک بود و اصلاً اهل غیبت و تهمت نبود.

**وجود اعداد و درصد در دو طرف رابطه:** اگر هر دو Arg1 و Arg2 شامل عدد، درصد و یا مقدار عددی باشد، به احتمال یک رابطه مقایسه‌ای وجود دارد (پیتلر و دیگران، ۲۰۰۹). منظور از Arg2 عبارت و جملاتی است که نقش‌نما به‌طور ساختاری در آن طرف قرار دارد و طرف دیگر به‌عنوان Arg1 شناخته می‌شود (در صورتی که نقش‌نما وجود نداشته باشد، قسمت نخست رابطه را Arg1 و قسمت دوم رابطه را Arg2 می‌نامیم).

**استفاده از جفت کلمات:** اگر رابطه تقابل بین دو قسمت متنی برقرار باشد و W1 و W2 به تمام کلمات این دو قسمت اشاره کند، فرض ما این است که جفت کلمات  $(w_i, w_j) \in W1 \times W2$  (کلمات دو طرف رابطه که به‌صورت جفت‌هایی در نظر گرفته می‌شوند) می‌تواند وجود رابطه تقابل را نشان دهد. در این راستا و به‌منظور شناسایی این نوع جفت کلمات کارهای زیر انجام شد:

۱-۴ از پیکره متنی تمام نمونه‌های رابطه تقابل که دارای یکی از نقش‌نماهای «ولی»، «اما»، «لکن»، «علی‌رغم»، «بالبین حال»، «بالبین همه» و «لیکن» هستند، استخراج شدند که کلمات قبل از آن به‌عنوان Arg1 و کلمات بعد از آن به‌عنوان Arg2 در نظر گرفته شد. (در مورد نقش‌نمای گفتمان «علی‌رغم» کلمات بعد از آن تا «» به‌عنوان Arg1 و کلمات پس از «» به‌عنوان Arg2 در نظر گرفته شدند).

۲-۴ این نقش‌نماها و هر واژه‌ای که نقشی به‌جز فعل، اسم، صفت یا قید دارد، حذف شدند.

۳-۴ هر واژه از pos آن جدا شد.

۴-۴ هر واژه در Arg1 را با تک‌تک کلمات Arg2 به‌صورت یک جفت مجزا در نظر گرفته و در تمام نمونه‌ها تعداد این جفت‌ها شمرده می‌شود.

حساب می‌آید. در این متون، کلمات یک جمله با توجه به جایگاه نحوی آن کلمه به مقوله‌های نحوی مختلف تقسیم شده‌اند و برای هر کلمه، ویژگی‌هایی بیان شده است. به‌عنوان مثال برای افعال، ویژگی‌هایی مانند زمان فعل، مرکب یا ساده‌بودن، شخص فعل و ... ذکر شده است و برای اسامی، ویژگی‌هایی مانند خاص یا عام‌بودن، جمع یا مفردبودن ذکر شده است. این پیکره صدمیلیونی نزدیک به ۱۰۰ میلیون واژه برچسب خورده دارد. ما از این پیکره ۸۸۰۰ نمونه از رابطه تقابل و سایر رابطه‌ها استخراج کردیم که ۵۰۰۰ نمونه رابطه تقابل و ۳۸۰۰ نمونه سایر رابطه‌ها بود. (جدول (۱-۳)) علاوه بر این تعداد ۱۲۰۰ جفت جملاتی را که ما فرض می‌کنیم رابطه‌ای بین آنها نیست، استخراج شدند. این جفت‌جملات هم‌جوار نیستند و هر کدام را به‌طور تصادفی از متون پیکره استخراج کردیم.

در مورد نقش‌نماهایی مانند «و» که کارکرد غیرنقش‌نمایی نیز دارند، دقت لازم در تهیه نمونه‌ها صورت گرفته است تا نمونه‌هایی جمع‌آوری شود که کارکرد نقش‌نمای مورد نظر را داشته باشند.

