پژوهش‌های زبانی، سال ۹، شمارهٔ ۱، بهار و تابستان ۱۳۹۷
(از ص ۹۹ تا ص ۱۱۷)

سامانهٔ رفع ابهام معنایی از حروف اضافه در زبان فارسی با استفاده از قالب‌های معنایی

زهرا مظفری
دانشجوی دکترای زبان‌شناسی همگانی، دانشگاه سیستان و بلوچستان
گیتی تاکی[۱]
استادیار زبان و ادبیات انگلیسی، دانشگاه سیستان و بلوچستان
مجتبی صباغ جعفری
استادیار گروه مهندسی کامپیوتر، دانشگاه ولی‌عصر رفسنجان
پاکزاد یوسفیان
استادیار زبان و ادبیات انگلیسی، دانشگاه سیستان و بلوچستان
تاریخ دریافت مقاله: ۹۶/۸/۱۳؛ تاریخ پذیرش مقاله ۹۶/۱۲/۲۷

چکیده

رفع ابهام معنایی از کلمات در بافت یکی از مهم‌ترین چالش‌ها در حوزهٔ پردازش زبان طبیعی و زبان‌شناسی رایانشی است. در این میان حروف اضافه، به‌خصوص در زبان فارسی، در پژوهش‌های مربوط به رفع ابهام معنایی همواره نادیده انگاشته شده‌اند. ازاین‌رو، پژوهش حاضر قصد دارد با ارائهٔ الگوریتمی جدید مبتنی بر قالب‌های معنایی، سامانه‌ای قاعده‌مند جهت رفع ابهام معنایی از حروف اضافه «از»، «در»، «با» و «تا» در زبان فارسی ارائه دهد. روش به کار گرفته‌شده در این پژوهش و الگوریتم پیشنهادی منحصربه‌فرد است. دادگان مورداستفاده در این پژوهش شامل ۱۰۰۰ جمله دادهٔ آموزشی، ۱۰۰ جمله دادهٔ توسعه و ۵۰۰ جمله دادهٔ تست است که از منابع اینترنتی و شبکه‌های اجتماعی همچون یوتیوب جمع‌آوری گردیده است. جهت انجام کار، تمام حروف اضافه موردنظر پژوهش برچسب‌دهی معنایی شده و برای آن‌ها در پیکرهٔ آموزشی، قالب‌های معنایی بر اساس زبان قالب بنیاد مینسکی (۱۹۷۵) تعریف شدند. همچنین برای کلمات قبل و بعد حروف اضافه در پیکره نیز قالب‌هایی تهیه و در فایل داده مدخل‌ها وارد سامانه گردیدند. الگوریتم پژوهش در سه مسیر با استفاده از اطلاعات موجود در قالب‌ها، اقدام به تعیین معنای حروف اضافه در جملات می‌کند. نتایج آزمایش‌های دادهٔ تست، نشان دهندهٔ دقت بالای عملکرد سامانه (۹۹/۱۶٪) در رفع ابهام معنایی از حروف اضافه در زبان فارسی است.

واژه‌های کلیدی: زبان‌شناسی رایانشی، پردازش زبان طبیعی، سامانهٔ رفع ابهام معنایی، حروف اضافه، قالب‌های معنایی

---

[۱]. رایانامهٔ نویسندهٔ مسئول: giti_taki2020@yahoo.com

۱- مقدمه

یکی از مسائل اساسی در حوزهٔ پردازش زبان که هم‌زمان با آغاز طراحی مترجم‌های ماشینی در دهه ۱۹۵۰ مورد توجه محققین بوده است، رفع ابهام معنایی از کلمات است که به‌عنوان شاخه‌ای جدا در زبان‌شناسی رایانشی، پژوهش‌های زیادی را به خود اختصاص داده است. رفع ابهام از معنی کلمه یا تشخیص اینکه کدام یک از معانی ممکن کلمه در یک بافت خاص به‌کاررفته، یکی از پایه‌ای‌ترین وظایف در پردازش زبان طبیعی است. در حوزهٔ زبان‌شناسی رایانشی و پردازش زبان طبیعی به این مسئله، رفع ابهام معنایی[۱] می‌گویند که این‌گونه تعریف می‌شود: «فرایند خودکار تعیین معنای دقیق کلمه در بافت توسط رایانه» (آگیر و ادموند[۲]،۲۰۰۷:۳).

تاکنون پژوهش‌های بسیاری در زمینهٔ سامانه‌های رفع ابهام معنایی به‌خصوص در زبان انگلیسی انجام‌شده است. با جستجو در بانک مقالات رشتهٔ زبان‌شناسی رایانشی[۳]، حدود ۷۰۰ مقاله در این زمینه می‌توان یافت که در آن‌ها به رفع ابهام معنایی از کلمات مختلف به‌ویژه اسم پرداخته‌شده است. همچنین در زمینهٔسامانه‌های رفع ابهام معنایی از حروف اضافه، پژوهش‌های درخور توجهی در زبان انگلیسی صورت گرفته است که از مهم‌ترین آن‌ها می‌توان به پژوهش لیتکوسکی و هارگراوس[۴] (۲۰۰۷) اشاره داشت. رفع ابهام معنایی از حروف اضافه در پژوهش آنان بر اساس مدل SemEval-2007 task بسیار موفقیت‌آمیز بوده است.

معدود پژوهش‌های انجام‌شده در زبان فارسی در زمینهٔ سامانه‌های رفع ابهام معنایی، بیشتر بر روی اسامی متمرکز بوده است و به حروف اضافه پرداخته نشده است. غافل از اینکه نادیده انگاشتن و بی‌اهمیت جلوه دادن حروف اضافه در حوزهٔ پردازش زبان طبیعی به معنای کاهش دقت و عملکرد سامانه‌ها خواهد بود. سامانه‌های رفع ابهام معنایی از حروف اضافه در ترجمهٔ ماشینی، در سامانه‌های بازیابی اطلاعات[۵] و موتورهای جستجو[۶] اهمیت زیادی دارند. به‌عنوان‌مثال، در سامانه‌های بازیابی اطلاعات، اگر تنها

---

[1]. word sense disambiguation

[2].Agirre & Edmonds

[۳]. مجموعه‌ای از مقالات زبان‌شناسی رایانشی است که به همت انجمن زبان‌شناسی رایانشی در وبگاه http://www.aclweb.org/ گردآمده است.

