

زبان‌شناسی رایانشی: رویکردی توصیفی به کاربردها و تکنیک‌های آن

زلفا ایمانی^۱

تاریخ دریافت مقاله: مهر ۱۳۹۹ تاریخ پذیرش نهایی: بهمن ۱۳۹۹

چکیده

نیاز روزافزون بشر به استفاده از رایانه و به دنبال آن تولد یک میان رشته جدید تحت عنوان زبان‌شناسی رایانشی که در یک دهه اخیر با تلفیق دو رشته زبان‌شناسی و علوم رایانه‌ای جایگاهی مجزا در دانشگاه‌های کشور به خود اختصاص داده است از یک سو و نیاز دانشجویان و علاقه‌مندان به آشنایی با و آگاهی از کاربرد زبان‌شناسی رایانشی و کمبود منابع فارسی در راستای توصیف این رشته از سوی دیگر، نگارنده را بر آن داشت تا ماحصل مطالعه و پژوهش خود پیرامون این رشته را در قالب مقاله پیش رو که به معرفی زبان‌شناسی رایانشی، عملکرد و تکنیک‌های موجود در آن پرداخته، به رشته تحریر در آورد و آن را در اختیار دانشجویان بالآخر پژوهش‌گران زبان‌شناسی همگانی و محققان حوزه‌های مرتبط با آن قرار دهد. یافته‌های پژوهش توصیفی پیش رو مشتمل بر ۲ بخش در حوزه زبان‌شناسی رایانشی و نظریه معناشناسی قالبی می‌باشد که با تکیه بر منابع دست اولی که در بر گیرنده شرح مفصلی از دو رویکرد زبان‌شناسی رایانشی، اصول تکنیک‌ها و چگونگی به کارگیری این میان رشته و همچنین نظریه معناشناسی قالبی است، گردآوری، مطالعه و ترجمه شده و اهم مطالب آن به شیوه‌ای تدوین گشته که مخاطب را با سیمای کلی این میان رشته‌ی جدید که در حال حاضر رشته‌ای مستقل به شمار می‌رود آشنا نماید.

واژگان کلیدی: زبان‌شناسی رایانشی، هوش مصنوعی، ترجمه ماشینی، پردازش زبان طبیعی، معناشناسی قالبی.

مقدمه

زبان یکی از ابزارهای اصلی و محوری در زندگی اجتماعی و حرفه‌ای‌مان به شمار می‌رود. این ابزار وسیله‌ای برای انتقال ایده‌ها، عقاید، اطلاعات، احساسات و عواطف عمل می‌کند. افزون بر این، در متقاعد کردن افراد، پرسش و پاسخ، دستور دادن، علوم رایانه‌ای و به ویژه هوش مصنوعی^۱ نقش بنیادی ایفا می‌نماید. زبان‌شناسی که به مطالعه زبان به عنوان نظامی از ارتباطات بشر می‌پردازد و با رویکردهای متعدد به مطالعه زبان و حوزه‌های مختلف آن از جمله معنی‌شناسی، نحو، واج‌شناسی، واژه‌شناسی و تحلیل گفتمان می‌نگرد (ریچاردز و شمیت، ۱۹۸۵) شاخه‌های گوناگونی دارد که یکی از این شاخه‌ها زبان‌شناسی رایانشی^۲ است (کریستال، ۲۰۰۳؛ فینچ، ۲۰۰۰).

زبان‌شناسی رایانشی همراه با پردازش زبان طبیعی^۳ در سال‌های اخیر بیشتر کانون توجه قرار گرفته است. زبان‌شناسی رایانشی حوزه مهمی در توسعه صنعت نیز محسوب می‌شود. پژوهش در زبان‌شناسی رایانشی و پردازش زبان طبیعی در سه دهه گذشته مسیر خود را از مدل‌های نظری در راستای یادگیری و آموزش به سوی سیستم‌های پردازش موجود در پیکره‌های بزرگ تغییر داده است (کلارک و همکاران، ۲۰۱۰: ۱).