(جدول ۱-۳): نقش‌نماهای مورد استفاده

نقش‌نما	نوع رابطه	تعداد	نقش‌نما	نوع رابطه	تعداد
اما	تقابل	۲۹۰۰	زیرا	علی	۶۰۰
ولی	تقابل	۱۶۰۰	یعنی	افزایشی	۱۰۰۰
لکن	تقابل	۵۰	پس	زمانی	۴۰۰
لیکن	تقابل	۲۰۰	و	افزایشی	۵۰
بالبین حال	تقابل	۲۵۰	همچنین	افزایشی	۵۰۰
			دراین میان	زمانی	۵۰
			دراین حال	زمانی	۱۰۰
			چون	علی	۱۵۰
			بنابراین	علی	۲۰۰
			به علاوه	افزایشی*	۵۰
			بعد	افزایشی	۴۰۰
			آنگاه	زمانی	۱۵۰
			علاوه بر این	افزایشی	۱۵۰

\* (مقدم کیا، ۱۳۸۴)

## ۴- استخراج ویژگی برای سامانه

### شناسایی رابطه تقابل

با توجه به محدودیت‌ها و مشکلاتی که برای تشخیص و شناسایی روابط گفتمان در زبان فارسی بیان شد، ویژگی‌های زیر جهت تشخیص و شناسایی رابطه تقابل پیشنهاد می‌شود:

۴-۵) تمامی جفت‌کلماتی را که زیر چهار بار تکرار شده است حذف و باقی جفت‌کلمات را به همراه تعداد آن ذخیره می‌شود.

مراحل ۴-۱ تا ۴-۵ را برای نقش‌نماهای «زیرا»، «یعنی»، «پس»، «همچنین»، «دراین‌میان»، «دراین‌حال»، «چون»، «بنابراین»، «به‌علاوه»، «بعد»، «آنگاه» و «علاوه‌براین» که دربرگیرنده بعضی روابط مهم دیگر در گفتمان فارسی هستند، تکرار می‌کنیم.

پس از انجام کارهای گفته‌شده و در حضور یک نمونه، ابتدا جفت‌کلمات آن را استخراج کرده و سپس به‌ازای هر جفت مقدار  $Nc/(Nc+No)$  را محاسبه می‌کنیم ( $Nc$  تعداد این جفت در کل نمونه‌های تقابل و  $No$  تعداد این جفت در کل نمونه‌های سایر روابط می‌باشد). مقدار به‌دست آمده را برای هر جفت با مقادیر به دست آمده برای سایر جفت‌کلمات جمع کرده و عدد حاصل تقسیم بر تعداد جفت‌های نمونه را به‌عنوان ویژگی دیگری در نظر می‌گیریم. طبیعی است که هرچه این عدد بزرگ‌تر باشد، احتمال اینکه نمونه از نوع تقابل باشد بیشتر است.

هم‌زمان بودن فعل‌های اصلی  $Arg1$  و  $Arg2$ : از آنجاکه در بعضی از رابطه‌ها مانند رابطه زمانی، زمان فعل‌های  $Arg1$  و  $Arg2$  ممکن است با هم تفاوت داشته باشد؛ هم‌زمان بودن فعل قبل از نقش‌نما و آخرین فعل بعد از نقش‌نما به‌عنوان ویژگی در نظر گرفته شده است.