[4].Litkowski & Hargraves

[5]. information retrieval

[6]. search engines

عناصر واژگانی مورد تأکید باشند، در بسیاری مواقع خواستهٔ کاربر برطرف نخواهد شد. جستار[1] «مهاجرت به ایران» و «مهاجرت از ایران» را در نظر بگیرید که تفاوت این دو تنها در حروف اضافه «به» و «از» هست؛ ولی همین حروف اضافه باعث شده‌اند، این دو عبارت دو مفهوم کاملاً متفاوت داشته باشند؛ اولی به معنای ورود به ایران و دومی به معنای خروج از ایران است. حال اگر در مرحلهٔ پیش‌پردازش سامانه، حروف اضافه حذف شوند، آنچه باقی می‌ماند دو کلمهٔ «مهاجرت» و «ایران» است. لذا سامانه اسناد یکسانی را برای هر دو جستار بازیابی خواهد کرد.

ازآنجایی‌که در زبان فارسی تمرکز اکثر سامانه‌های رفع ابهام معنایی بر روی کلمات واژگانی همچون اسم بوده است؛ لذا پژوهش حاضر قصد دارد با طراحی قالب‌های معنایی[2] که عمدتاً برای مفاهیم واژگانی مطرح‌شده‌اند و طرح اولیهٔ آن‌ها به آراء مینسکی[3] (۱۹۷۵) در زمینهٔ هوش مصنوعی بازمی‌گردد، سامانه‌ای جهت رفع ابهام معنایی از حروف اضافه سادهٔ پیشین پرکاربرد «در»، «از»، «با» و «تا» در زبان فارسی ارائه دهد که از این نظر الگویی جدید در میان روش‌های مختلف رفع ابهام معنایی در پردازش زبان طبیعی به شمار می‌رود. تنها پژوهشی که با استفاده از قالب‌های معنایی سامانه‌ای برای ابهام‌زدایی از حروف اضافه در متن ارائه می‌دهد، مربوط به آرنا و همکاران (۲۰۱۴) هست که بر روی تعدادی از حروف اضافه در زبان اسپانیایی انجام‌شده است. نتایج پژوهش (۹۸٪ پاسخ صحیح)، نشان‌دهندهٔ میزان دقت بالای الگوریتم ارائه‌شده در سامانهٔ ابهام‌زدایی از حروف اضافه است.

پیکرهٔ پژوهش حاضر برگرفته از منابع اینترنتی و شبکه‌های اجتماعی همچون یوتیوب می‌باشد. معانی مختلف حروف اضافه موردنظر نیز از فرهنگ‌های لغت و کتب مطرح در این زمینه همچون کتاب حرف‌اضافه و ربط (خطیب رهبر: ۱۳۶۷)، فرهنگ فشرده سخن (انوری:۱۳۸۲) و دستور زبان فارسی (انوری و گیوی: ۱۳۸۹) استخراج‌شده‌اند. این پژوهش شامل چهار بخش است: در بخش مقدمه، به بیان مسئله و ذکر آثار مرتبط پرداخته‌شده است. در بخش دوم بنیان‌های نظری پژوهش مرور می‌شوند و سپس در بخش سوم الگوریتم رفع ابهام، سامانهٔ رفع ابهام معنایی و نحوهٔ ابهام‌زدایی مطرح و در آخرین بخش به بحث و نتیجه‌گیری پرداخته می‌شود.

---

1. query
2. semantic frames
3. Minsky

۲- مبانی نظری پژوهش

یکی از روش‌هایی که از دههٔ ۱۹۶۰ در سامانه‌های رفع ابهام معنایی کاربرد داشته، استفاده از زبان قالب- بنیاد[1] برای رفع ابهام معنایی از کلمات است که روشی مطرح در حوزهٔ هوش مصنوعی است. روشی که پژوهش حاضر قصد دارد بر مبنای آن به رفع ابهام معنایی از حروف اضافه در زبان فارسی بپردازد، استفاده از ساختار قالب در زبان قالب- بنیاد مینسکی (۱۹۷۵) است.

ایدهٔ زبان- قالب بنیاد به‌عنوان الگویی برای شبیه‌سازی فرایند درک انسانی در سامانه‌های رایانه‌ای متعلق به مینسکی (۱۹۷۵) می‌باشد. مبحث قالب‌ها در آثار گافمن (۲۰۱۳) و بوچلر (۲۰۰۰) نیز مطرح شده است. گافمن (۲۰۱۳) قالب‌ها را ساختارهای شناختی می‌داند که به‌واسطهٔ آن‌ها، انسان‌ها اطلاعات پیچیده را درک می‌کنند. بوچلر (۲۰۰۰) معتقد است ما قالب‌ها را ایجاد می‌کنیم تا موقعیتی که در آن هستیم را نام‌گذاری کنیم؛ جنبه‌هایی از موقعیت را تفسیر کنیم و به‌واسطهٔ این تفسیر با دیگران ارتباط برقرار نماییم. درعین‌حال، در آثار فیلمور[2] (۱۹۷۶، ۱۹۷۷، ۱۹۸۲، ۱۹۸۵، ۱۹۹۷) نیز که یکی از مؤسسان زبان‌شناسی شناختی است، مفهوم قالب مطرح‌شده است. اساس رویکرد فیلمور بر این است که معنی واژه‌ها را باید در ارتباط با بازنمایی‌های قالب‌های معنایی، طرح‌واره‌های ساخت مفهومی و الگوهای عقاید و باورها توصیف کرد. به بیان فیلمور (۸۸:۱۹۸۲) «تمام تجربیاتی که هر گوینده از سناریوها و نهادهای اجتماعی در ذهن دارد، معنی واژه را مشخص می‌کند».

مفهوم قالب و اجزای تشکیل‌دهنده آن، در مثال معروف مینسکی (۱۹۷۵) این‌طور شرح داده‌شده است: یک خوانندهٔ آمریکایی با شنیدن جمله (۱) به‌درستی درمی‌یابد که منظور از کلمهٔ «مهمانی» در این جمله، همان مهمانی جشن تولد است، اما هیچ واژه‌ای در جمله (۱)، مستقیماً مفهوم جشن تولد را برنمی‌انگیزد. احتمالاً قالب واژهٔ «مهمانی جشن تولد» در این فرهنگ شامل اطلاعاتی در خصوص شرکت‌کنندگان، مکان و نوع سرگرمی‌های مرسوم جشن تولد است؛ مانند بادبادک، شمع، کیک، آواز و .... واژهٔ «بادبادک» (kite) در جمله فوق، سرنخی است برای درک مفهوم واژهٔ «مهمانی»؛ زیرا کلمهٔ «مهمانی» در قالب معنایی خود در ذهن فرد اطلاعاتی دارد که با قالب واژهٔ

[1]. frame-based
[2]. Fillmore

«بادبادک» مرتبط است. لذا اطلاعات معنایی که در قالب واژهٔ «مهمانی بچه‌ها» مطرح می‌شود، بیشتر از تعریف آن واژه در فرهنگ لغت است. شکل (۱) به بخشی از این قالب اشاره دارد.