زبان‌شناسی رایانشی، هوش مصنوعی و پردازش زبان طبیعی موضوعات محوری و کلیدی این پژوهش است که توصیف و کارکرد هر یک به تفصیل مطرح شده است. افزون بر این، با اشاره به ترجمه ماشینی که یکی از دستاوردهای مهم این حوزه به شمار می‌رود و امروزه استفاده از آن بسامد بالایی را در میان دانشجویان و استادان به ویژه افرادی که رشته تخصصی‌شان زبان انگلیسی نیست و همچنین با اشاره به معناشناسی قالبی فیلمور (۱۹۷۷، ۱۹۸۲، ۱۹۸۵) و تأثیری که پروژه وی بر زبان‌شناسی رایانشی نهاده، مقاله حاضر به پایان می‌رسد.

ظهور زبان‌شناسی رایانشی: (مارتین کی، ۲۰۰۳) معتقد است که زبان‌شناسی رایانشی برای اولین بار در سال ۱۹۴۹ برای ترجمه ماشینی^۴ قدم در عرصه ظهور نهاد. وی در ادامه می‌افزاید که اولین کنفرانس ترجمه ماشینی در سال ۱۹۵۲ برگزار و اولین مجله ترجمه ماشینی در سال ۱۹۵۴ منتشر شد. اما عبارت زبان‌شناسی رایانشی در سال ۱۹۶۵ ایجاد شد و در سال ۱۹۷۴ عناوین مجله ترجمه ماشینی و زبان‌شناسی رایانشی به مجله زبان‌شناسی رایانشی تغییر یافت و از سال ۱۹۸۰ تا به امروز، عبارت زبان‌شناسی رایانشی

همچنان مورد استفاده قرار گرفته است.

حوزه زبان‌شناسی رایانشی را می‌توان کاربرد علوم رایانه در مدل‌سازی ارتباطات زبان طبیعی به عنوان یک سیستم نرم‌افزاری تعریف کرد. این حوزه دربرگیرنده تحلیل زبان‌شناختی زبان طبیعی با استفاده از رایانه است. با توجه به اینکه زبان‌های طبیعی مبتنی بر قراردادهای اجتماعی می‌باشند، زبان‌شناسی رایانشی در فاصله میان علوم طبیعی و ریاضیات از یک سو و علوم انسانی و علوم اجتماعی از سوی دیگر قرار دارد. در اوایل دهه چهل میلادی، در حوزه علوم رایانه، تمایزی میان محاسبات عددی و غیر عددی ایجاد شد. منظور از محاسبات عددی، محاسبه اعداد است که در زمینه‌های گوناگونی از جمله فیزیک، شیمی و ستاره‌شناسی، گسترش روز افزون علم را به دنبال داشته است. علاوه بر این، بانک‌داری، خلبانی، سهام‌داری، تولید و امثال آن نیز به محاسبات عددی وابسته است. در واقع، بدون در اختیار داشتن رایانه و نرم‌افزارهای مرتبط با آن، فعالیت در این زمینه‌ها امکان‌پذیر نیست (هاوسر، ۱۹۹۹: ۳).

در مقابل، منظور از محاسبات غیر عددی، توجه به پدیده‌هایی از قبیل ادراک و شناخت است. محاسبات غیر عددی در سال‌های اخیر در زمینه‌های هوش مصنوعی و علوم شناختی کاربرد فراوان داشته است. هوش مصنوعی و علوم شناختی حوزه‌های جدید و میان‌رشته‌ای به شمار می‌روند که به مدل‌سازی داده‌های طبیعی و پردازش الکترونیکی آن می‌پردازند و ارتباط زبان‌شناسی رایانشی با این دو حوزه در این است که زبان‌شناسی رایانشی به عنوان یک ربات سخنگو محاسبات غیر عددی را برای تحلیل زبان طبیعی به کار می‌گیرد (همان).