**کلمات دوتایی پشت سرهم:** به‌منظور کشف کلمات دوتایی پشت سرهم که در رابطه تقابل بیشتر استفاده می‌شود، از جفت‌کلمات استفاده می‌کنیم؛ بدین منظور تمام کلمات دوتایی پشت سرهم را در  $Arg1$  و  $Arg2$  به‌ازای تمام نمونه‌های رابطه تقابل با نقش‌نماهای گفته‌شده و همچنین برای تمام نمونه‌های سایر روابط با نقش‌نماهای گفته‌شده، استخراج و به همراه تعداد آن ذخیره می‌کنیم؛ سپس و در حضور یک نمونه که قصد استخراج ویژگی برای آن داریم، مشابه استفاده از جفت‌کلمات و برای هر دو واژه پشت سرهم در نمونه، مقدار  $Nc/(Nc+No)$  ( $Nc$  تعداد کلمات دوتایی پشت سرهم در کل نمونه‌های تضاد و  $No$  تعداد کلمات دوتایی پشت سرهم در کل نمونه‌های سایر روابط می‌باشد) را محاسبه می‌کنیم و ویژگی را حاصل جمع تمام مقادیر محاسبه‌شده برای تمام کلمات دوتایی پشت سرهم، تقسیم بر تعداد کلمات دوتایی پشت سرهم موجود در نمونه قرار می‌دهیم.

**کلمات سه‌تایی پشت سرهم:** مشابه ویژگی قبلی، اما برای سه واژه پشت سرهم این کار را انجام می‌دهیم.

## ۵- روش‌های یادگیری با سرپرستی

به‌طورکلی در مبحث شناسایی الگو با جفت‌هایی به‌صورت  $\{X, \omega\}$  سروکار داریم که  $X$  بردار ویژگی است و  $\omega$  دسته هرکدام از بردارهای ویژگی است. در روش‌های یادگیری با سرپرستی هر دوی  $X$  و  $\omega$  را در اختیار داریم؛ اما در روش‌های بدون سرپرستی ما فقط  $X$  را داریم. در اینجا از سه روش ماشین بردار پشتیبان، دسته‌بندی‌کننده  $K$  نزدیک‌ترین همسایه و الگوریتم پنجره پارزن<sup>۱</sup> استفاده می‌کنیم که در ادامه به‌اختصار بررسی می‌شود.

ماشین بردار پشتیبان<sup>۲</sup> ( $O$ ) یکی از انواع روش‌های دسته‌بندی است که در هر لحظه توانایی دسته‌بندی دو طبقه را دارد. رویکرد SVM به این صورت است که در مرحله آموزش، سعی می‌شود مرز تصمیم‌گیری<sup>۳</sup> به‌گونه‌ای انتخاب گردد که حداقل فاصله آن با هر یک از دسته‌های موردنظر بیشینه شود. این نحوه انتخاب مرز براساس پیداکردن نقاطی برای محاسبه بردارهای پشتیبان انجام می‌شود. با فرض اینکه دسته‌ها به‌صورت خطی جدایی‌پذیر باشند، ابرصفحه‌هایی با حداکثر حاشیه<sup>۴</sup> برای تفکیک دسته‌ها به دست می‌آید. به‌طورکلی ابرصفحه‌های زیادی برای دسته‌بندی وجود دارد که در بین آنها SVM ابرصفحه-ای را پیدا می‌کند که فاصله بیشتری را بین دو دسته ایجاد کند (بورگر، ۲۰۰۸).

در بسیاری از کاربردهای واقعی ابرصفحه‌ای که بتواند به‌صورت خطی جداسازی را در دو دسته انجام دهد، وجود ندارد؛ برای حل این مشکل یک راه‌حل این است که داده‌ها از فضای اصلی به فضای دیگری (فضای با بعد بالاتر) انتقال دهیم و بعد ابرصفحه‌ای را تعیین نماییم که بتواند به‌صورت خطی این جداسازی را انجام دهد که در این حالت بار محاسباتی زیاد می‌شود؛ از این‌رو فرآیند نگاشت را توسط تابع هسته<sup>۵</sup> انجام خواهیم داد. تابع هسته تابعی است که الگوها را از یک فضا به یک فضای دیگر نگاشت می‌کند تا SVM بتواند با استفاده از یک ابرصفحه الگوها را دسته‌بندی کند.