(1).Mary was invited to Jack's party. She wondered if he would like a kite.

واژه موردنظر: مقوله نحوی مهمانی بچه‌ها: {اسم }
(تعریف) {تجمع افراد در مکانی برای اوقات خوش یا جشن گرفتن یک رویداد که در آن مرسوم است چیزی بخورند و یا بنوشند ...}
(عامل معلوم) {شخصی که جشن می‌گیرد}
(عامل مجهول) {افرادی که به مهمانی می‌روند، بستگان، دوستان}
(واژگان شامل) {رویداد، موفقیت، تجمع}
(زیرشمول) {مهمانی جشن تولد، مهمانی غسل‌تعمید}

شکل ۱- بخشی از قالب واژه «مهمانی بچه‌ها» (مینسکی،۱۹۷۵)

اطلاعاتی که در خصوص معنای واژهٔ «مهمانی بچه‌ها» در قالب فوق گنجانده‌شده است، ابتدا شامل مقولهٔ نحوی کلمه است؛ زیرا باید مشخص باشد این کلمه به کدام مقوله نحوی اسم، فعل، صفت و یا قید متعلق است. سپس، تعریف این واژه در قالب گنجانده می‌شود که این تعریف عمدتاً همان نوع تعریفی است که در فرهنگ لغات جامع آورده می‌شود. عامل معلوم اشاره به شخصی دارد که این نوع مهمانی را برگزار می‌کند و عامل مجهول اشاره به نوع افرادی دارد که معمولاً به این نوع مهمانی‌ها می‌روند. واژگان شامل[1] و زیر شمول[2] اشاره به مفهوم شمول معنایی دارد که یکی از انواع رابطه‌های واژگانی بین کلمات است. صفوی (۹۹:۱۳۹۲) بیان می‌دارد که هرگاه مفهومی بتواند یک یا چند مفهوم دیگر را شامل شود، در این صورت بین آن مفاهیم رابطه شمول معنایی برقرار است. به‌عنوان نمونه، مفهوم واژهٔ «گل» مفهوم واژه‌های «لاله»، «سنبل»، «میخک» و مانند آن را در برمی‌گیرد. در این صورت مفهوم گل، شامل و مفاهیم دیگر زیر شمول به‌حساب می‌آیند که بین آن‌ها رابطه هم شمولی[3] برقرار است. مفهوم شامل یکی از اجزای مهم در قالب‌ها است.

[1].hypernm
[2].hyponym
[3].hyponymy

هر یک از معانی مربوط به یک کلمه، قالب مخصوص خود را دارد. به‌عنوان نمونه، کلمه‌ای که ۵ معنا دارد، لذا ۵ قالب دارد. لازم به ذکر است که می‌توان هر نوع اطلاعات معنایی مربوط به مفهوم واژهٔ موردنظر را به قالب‌ها اضافه کرد و آن‌ها را به لحاظ معنایی غنی‌تر نمود (آرنا و همکاران،۲۰۱۴). مسئله مهم این است که اطلاعات قالب‌ها ثابت نیستند و با توجه به نوع سامانه‌های هوشمندی که طراحی می‌شوند، می‌توانند متفاوت باشند و غنی‌تر شوند.

این پژوهش قصد دارد با تعریف این نوع قالب‌ها برای حروف اضافهٔ سادهٔ پیشین پرکاربرد «از»، «در»، «تا» و «با» در فارسی، سامانه‌ای برای رفع ابهام معنایی آن‌ها یا بهتر بگوییم برای تعیین خودکار معنای آن‌ها در بافت ارائه دهد.

۳- سامانهٔ رفع ابهام معنایی

جهت طراحی سامانهٔ قاعده‌مند رفع ابهام معنایی از حروف اضافه، ابتدا حروف اضافه موردنظر در پیکرهٔ متنی پژوهش (۱۰۰۰ جمله) که داده اولیه و آموزشی[1] در سامانه هستند، به لحاظ معنایی برچسب‌گذاری شدند.

معناهای متعدد حرف‌اضافه «از» در پیکرهٔ متنی پژوهش: ۱- به‌واسطه ۲- منشأ ۳- جنس، نوع ۴- درباره، مربوط‌به ۵- علت، سبب ۶- شامل ۷- اندازه، مقدار ۸- ابتدای زمان، دوره زمانی ۹- نسبت، مقایسه ۱۰- روش، از طریق ۱۱- داخل محدوده مجازی یا مکانی ۱۲- تمایز ۱۳- بخش، تکه ۱۴- جزء ۱۵- مکان ۱۶- تعلق، متعلق به ۱۷- به وجود آمده، به‌وسیله ۱۸- از جهت، ازنظر ۱۹- وضعیت، چگونگی

معناهای متعدد حرف‌اضافه «در» در پیکرهٔ متنی پژوهش: ۱- محدوده زمانی ۲- محدوده مکانی ۳- راجع به چیزی، مربوط‌به ۴- محدوده مکانی مجازی ۵- داخل، درون ۶- درون حوزه مجازی ۷- سبب، علیت ۸- در حال، وضعیت و چگونگی ۹- نوع ۱۰- به هنگام ۱۱- بیان تناوب ۱۲- اندازه ۱۳- علیه چیزی ۱۴- در مقایسه با

معناهای متعدد حرف‌اضافه «با» در پیکرهٔ متنی پژوهش: ۱- بیان ارتباط بین دو یا چند شخص یا مفهوم ۲- از طریق ۳- به‌وسیله، به‌واسطه ۴- دارندگی، دارای ۵- همراه، همراهی ۶- وضعیت و حالت ۷- علیه، در مقابل ۸- جنس، نوع ۹- زمان ۱۰- در مورد، در زمینه ۱۱- مکان ۱۲- تناسب ۱۳- تمایز

معناهای متعدد حرف‌اضافه «تا» در پیکرهٔ متنی پژوهش: ۱- انتهای زمان، فاصله زمانی ۲- فاصله مکانی ۳- مقدار ۴- بیان تفاوت، انواع ۵- فاصله مکانی مجازی

شکل (۲) به نمونه‌هایی از پیکرهٔ آموزشی برچسب خورده در این پژوهش اشاره دارد. سپس جهت آزمایش سامانه و رفع نواقص قالب‌ها، حدود ۱۰۰ جمله تحت عنوان

[1]. training data

دادهٔ توسعه[1] گردآوری شدند که به‌واسطهٔ آن‌ها سامانه مورد آزمایش قرار گرفت و نواقص قالب‌ها برطرف شد. درنهایت جهت آزمایش نهایی، حدود ۵۰۰ جمله تحت عنوان دادهٔ تست[2] (شکل۳) از منابع مختلف اینترنتی و یوتیوب گردآوری شدند که به‌واسطهٔ آن‌ها سامانه مورد آزمایش نهایی قرارگرفته؛ نتایج نهایی حاصل شد.