پردازش زبان طبیعی: پردازش زبان طبیعی از تعامل علوم گوناگونی از جمله علوم رایانه، هوش مصنوعی و روان‌شناسی شناختی پدید آمده است. منظور از پردازش زبان طبیعی، تولید نرم‌افزاری مجهز به مشخصه‌های مرتبط با پردازش زبان است. به موازات آن، پردازش گفتار^۵ صورت می‌گیرد که دربرگیرنده تکنیک‌هایی است که به شناسایی و تولید واحدهای زبانی همچون واج، هجا و کلمه می‌پردازد. در واقع، تفاوت عمده‌ای میان پردازش گفتار و پردازش زبان طبیعی وجود ندارد. تکنیک‌های بسیاری که در پردازش گفتار به کار گرفته می‌شوند مسیر خود را به سوی کاربرد در پردازش زبان طبیعی ادامه داده‌اند. ضمن اینکه زبان‌شناسی

^۴. Machine translation

^۵. Speech processing

^۱. Artificial intelligence

^۲. Computational linguistics

^۳. Natural language processing

«علم و مهندسی ساخت ماشین‌های هوشمند».
هوش مصنوعی شاخه‌ای از علوم رایانه‌ای است که با مطالعه و طراحی عوامل هوشمندی که قادر به درک محیط اطراف خود هستند و اقداماتی را در جهت دستیابی به موفقیت انجام می‌دهند سر و کار دارد (سینگ و همکاران، ۲۰۱۳).

در واقع، منظور از هوش مصنوعی، طراحی و تولید عوامل هوشمند است و منظور از عامل هوشمند، یک سیستم مصنوعی یا طبیعی با قابلیت‌های ادراکی است که به آن این امکان را می‌دهد تا در محیطی خاص عمل کند، به ارضای خواسته‌های خود پردازد و یا با موفقیت به اهداف برنامه‌ریزی شده‌اش دست یابد (زکریا کردی، ۲۰۱۶).

فعالیت در عرصه هوش مصنوعی معمولاً به چندین شاخه تقسیم می‌شود که از آن جمله است برنامه‌ریزی، ادراک و یادگیری. هر سه شاخه با پردازش زبان طبیعی رابطه مستقیم دارند. منظور از برنامه‌ریزی در این عرصه، یافتن گام‌هایی جهت نیل به یک هدف خاص و مشخص است (همان).

در پردازش زبان طبیعی، منظور از برنامه‌ریزی، ضرورت انجام فعالیت‌های پیچیده‌ای است که دربرگیرنده منابع مختلف دانش است که بایستی به منظور دستیابی به هدف نهایی با یکدیگر مشارکت داشته باشند (همان).

به طور کلی، هوش مصنوعی یعنی توسعه سیستم‌های رایانه‌ای در راستای انجام اموری که نیازمند هوش انسانی باشد. نمونه‌هایی از این ادراک بصری، تشخیص گفتار، تصمیم‌گیری و ترجمه از زبانی به زبان دیگر است. ماشینی که از هوش مصنوعی قوی برخوردار باشد قادر به اندیشیدن است و درست همانند انسان عمل می‌کند. چنین ماشینی توانایی یادگیری بر اساس تجربه را دارد. در دهه پنجاه میلادی، آلن تورینگ^۸ آزمون تورینگ^۹ را طراحی نمود که برای تعیین سطح هوشمندی رایانه به کار می‌رفت.

روش پژوهش

از آنجا که پژوهش پیش رو یک مطالعه توصیفی صرف است، یافته‌های آن که مشتمل بر ۲ بخش در حوزه زبان‌شناسی رایانشی و نظریه معناشناسی قالبی می‌باشد با تکیه بر منابع دست اولی که در برگیرنده شرح تفصیلی دو رویکرد زبان‌شناسی رایانشی، اصول تکنیک‌ها و چگونگی به کارگیری این میان رشته و همچنین

پیکره‌های^۱ که به روش‌های گردآوری، حاشیه‌نویسی^۲ و استفاده از پیکره‌ها هم در پژوهش‌های زبان‌شناختی و هم در پردازش زبان طبیعی مربوط می‌شود، یکی از شاخه‌های پردازش زبان طبیعی به شمار می‌رود چرا که نقش بنیادی در ساخت و تولید سیستم‌های پردازش زبان طبیعی ایفا می‌کند (زکریا کردی، ۲۰۱۶).