<sup>۱</sup>Parzen Window

<sup>۲</sup>Support Vector Machine (SVM)

<sup>۳</sup>Decision Boundary

<sup>۴</sup>Maximum Margin

<sup>۵</sup>Kernel Function



تمام محاسبات مربوط به استخراج بردارهای ویژگی با نرم‌افزار Microsoft Visual Studio 2008 انجام شد و نتایج به نرم‌افزار MATLAB 7.8 منتقل شد؛ داده‌ها به دو دسته آموزش و آزمایش تقسیم کردیم. بر روی مجموعه آموزش که مجموعه بزرگ‌تری است (بین ۷۰ تا ۷۵ درصد داده‌ها را به بخش آموزش اختصاص می‌دهیم) سامانه را آموزش داده و با مجموعه آزمایش کارایی سامانه مورد ارزیابی قرار گرفت.

### ۱-۶- شناسایی رابطه تقابل

داده‌های آموزشی به وسیله سه الگوریتم گفته شده طبقه‌بندی می‌شوند. معیاری که برای سنجش میزان کارایی سامانه به کار رفته است میزان صحت سامانه در تشخیص درست داده‌های آزمون می‌باشد که عبارت از نسبت تعداد نمونه‌های درست بر کل نمونه‌های آزمون است.

در جدول (۱-۶)، نتایج میزان صحت دسته‌بند ماشین بردار پشتیبان را به‌ازای چند تابع هسته مختلف مشاهده می‌کنید. همچنین نتایج میزان صحت دسته‌بند، نزدیک‌ترین همسایه به‌ازای چند K مختلف و نتایج میزان صحت دسته‌بند پنجره پارزن به‌ازای چند h مختلف را مشاهده می‌کنید.

(جدول ۱-۶): میزان صحت دسته‌بندهای SVM، KNN و پنجره پارزن.

SVM		KNN		Parzen Window	
میزان	نوع هسته	میزان	مقدار K	میزان	مقدار h
۸۴/۷۶	خطی	۸۱/۶۶	۳	۸۰/۶۳	۰/۱۰
۸۴/۲۵	درجه ۲	۸۳/۶۸	۱۰	۸۴/۸۶	۰/۱
۸۴/۹۵	RBF	۸۵/۰۸	۳۰	۸۱/۸۳	۰/۲۵
		۸۳/۹۶	۵۰	۷۳/۹۲	۰/۵
		۸۳/۴۴	۸۰	۷۳/۰۷	۱

### ۲-۶- بررسی تأثیر ویژگی‌های انتخابی بر روی

#### شناسایی رابطه تقابل

به‌منظور بررسی تأثیر ویژگی‌های انتخابی بر روی نرخ شناسایی رابطه تقابل، آزمایش‌هایی صورت گرفت. در این آزمایش‌ها از دسته‌بند نزدیک‌ترین همسایگی با مقدار همسایگی برابر سی که دقت خوبی در شناسایی رابطه تقابل نشان داده است، استفاده شد.

در این آزمایش‌ها فضای ویژگی اولیه به زیرفضاهایی با ابعاد کمتر تقسیم می‌شود که همه حالات ممکن در نظر گرفته شده است، هر بار چهار ویژگی انتخاب شده و مجموعه

در دسته‌بندی به روش K نزدیک‌ترین همسایه<sup>۱</sup> می‌خواهیم یک بردار ویژگی نامعلوم x را در یکی از دسته شناخته شده دسته‌بندی کنیم. برای استفاده از دسته‌بندی به روش K نزدیک‌ترین همسایه ابتدا معیاری را برای اندازه‌گیری فاصله در نظر می‌گیریم (مانند فاصله اقلیدسی)؛ سپس به‌روش زیر عمل می‌کنیم:

- ❖ فاصله x را با بردارهای آموزشی حساب می‌کنیم؛ بدون در نظر گرفتن دسته آنها، K داده که نزدیک‌ترین فاصله را تا x دارند، انتخاب می‌کنیم. K فرد انتخاب می‌شود.
- ❖ از این K نمونه، تعداد اعضای هر دسته را مشخص می‌کنیم و تعداد اعضای  $w_i$  از این K نمونه را  $K_i$  می‌نامیم ( $i=1, \dots, c$ ). مشخص است که  $\sum_i k_i = k$ .
- ❖ x را در دسته‌ای دسته‌بندی می‌کنیم که بزرگ‌ترین  $K_i$  را داشته باشد.