رانندگی در شب بسیار خطرناک است.	محدوده زمانی
با یک قاشق وزر دادن را ادامه دهید.	بوسیله، بواسطه
اشراف تا دندان مسلح همه را کشتند.	فاصله مکانی مجازی
کودک می تواند از شکم به کمر چرخ بزند.	وسیله، روش، از طریق
او می تواند وسایل را از روی زمین بردارد.	مکان
با ما همراه باشید.	همراهی، همراه
سعی کنید در حین ورز دادن گوشت را زیادی له نکنید.	درحال، وضعیت و چگونگی
آنها خود را با واکنش‌های متقابل اجتماعی سازگار می‌کردند.	علیه، درمقابل
ما تا چه حد جایگاه ارزشمند معلمی را شناخته‌ایم.	مقدار

شکل ۲- نمونه‌هایی از پیکرهٔ آموزشی برچسب خورده

در روز پنج شنبه تا صبح به تدریج کاهش دما را در نیمه شمالی خواهیم داشت.	محدوده زمانی
انتهای زمانی، فاصله زمانی	محدوده مکانی
با تبدیل انرژی بادی به نیروی الکتریسیته می توان برق تهیه نمود.	از طریق
حافظ شیرازی شاعری با معلومات غیبی بود.	دارندگی، دارای
مردم آمریکا طی تظاهراتی اعتراضات خود را با نظام سرمایه داری اعلام کردند.	علیه، درمقابل
چرا مخالفان سند بیست ساله در دولت سکوت کرده اند.	داخل، درون
دکور کافی شاپ های مدرن همه از جنس چوب است.	جنس، نوع
بیش از میلیون ها مسلمان در هند بی خانمان شده اند.	اندازه و مقدار محدوده مکانی

شکل ۳- نمونه‌هایی از دادهٔ تست

[1]. development data
[2]. test data

در مرحله بعدی، قالب‌های معنایی برای حروف اضافه «از»، «با»، «تا» و «در» تعریف شدند. مواردی که معمولاً در قالب‌ها همواره به‌عنوان موارد ثابت گنجانده می‌شوند، شامل مقولهٔ نحوی کلمه، تعریف یا همان معنای کلمهٔ موردنظر، چند مثال از کلمه در بافت و واژگان شامل. واژگان شامل هر کلمه نیز از وردنت[۱] و همچنین نسخهٔ فارسی آن، فارس‌نت[۲] قابل‌استخراج است. درعین‌حال، چنانچه بخواهیم برای حروف اضافه نیز همین نوع اطلاعات را در قالب‌ها وارد کنیم، مسئلهٔ اساسی واژگان شامل است که در فارس‌نت به دلیل اینکه حروف اضافه عناصر نقشی هستند، به آن‌ها پرداخته نشده و شاملی برای آن‌ها ذکر نشده است. به همین منظور، برای حروف اضافه در قالب‌ها دو مورد شامل بعد[۳] و شامل قبل[۴] گنجانده شد. این دو مفهوم را که اول‌بار آرنا (۲۰۱۴) برای قالب‌های حرف‌اضافه مطرح نمود، این امکان را فراهم کرد تا سامانهٔ رفع ابهام معنایی برای برقراری ارتباط بین قالب‌ها از آن‌ها استفاده کند. در قالب حرف‌اضافه شاملی که برای کلمه بعد از حرف‌اضافه می‌تواند وجود داشته باشد، تحت عنوان شامل بعد در قالب مربوطه آورده می‌شود و شاملی که برای کلمه قبل از حرف‌اضافه وجود دارد، تحت عنوان شامل قبل در قالب حرف‌اضافه درج می‌شود. ازاین‌رو، برای درج شامل قبل و بعد در قالب حروف اضافه، کلمات قبل و بعد حروف اضافه در جملات که محتوای واژگانی داشته باشند بسیار اهمیت دارد؛ زیرا شامل آن کلمات باید از فارس‌نت استخراج و در قالب حرف‌اضافهٔ موردنظر درج شود. ازآنجاکه روی‌هم‌رفته، حدود ۵۱ معنای مختلف برای حروف اضافه موردنظر پژوهش در پیکرهٔ متنی برچسب معنایی داده شدند، لذا حدود ۵۱ قالب معنایی برای حروف اضافه تعریف شدند که سامانه به‌واسطهٔ اطلاعات موجود در آن‌ها اقدام به رفع ابهام از حروف اضافه در جملات می‌کند. در این قسمت، قالب هر حرف‌اضافه تنها در یکی از معانی‌اش به‌عنوان نمونه آورده شده است.

---

[1].WordNet

[۲]. فارس‌نت پایگاه دانشی است حاوی اطلاعات در مورد واژه‌ها، مفاهیم آن‌ها، اطلاعات نحوی و روابط معنایی میان آن‌ها

[3]. pro-hpernym

[4]. pre-hypernym

قالب معنایی «از»:
مقوله دستوری: «حرف‌اضافه»
تعریف: «کلمه‌ای نقشی که نشان می‌دهد طول مدت یک دوره زمانی یا نقطه شروع یک بازه زمانی»
«ابتدای زمان، دوره زمانی»
نمونه: «از امروز»، «از امسال»، «از دیرباز»، «از این لحظه»
شامل کلمه بعد: «زمان»، «دوره زمانی»
شامل کلمه قبل: «زمان»، «دوره زمانی»

قالب معنایی «در»:
مقوله دستوری: «حرف‌اضافه»
تعریف: «کلمه‌ای نقشی نشان‌دهنده مکان»
«محدوده مکانی»
نمونه: «در تهران»، «در اتوبوس»، «در شهر»
شامل کلمه بعد: «نهاد»، «موتور»، «مرکز»، «منطقه»، «جا»، «قسمت»، «مکان»، «زمین»
شامل کلمه قبل: «زمین»

قالب معنایی «تا»:
مقوله دستوری: «حرف‌اضافه»
تعریف: «کلمه‌ای نقشی که نشان می‌دهد طول مدت یک دوره زمانی یا انتهای زمانی را»
«انتهای زمانی، فاصله زمانی »
نمونه: «تا فردا»، «تا سال بعد»، «تا امشب»
شامل کلمه بعد: «زمان»، «دوره زمانی»
شامل کلمه قبل: «زمان»، «دوره زمانی»

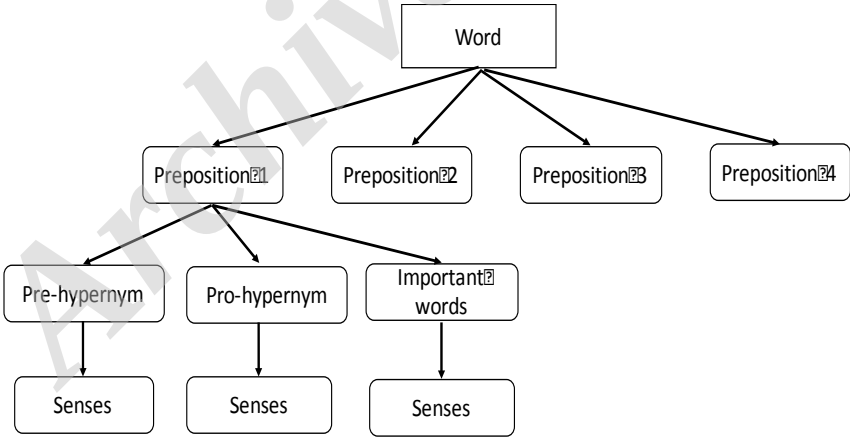
قالب معنایی «با»:
مقوله دستوری: «حرف‌اضافه»
تعریف: «کلمه‌ای نقشی که نشان می‌دهد همراهی و یا همراه شدن با کسی یا چیزی»
«همراهی, همراه»
نمونه: «با پدرم»، «با خودم»، «با چای»، «با آب»
شامل کلمه بعد: «ضمیر»، «همراه»، «توده»، «نوشیدنی»، «حامل»، «سیال»، «بخش»، «روش»، «جمعیت»، «ملزومات»، «تحول»
شامل کلمه قبل: «نوشیدنی»، «همراه»، «جمعیت»

۳-۱- نحوه محاسبه شامل یک کلمه

در ابتدا سعی بر آن شد که شامل کلمات قبل و بعد حرف‌اضافه، مستقیماً به‌طور خودکار از فارس‌نت استخراج شوند. برای این کار یک سرویس XMLRPC برای استفاده از API های فارس‌نت به زبان Jython نوشته شد که کار آن برقراری ارتباط بین الگوریتم رفع

ابهام و فارس‌نت بود. با این روش، شامل‌ها با موفقیت از فارس‌نت استخراج شدند؛ اما تعداد شامل‌ها برای هر کلمه بسیار زیاد بود و با توجه به سطح انتزاع آن‌ها، شامل‌های متعددی به‌عنوان شامل کلمه موردنظر، از فارس‌نت دریافت می‌شد که باعث پیچیدگی تصمیم‌گیری سامانه می‌گردید؛ بنابراین تصمیم گرفته شد یک فایل داده برای محاسبه شامل‌هایی که از فارس‌نت استخراج می‌شوند، ایجاد شود. این فایل داده که مدخل[۱] نام‌گذاری شد، قالب‌هایی هستند که در هر قالب کلمه موردنظر، شامل آن کلمه و حرف‌اضافه‌ای که در کنار آن کلمه در جمله قرار دارد، درج‌شده است. این فایل در قالب json تهیه‌شده که بتوان به‌راحتی محتوای آن را خواند و اطلاعات آن را به الگوریتم اصلی ارسال نمود.

از آنجاکه قالب‌ها ساختار گسترده[۲] دارند، جستجوی یک شامل در میان قالب‌های حرف‌اضافه پیچیده و زمان‌بر است. لذا قالب‌ها به‌صورت یک ساختار درختی (شکل ۴) و به عبارتی اندیس معکوس[۳] درآورده شدند تا جستجو با سرعت بیشتری انجام‌شده و از پیچیدگی آن نیز کاسته شود. روند جستجو بدین‌صورت است که ابتدا درخت کلمه موردنظر پیدا می‌شود؛ سپس با توجه به حرف‌اضافه‌ای که در حال رفع ابهام است، یک شاخه در درخت دنبال می‌گردد و با توجه به مسیر رفع ابهام (در بخش بعدی توضیح داده می‌شود) معنای موردنظر به الگوریتم رفع ابهام ارسال می‌شود.