این الگوریتم قادر به نمایش پیچیده‌ترین مرزهای تصمیم‌گیری است و فهم آن بسیار ساده است؛ ولی از نظر محاسبات خیلی کارآمد نیست.

در روش پنجره پارزن، به دور هر داده‌ای که قرار است دسته‌بندی شود یک فوق مکعب<sup>۲</sup>، d بعدی به حجم V و به مرکزیت آن داده در نظر می‌گیریم و در آن حجم داده‌های موجود را می‌شماریم. اگر  $h_n$  طول یک یال از این فوق مکعب باشد، حجم آن به صورت  $V_n = h_n^d$  محاسبه می‌شود. احتمال شباهت هر داده به تمام دسته‌ها محاسبه می‌شود و دسته‌ای که احتمال بیشتری داشته باشد، به‌عنوان دسته داده مورد نظر، پذیرفته می‌شود.

### ۶- دسته‌بندی داده‌ها

نقش‌ما را در نمونه‌های استخراج شده از پیکره پردازش هوشمند علائم که به‌صورت [ARG2] D\_Marker { [ARG1] بود حذف کردیم و تنها از دو قسمت [ARG1] و [ARG2] به‌عنوان منابع استخراج ویژگی استفاده کردیم. (در مورد ۱۲۰۰ جفت جمله بدون رابطه [ARG1] را جمله نخست و [ARG2] را جمله دوم در نظر گرفته شد). به‌ازای هر کدام از این ۱۰۰۰۰ نمونه، ۶ ویژگی گفته شده را استخراج کردیم. ۵۰۰۰ نمونه رابطه تقابل در دسته ۱ و داده‌های سایر روابط در دسته ۲ تقسیم‌بندی شد (۱۲۰۰ جفت جمله را نیز جزء داده‌های سایر روابط دسته‌بندی کردیم).

<sup>۱</sup>K Nearest Neighbors  
<sup>۲</sup> Hyper Cube

این چهار ویژگی به همراه دسته هر نمونه به دسته‌بند نزدیک‌ترین همسایه داده شد و نتایج به دست آمده و مورد ارزیابی قرار گرفت. نتایج حاصل از این آزمایش در جدول (۲-۶) مشاهده می‌شود.

ویژگی‌ها به همان ترتیبی که در بخش ۴ مقاله معرفی شده در نظر گرفته شده است. نتایج دسته‌بندی کننده نشان می‌دهد در حالتی که ویژگی‌ها شامل جفت کلمات و کلمات دوتایی پشت سرهم نیست، درصد صحت خیلی پایین‌تر و در حالتی که از این دو ویژگی استفاده شده است درصد صحت بالاتر می‌رود؛ که در این بین تأثیر ویژگی جفت‌واژه بیشتر از دیگری است؛ بنابراین مؤثرترین ویژگی به حساب می‌آید.