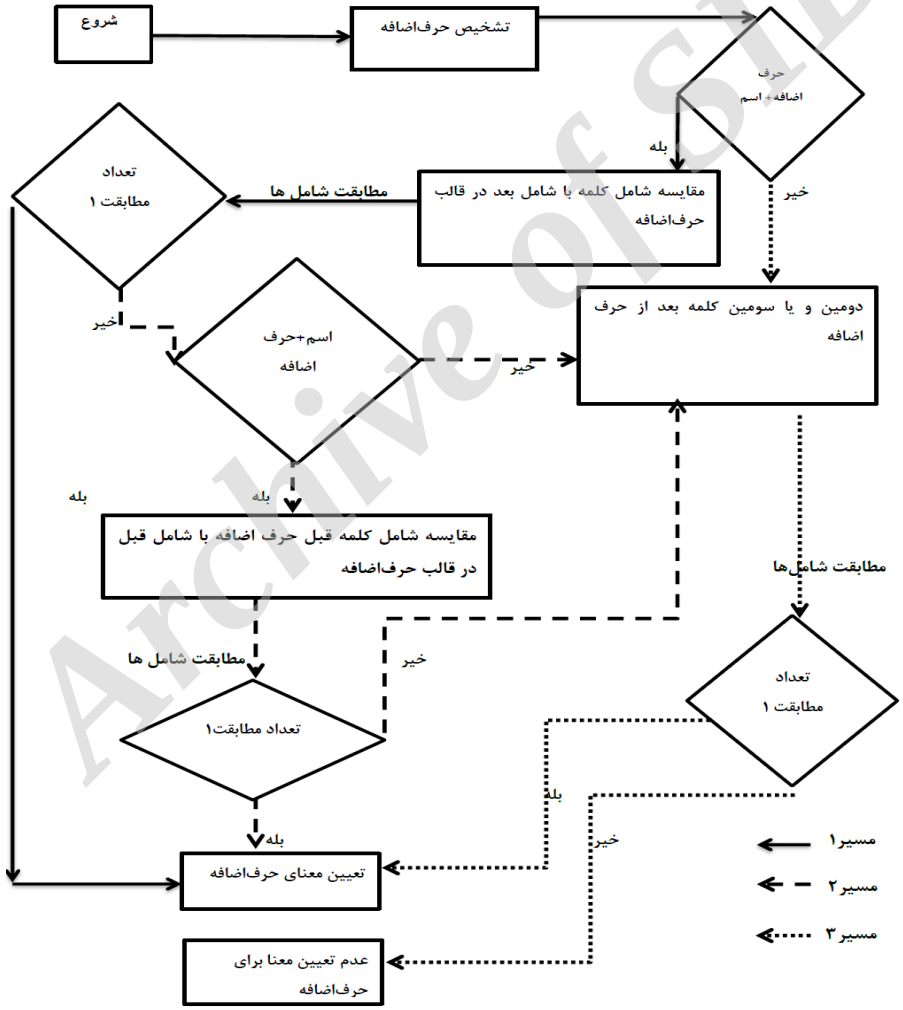


شکل ۴- ساختار درختی اندیس معکوس برای جستجوی کلمات در قالب‌ها

[1]. entry
[2]. flat
[3]. inverted index

۳-۲- الگوریتم رفع ابهام معنایی از حروف اضافه

سامانه به‌واسطهٔ کلمات قبل و یا بعد از حرف‌اضافه که عمدتاً نقش نحوی اسمی دارند، اقدام به تعیین معنا می‌کند. اساس کار سامانه، برقراری ارتباط بین واژگان شامل موجود در قالب حرف‌اضافه و واژگان شامل در مدخل‌هاست. چنانچه بین واژهٔ شامل کلمه قبل یا بعد حرف‌اضافه در مدخل‌ها و واژهٔ شامل موجود در قالب حرف‌اضافه انطباق و یکسانی وجود داشته باشد، معنای موردنظر برای حرف‌اضافه تعیین می‌گردد. شکل (۵) فلوچارت الگوریتمی است که سامانه به‌واسطهٔ آن، معنای حروف اضافه را در جملات تعیین می‌کند.



شکل ۵- الگوریتم مسیرهای رفع ابهام معنایی

در الگوریتم فوق، سه مسیر مشخص‌شده است. در ابتدا، جمله ورودی تقطیع شده، برچسب نحوی (POS[1]) زده می‌شود و ریشه کلمه استخراج می‌شود. سپس عملیات رفع ابهام به این شرح آغاز می‌گردد: مسیر ۱: در این مسیر، سامانه حروف اضافه موردنظر پژوهش را در فایلی که برچسب نحوی آن به‌صورت خودکار انجام‌شده شناسایی می‌کند. سپس سراغ کلمه بعد از حرف‌اضافه می‌رود. اگر واژهٔ شامل این کلمه در لیست شامل‌های بعد که در قالب حرف‌اضافه وجود دارد موجود باشد، تطابق قالب‌ها رخ‌داده، سامانه معنای موجود در آن قالب را به حرف‌اضافه می‌دهد. چنانچه این تطابق برقرار نشود، سامانه وارد مسیر ۲ می‌گردد. در مسیر ۲ کلمه قبل حرف‌اضافه شناسایی می‌شود. چنانچه واژهٔ شامل کلمه قبل در لیست شامل‌های قبل در قالب حرف‌اضافه موجود باشد، معنای حرف‌اضافه موردنظر تعیین می‌شود؛ در غیر این صورت سامانه وارد مسیر ۳ می‌گردد. در این مسیر، سامانه دو کلمه بعد یا سه کلمه بعد حرف‌اضافه را جستجو می‌نماید و واژگان شامل آن کلمات را با شامل‌های بعد در قالب حرف‌اضافه موردنظر مقایسه می‌کند. چنانچه مطابقت برقرار بود، معنای حرف‌اضافه تعیین می‌شود؛ در غیر این صورت معنایی برای حرف‌اضافه مشخص نمی‌گردد و الگوریتم پایان می‌یابد.

۳-۳- آزمایش داده‌های تست و نتایج

در این قسمت، نحوه عملکرد الگوریتم رفع ابهام معنایی از حروف اضافه در چند نمونه جمله از داده آزمون شرح داده می‌شود. در جمله (۲)، ابتدا سامانه دو حرف‌اضافه «در» و «از» را تشخیص می‌دهد. سپس الگوریتم وارد مسیر ۱ شده کلمه بعد از حرف‌اضافه «در» را شناسایی می‌کند؛ زیرا قالب معنایی این کلمه در فایل داده مدخل‌ها موجود است (شکل ۶).

(۲) هر چه در آنجا شامل جواهر و نفایس بود بردند.

```
"entry": ["ساخته", "محلی", "شهر", "اروپا", "آنجا", "کوهپایه"],
"tag": "NP",
"prep": "در",
"sense": "محدوده مکانی",
"hypernym": [ "زمین"]
```

شکل ۶- بخشی از قالب کلمات در مدخل‌ها

[1]. part of speech

طبق مسیر ۱ در الگوریتم، سامانه شامل موجود در این مدخل که در اینجا کلمه «زمین» هست را با شامل‌های موجود در قالب‌های معنایی حرف‌اضافه «در» مقایسه می‌کند. همان‌طور که در شکل (۷) مشخص است، شامل بعد «زمین» در قالب معنایی حرف‌اضافه «در» بامعنای محدوده مکانی موجود است، لذا الگوریتم انطباق بین دو شامل را تشخیص می‌دهد و نهایتاً معنای محدوده مکانی را برای حرف‌اضافه «در» تعیین می‌کند (شکل ۸).