جدول ۲-۶: استفاده از ۴ ویژگی به منظور شناسایی رابطه تقابل

درصد صحت	واژه‌های منفی	وجود اعداد	جفت کلمات	هم‌زمانی فعل‌ها	کلمات دوتایی	کلمات سه تایی
79.752	√	√	√	√	-	-
84.536	√	√	√	-	√	-
81.08	√	√	√	-	-	√
73.424	√	√	-	√	√	-
56.008	√	√	-	√	-	√
74.08	√	√	-	-	√	√
84.816	√	-	√	√	√	-
80.816	√	-	√	√	-	√
83.864	√	-	√	-	√	√
73.872	√	-	-	√	√	√
84.144	-	√	√	√	√	-
81.36	-	√	√	√	-	√
84.168	-	√	√	-	√	√
73.112	-	√	-	√	√	√
84.664	-	-	√	√	√	√

### ۳-۶- دسته‌بندی داده‌ها به وسیله دسته‌بندی

#### کننده ترکیبی

ترکیب طبقه‌بندی کننده‌ها، مجموعه‌ای از طبقه‌بندی کننده‌ها است که در آن تصمیمات منفرد هر یک از آنها به روش‌های مختلف با یکدیگر ترکیب می‌شوند. در این قسمت از یک طبقه‌بند ترکیبی جهت آموزش سامانه خود استفاده می‌کنیم. در این طبقه‌بند ترکیبی از سه سناریو استفاده می‌شود. در سناریوی اول داده‌های آموزشی به ۳ دسته تقسیم می‌شود و سپس با هر کدام، یک طبقه‌بندی کننده آموزش می‌بیند و در نهایت خروجی این طبقه‌بندی کننده‌ها با روش رأی‌گیری اکثریت تعیین می‌شوند؛ یعنی اگر فرض

کنیم  $s^1$ ،  $s^2$  و  $s^3$  خروجی ۳ طبقه‌بندی کننده موجود برای طبقه‌بندی نمونه  $X$  باشند، در این صورت برچسبی به عنوان برچسب نهایی انتخاب می‌شود که دارای بیشترین تعداد رأی در بین خروجی طبقه‌بندی کننده‌ها باشد.

در سناریوی دوم از نتایج به دست آمده در جدول (۱-۶) استفاده می‌کنیم و فضای ویژگی اولیه را به زیرفضاهای کوچک‌تر چهارتایی تقسیم می‌کنیم. پنج زیرمجموعه از ویژگی‌های چهارتایی را که بیشترین درصد صحت را ایجاد کرده‌اند، انتخاب می‌کنیم و برای هر کدام از این ویژگی‌ها یک دسته‌بند نزدیک‌ترین همسایگی در نظر می‌گیریم؛ سپس دسته‌بندها را با داده‌های آموزش، آموزش داده و در نهایت یک دسته‌بند ترکیبی از طریق این پنج دسته‌بند ایجاد می‌کنیم؛ در این روش نیز به کمک سازوکار رأی‌گیری اکثریت خروجی طبقه‌بندی کننده ترکیبی مشخص می‌شود.

در سناریوی سوم، سناریوی دوم را با دسته‌بندهای متفاوت در نظر می‌گیریم؛ نتایج کار در جدول (۳-۶) نشان داده شده است.

جدول ۳-۶: میزان صحت دسته‌بندهای ترکیبی.

سناریو	درصد صحت
سناریوی ۱	۸۵/۲۰
سناریوی ۲	۸۷/۱۳
سناریوی ۳	۸۵/۸۰

### ۷- نتیجه‌گیری و کارهای آینده

در این مقاله ویژگی‌هایی برای شناسایی رابطه تقابل در گفتمان زبان فارسی ارائه شد. از آنجاکه در رابطه تقابل ممکن است در مواردی نقش‌نما حذف شود، بنابراین سامانه‌ای مستقل از نقش‌نما ارائه شد که با استخراج ویژگی‌هایی سعی در شناسایی رابطه تقابل دارد. نتایج دسته‌بندهای SVM و KNN و پنجره پارزن بر روی ویژگی‌های استخراج شده، امیدوارکننده بود و در بهترین حالت بیشتر از ۸۷ درصد صحت داشت. به‌عنوان پیشنهاد کار برای آینده می‌توان ویژگی‌های بهتری را استخراج کرد، مانند پیدا کردن کلمات متضاد در دو طرف رابطه و یا پیدا کردن الگوها و عباراتی که نقش‌نما نیستند؛ اما در رابطه تقابل ممکن است به کار گرفته شوند. همچنین به‌علت نبود مجموعه‌داده مخصوص روابط گفتمان، ایجاد چنین داده‌ای از نیازهای اولیه پردازش‌های گفتمانی در زبان فارسی است؛ که با استفاده از آن می‌توان