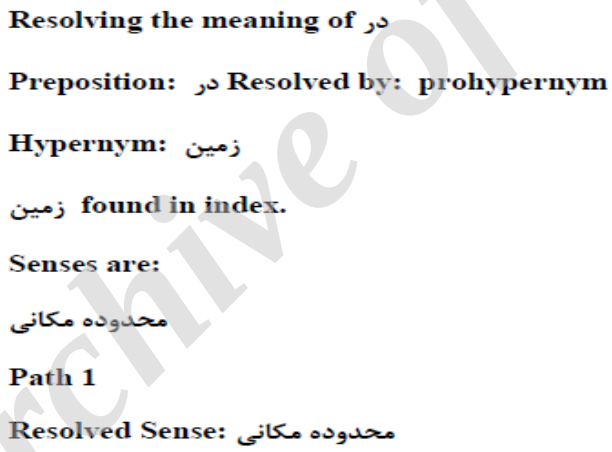
"prep": "در",

"sense": "محدوده مکانی",

"prohypernym": ["نهاد", "موتور", "مرکز", "منطقه", "جا", "قسمت", "مکان", "زمین"],

"prehypernym": ["زمین"],

شکل ۷- قالب معنایی حرف‌اضافهٔ «در» بامعنای محدودهٔ مکانی



شکل ۸- معنای تعیین‌شده توسط الگوریتم برای حرف‌اضافه «در»

در جمله (۳) الگوریتم بعد از تشخیص حرف‌اضافه «از»، ابتدا وارد مسیر ۱ می‌شود و کلمه «ابن» را شناسایی می‌کند. ازآنجاکه این کلمه در فایل داده مدخل‌ها یافت نمی‌شود، الگوریتم وارد مسیر ۲ شده، کلمه قبل حرف‌اضافه را جستجو می‌کند. با جستجو در مدخل‌ها، قالب این کلمه یافت می‌شود (شکل ۹). سپس، شامل موجود در این قالب (کلمه زمان) با شامل‌های قبل در قالب‌های حرف‌اضافه «از» مقایسه می‌گردد، (شکل ۱۰). چنانچه انطباق برقرار باشد، معنای موردنظر تعیین می‌شود (شکل ۱۱).

(۳) پس از ابن مسعود، پسرش عبدالعزیز بن سعود به امارت رسید.

"entry": ["زمان", "پیش", "پس", "دیرباز", "قرن", "سده", "قرون", "دیر", "پایان"],

"tag": "NP",

"prep": "از",

"sense": "ابتدای زمان،دوره زمانی",

"hypernym": ["زمان"]

شکل ۹- بخشی از قالب کلمات در مدخل‌ها

"prep": "از",

"sense": "ابتدای زمان،دوره زمانی",

"prohypernym": ["زمان", "دوره زمانی"],

"prehypernym": ["زمان", "دوره زمانی"],

شکل ۱۰- قالب حرف‌اضافه «از» بامعنای ابتدای زمان، دورهٔ زمانی

Resolving the meaning of از

Preposition: از Resolved by: prohypernym

Preposition: از Resolved by: prehypernym

Hypernym: زمان

زمان found in index.

Senses are:

ابتدای زمان،دوره زمانی

Path 2

Resolved Sense: ابتدای زمان،دوره زمانی

شکل ۱۱- معنای تعیین‌شده توسط الگوریتم برای حرف‌اضافه «از»

در جمله (۴)، ابتدا الگوریتم بعد از ناکامی در مسیرهای ۱ و ۲، به‌دلیل نیافتن کلمات موردنظر در مدخل‌ها، وارد مسیر ۳ می‌شود. طبق این مسیر، الگوریتم باید

دومین کلمه (طرف) بعد از حرف‌اضافه را جستجو کند؛ چنانچه نتیجه نگرفت، کلمه سوم بعد از حرف‌اضافه جستجو می‌گردد. در اینجا کلمه «هوا» در مدخل‌ها (شکل ۱۲) یافت می‌شود. لذا با مقایسه شامل‌ها در قالب حرف‌اضافه (شکل ۱۳) و شامل موجود در مدخل‌ها، معنای موردنظر مطابق شکل (۱۴) برای حرف‌اضافه «از» تعیین می‌گردد.

(۴) بمب‌های عربستان از همه طرف هوا و زمین بر سر مردم بحرین فرومی‌ریزد.

"entry": ["هوا"],

"tag": "NP",

"prep": "از",

"sense": "وسیله،روش،ازطریق",

"hypernym": ["جو"]

شکل ۱۲- بخشی از قالب کلمات در مدخل‌ها

"prep": "از",

"sense": "وسیله،روش،از طریق",

"prohypernym": ["جو", "روش", "قسمت بدن", "آیین", "واسطه"],

"prehypernym": [],

شکل ۱۳- قالب حرف‌اضافه «از» به معنای وسیله، روش، از طریق

Resolving the meaning of از

Preposition: از Resolved by: prohypernym

Third post word: هوا

Preposition: از Resolved by: prohypernym

Hypernym: جو

جو found in index.

Senses are:

وسیله،روش،از طریق

Path 3

Resolved Sense: وسیله،روش،از طریق

شکل ۱۴- معنای تعیین‌شده توسط الگوریتم برای حرف‌اضافه «از»

۴- نتیجه

نتایج آزمایش ۵۰۰ داده آزمون در سامانهٔ رفع ابهام معنایی از حروف اضافه «از»، «در»، «با» و «تا» نشان‌دهنده دقت بالای سامانه در رفع ابهام معنایی است. درصد پاسخ‌های صحیح به‌دست‌آمده حدود Accuracy = 99.16% است. ازآنجاکه تابه‌حال سامانه‌ای برای رفع ابهام معنایی با استفاده از قالب‌های معنایی در زبان فارسی طراحی نشده است، روش به کار گرفته‌شده در این پژوهش و الگوریتم پیشنهادی منحصربه‌فرد است. میزان خطاهای به‌دست‌آمده در این سامانه را می‌توان به چندین عامل نسبت داد. گاهی اوقات وجود حروف‌اضافه در جمله به‌دلیل فعل جمله است؛ به این معنا که فعل برای کامل کردن معنا به حروف‌اضافه نیاز دارد و این حروف گرچه به‌ظاهر حرف‌اضافه‌اند، اما به‌تنهایی معنای مستقلی ندارند، مانند عبارات فعلی «از کوره دررفتن»، «از دست دادن»، «درنظر آمدن» و یا افعال پیشوندی همچون «درافتادن» و یا «درآمیختن». از همین رو، زمانی که جملاتی با این نوع حروف‌اضافه به سامانه داده می‌شوند، به‌دلیل اینکه سامانه آن‌ها را نیز حرف‌اضافه ساده به‌حساب می‌آورد و نمی‌تواند معنایی برایشان تعیین کند، خطا می‌دهد.