Pitler, Emily, Louis, Annie, and Nenkova, Ani, "Automatic sense prediction for implicit discourse relations in text", Proceedings of the Joint Conference of the 47th Annual Meeting of the Association for Computational Linguistics and the 4th International Joint Conference on Natural Language Processing of the Asian Federation of Natural Language Processing (ACL-IJCNLP), 2009, pp 683-691.

Prasad, Rashmi, Dinesh, Nikhil, Lee, Alan, Miltsakaki, Eleni, Robaldo, Livio, Joshi, Aravind, Webber, Bonnie, "The Penn Discourse Treebank 2.0", 6<sup>th</sup> International Conference on Language Resources and Evaluation, 2008.

Saito, Manami, Yamamoto, Kazuhide, Sekine, Satoshi, "Using Phrasal Patterns to Identify Discourse Relations", Human Language Technology Conference of the North American Chapter of the ACL, New York, USA, 2006, 133-136.

Zhou, Zhi-Min, Lan, Man, Yu, Xu, Niu, Zhenguy, Su, Jian and Tan, Chew Lim, "Predicting discourse connectives for implicit discourse relation recognition", 23rd International Conference on Computational Linguistics (COLING 2010), Beijing, China, 2010, pp 343-351.



#### حبیب خدادادی مدرک کارشناسی

خود را در رشته کامپیوتر-نرم‌افزار در سال ۱۳۸۷ از دانشگاه بیرجند و مدرک کارشناسی ارشد خود را در رشته کامپیوتر-هوش مصنوعی در سال

۱۳۸۹ از دانشگاه آزاد اسلامی واحد مشهد دریافت کرده است. وی هم‌اکنون عضو هیأت علمی دانشگاه آزاد اسلامی واحد میناب است. زمینه‌های پژوهشی مورد علاقه وی پردازش زبان طبیعی، رمزنگاری و پردازش تصویر است.

نشانی رایانامه ایشان عبارت است از:

habibekhodadadi@gmail.com



#### سعید راحتی قوچانی دانش‌آموخته

کارشناسی الکترونیک سال ۱۳۶۹ دانشکده فنی دانشگاه تهران و کارشناسی ارشد مخابرات ۱۳۷۲

دانشگاه آزاد اسلامی تهران جنوب و دکترای مخابرات ۱۳۷۷ دانشگاه آزاد اسلامی واحد علوم و تحقیقات است. وی پژوهش‌گر برتر جشنواره فردوسی

پژوهش‌های متعددی را در پردازش رایانه‌ای گفتمان برای زبان فارسی انجام داد.

## ۸- مراجع

مقدم‌کیا، رضا، "بعد، نقش‌نمای گفتمان در زبان فارسی"، نامه فرهنگستان، ۱۳۸۴، ۲۳، ۸۱-۹۸.

ذوقدارمقدم، رضا و دبیرمقدم، محمد، "نقش‌نامه‌های گفتمان مقایسه نقش‌نمای اما در زبان انگلیسی با نقش‌نمای اما در زبان فارسی"، پژوهش زبان‌های خارجی، ۱۳۸۱، ۱۲، ۵۵-۷۶. خدادادی، حبیب، راحتی قوچانی، سعید و استاجی، اعظم، "استخراج ویژگی‌هایی برای شناسایی رابطه تقابل در حالت حذف نقش‌نمای گفتمان فارسی"، شانزدهمین کنفرانس بین‌المللی سالانه انجمن کامپیوتر ایران، تهران، دانشگاه صنعتی شریف، ۱۳۸۹.