مورد بعدی، مربوط به ماهیت حروف‌اضافه در زبان فارسی است. در این پژوهش، تنها به حروف‌اضافه سادهٔ پیشین پرداخته شد؛ اما همین حروف‌اضافه، در بعضی جملات بدون هیچ نشانه‌ای نقش‌های دیگری دارند و دیگر حرف‌اضافه نیستند. به‌عنوان نمونه، حرف‌اضافه «تا» در جمله (۵) حرف ربط است نه حرف‌اضافه. زمانی که این جمله به سامانه داده می‌شود، حرف‌اضافه «با» در این جمله به‌درستی تشخیص داده‌شده و معنای آن مطابق شکل (۱۵) تعیین می‌شود؛ اما حرف «تا» در این جمله حرف ربط است که سامانه به‌عنوان حرف‌اضافه آن را تشخیص می‌دهد؛ ولی نمی‌تواند معنایی برای آن تعیین کند (شکل ۱۶).

(۵) با بیماری سرطان سال‌ها جنگید تا عاقبت مغلوب شد.

Resolving the meaning of با

Preposition: با Resolved by: prohypernym

Hypernym: وضعیت

وضعیت found in index.

Senses are:

علیه,درمقابل

Path 1

Resolved Sense: علیه,درمقابل

شکل ۱۵- معنای تعیین‌شده توسط الگوریتم برای حرف‌اضافه «با»

Resolving the meaning of تا

Preposition: تا Resolved by: prohypernym

Preposition: تا Resolved by: prehypernym

Second post word: شد.

Preposition: تا Resolved by: prohypernym

No unique meaning found.

شکل ۱۶- عدم تعیین معنا

مسئله اساسی این است که برای همه نوع حروف‌اضافه، برچسب نحوی (مقوله نحوی) که به‌طور خودکار در پیکره‌ها زده می‌شود، برچسب p (preposition) است. لذا سامانه تمام این موارد را حرف‌اضافه ساده به‌حساب می‌آورد و به دنبال معنایی برای آن‌ها می‌گردد و چون معنایی یافت نمی‌شود، خطا می‌دهد.

راهکاری که می‌توان برای حل این مشکل پیش نهاد داد این است که برچسب‌های نحوی که به‌طور خودکار در پیکره‌ها خورده می‌شود، جزئی‌تر شده و برای هر یک از این موارد برچسبی متفاوت خورده شود تا از هم قابل‌تشخیص باشند. به‌عنوان‌مثال، برچسب نحوی حرف‌اضافه ساده متفاوت از برچسب نحوی حرف‌اضافه مرکب باشد. گرچه این کار نیازمند کار زبان‌شناسی گسترده، مطالعه دقیق و زمان‌بر است.

الگوریتم سامانهٔ رفع ابهام معنایی نوعی الگوریتم اکتشافی[1] است؛ لذا برای کارایی بالاتر قالب‌های بسیاری باید تعریف شود تا سطح عمومیت سامانه را بالا ببرد که نیازمند کار انسانی گسترده است. قالب‌های معنایی تعریف‌شده خود می‌توانند در تهیه هستان‌شناسی از حروف‌اضافه و یا کلمات دیگر مورداستفاده قرار بگیرند. از طرفی می‌توان برای کاربرد قالب‌ها در سایر برنامه‌های رایانشی، هرگونه اطلاعات موردنیاز دیگری در آن‌ها قرارداد و قالب‌ها را غنی‌تر نمود. در کارهای آتی سعی بر آن است که به‌جای استفاده از فایل داده مدخل‌ها که برای کلمات قبل و بعد حروف‌اضافه تعریف‌شده بودند، مشکل تعدد شامل‌ها و هم‌پوشانی آن‌ها در فارس‌نت حل شود تا سامانه بتواند شامل هر کلمه در قالب موردنظر را مستقیماً از فارس‌نت دریافت کند؛ در این صورت سطح عمومیت سامانه نیز بالاتر خواهد رفت.

پی‌نوشت

از آقای دکتر Arena- Guzman عضو موسسه بین‌المللی مهندسان برق و الکترونیک معروف به IEEE به خاطر راهنمایی‌های سخاوتمندانه ایشان در انجام این پژوهش سپاسگزاریم.

منابع

انوری، حسن (۱۳۸۲). *فرهنگ فشرده سخن*. تهران، انتشارات آگاه.

انوری، حسن و احمدی گیوی، حسن (۱۳۸۹). *دستور زبان فارسی*، تهران، انتشارات فاطمی.

خطیب رهبر، خلیل (۱۳۶۷). *دستور زبان فارسی: کتاب حروف‌اضافه و ربط*. تهران، انتشارات فاطمی.

صفوی، کوروش (۱۳۹۲). *درآمدی بر معناشناسی*، تهران، هرمس.

Agirre, E and Edmonds, Ph. 2007. *Word sense disambiguation algorithms and applications*. Springer.com.

Arenas, G.A., Villanueva, D., Rasgado, A.D. and Juarez, O. 2014.Using frames to disambiguate prepositions, *Science Direct*, Vol 40(2),598-610.

Buecheler, S. 2000. *Social movement in advanced capitalism*, Oxford: Oxford University Press.

Fillmore, C. J. 1976. Frame semantics and the nature of language, *In Annals of the NewYork Academy*: *Conferences on the Origin and Development of Language and Speech*, Vol.280, 20-32.

———— 1977b. Topics in lexical semantics, *In Current Issues in Linguistics*, Bloomington, Indian University Press, 76-138.

———— 1982b. Frame semantics, *In the Linguistics Society of Korea*, eds Linguistics in the Morning Calm, Seol, Hanshim, 37- 111.

[1]. huristic

———— 1985a. Frames and the semantics of understanding. *Quaderni de Semantica*, 6(2). 222-254.

———— 1997. Toward a frame based lexicon: the semantics of RISK and its neighbors in frames, fields and its contrast, *New Essays in Semantic and Lexical Organization*, HillsdayNewjersy, 75-102.

Goffman, E. 2013. *The presentation of self in the online world*, Harvard University Press.

Litkowski, K. and Hargraves, O. 2007. Word sense disambiguation of prepositions, *In proceeding of the 4th International Workshop on Semantic Evaluations*, 24-29.

Minsky, M. 1975. A framework for representing knowledge, *The Psychology of Computer Vision*, MIT press, (19).