خدادادی، حبیب، راحتی قوچانی، سعید و استاجی، اعظم، "بازشناسی رابطه تقابل در زبان فارسی در حالت حذف نقش‌نمای گفتمان به کمک ماشین بردار پشتیبان"، اولین کنفرانس ملی محاسبات نرم و فن‌آوری اطلاعات، ماهشهر، دانشگاه آزاد اسلامی واحد ماهشهر، ۱۳۸۹.

Bijankhan, Mahmood, Sheykhzadegan, Javad, Bahrani, Mohammad And Ghayoomi, Masood, Lessons from Building a Persian Written Corpus: Peykare. Language Resources and Evaluation, 2011, 45(2), 143-164.

Burger, Christopher, "A Tutorial on Support Vector Machines for Pattern Recognition", Bell Labs/Lucent, 1998, 955-974.

Lin, Ziheng, Kan, Min-Yen and Ng, Hwee Tou, "Recognizing Implicit Discourse Relations in the Penn Discourse Treebank", Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing (ACL and AFNLP), 2009, pp 343-351.

Marcu, Daniel and Echiabi, Abdessamad, "An unsupervised approach to recognizing discourse relations", In Proceedings of the 40th Annual Meeting on Association for Computational Linguistics, 2001, pp 368-375.

Pitler, Emily and Nenkova, Ani, "Using Syntax to Disambiguate Explicit Discourse Connectives Text", Proceedings of the ACL-IJCNLP Conference Short Papers, 2009.

۱۳۷۹، رتبه نخست پژوهش سال ۱۳۸۶ و رتبه دوم پژوهش سال ۱۳۸۵ دانشگاه آزاد اسلامی واحد مشهد شده است. وی از سال ۱۳۷۸ به‌عنوان استادیار مخابرات دانشگاه آزاد اسلامی مشهد مشغول به کار است. همچنین ایشان از سال ۱۳۹۰ به‌عنوان دانشیار با دانشگاه آزاد اسلامی، واحد مشهد همکاری دارد. تاکنون بیش از یک صد مقاله در کنفرانس‌های داخلی و خارجی و نشریات به چاپ رسانده است. گرایش‌های پژوهشی ایشان پردازش زبان و گفتار و آموزش شبکه‌های عصبی و کاربرد آن در مدل‌سازی سامانه‌های زیستی می‌باشد.

نشانی رایانامه ایشان عبارت است از:

[rahati@mshdia.ac.ir](mailto:rahati@mshdia.ac.ir)



#### اعظم استاجی عضو هیأت علمی

گروه زبان‌شناسی دانشگاه فردوسی مشهد است. وی تحصیلات کارشناسی خود را در رشته زبان و ادبیات انگلیسی، دانشگاه شهید بهشتی تهران

به اتمام رساند. تحصیلات کارشناسی ارشد وی در رشته زبان‌شناسی همگانی، دانشگاه فردوسی مشهد در سال ۱۳۷۶ به پایان رسید. در سال ۱۳۷۷ تحصیلات دوره دکتری خود را در رشته زبان‌شناسی همگانی، دانشگاه فردوسی مشهد آغاز کرد و در سال ۱۳۸۳ به‌عنوان نخستین فارغ‌التحصیل دوره دکتری رشته زبان‌شناسی همگانی دانشگاه فردوسی مشهد، تحصیلات خود را به اتمام رساند. وی از سال ۱۳۸۳ تاکنون در دانشگاه فردوسی مشهد مشغول کار است. زمینه‌های مورد علاقه وی زبان‌شناسی تاریخی، واج‌شناسی، تحلیل گفتمان و زبان‌شناسی رایانه‌ای است.

نشانی رایانامه ایشان عبارت است از:

[estaji@um.ac.ir](mailto:estaji@um.ac.ir)