رویکردهای معطوف به تجزیه و تحلیل زبان طبیعی: سه رویکرد به تجزیه و تحلیل زبان طبیعی وجود دارد (هاوسر، ۱۹۹۹) که عبارت است از:

۱- دستور سنتی

۲- زبان‌شناسی نظری

۳- زبان‌شناسی رایانشی

در ادامه، به توصیف و تبیین هر یک از رویکردهای فوق می‌پردازیم.
دستور سنتی: این رویکرد از روش توصیف و طبقه‌بندی غیر صوری بر مبنای سنت و تجربه بهره می‌جوید. هدف آن گردآوری و طبقه‌بندی قاعده‌مندی‌ها^۳ و بی‌قاعدگی‌های^۴ موجود در زبان طبیعی است و غالباً در آموزش زبان از جمله زبان لاتینی مورد استفاده قرار می‌گیرد.

زبان‌شناسی نظری: این رویکرد از روش منطقی ریاضی برای توصیف زبان‌های طبیعی به واسطه سیستم‌های قاعده صوری^۵ بهره می‌جوید تا صرفاً به عبارات خوش‌ساخت زبان دست یابد. هدف آن توصیف توانایی زبانی ذاتی انسان (توانش) است که به مدد آن جنبه‌هایی از کاربرد زبان در تعامل (کنش) استخراج می‌شود و به علت ناکارآمدی رایانه‌ای‌اش کاربرد نسبتاً محدودی دارد.

زبان‌شناسی رایانشی: این رویکرد روش‌های دستور سنتی و زبان‌شناسی نظری را با روشی کارآمد ترکیب می‌کند تا برنامه‌های رایانه‌ای برای تولید کارآمد دستور صوری^۶ را تولید نماید و این دستور را به طور خودکار بیازماید. در این رویکرد، این مدل‌سازی ساز و کار تعامل زبان طبیعی است که مستلزم تحلیل‌های کاربردشناختی، واژگانی، ساختوازی، معنایی و نحوی یک زبان طبیعی مفروض می‌باشد و کاربرد آن در تمام ابعاد مرتبط با تعامل رایانه و انسان است.

هوش مصنوعی: عبارت هوش مصنوعی را اولین بار جان مک کارتی^۷ در سال ۱۹۵۶ به کار برد. تعریفی که وی از این عبارت ارائه می‌کند چنین است:

^۶. Formal grammar

^۷. John McCarthy

^۸. Alan Turing

^۹. Turing Test

^۱. Corpus linguistics

^۲. Annotation

^۳. Regularities

^۴. Irregularities

^۵. Formal rule systems

اطلاعات و داده‌ها را به سهولت و با سرعت پیدا کند. به عنوان نمونه، (WWW) یک ابر پایگاه متنی محسوب می‌گردد که دسترسی به حجم قابل توجهی از داده‌ها را روزانه برای کاربران سراسر دنیا میسر می‌سازد.

۲- ترجمه ماشینی: که بالأخص در اتحادیه اروپا که در حال حاضر دارای ۲۴ زبان رسمی است، کاربرد بالقوه‌ای در سیستم‌های ترجمه خودکار و حتی نیمه‌خودکار (نیمه اتوماتیک) داشته است.

۳) تولید خودکار متن: که شرکت‌های بزرگی که به تولید محصولات جدیدی از قبیل خودروها، دستگاه‌های پخش دیسک فشرده، تجهیزات کشاورزی و امثال آن مبادرت می‌ورزند بایستی به طور مستمر و مداوم به توصیف محصولات خود و دستورالعمل‌های نگهداری از آن بپردازند.

۴- بررسی خودکار متن: که در این راستا، از بررسی‌گرهای املائی کلمه که بر اساس فهرست صورت کلمه‌ها و تشخیص صورت کلمه بر اساس تقطیع‌گر ساختوازی می‌باشد گرفته تا بررسی‌گرهای نحوی که بر مبنای تقطیع‌گرهای نحوی عمل می‌کنند و خطای موجود در ترتیب کلمات، ساخت دستوری و امثال آن را می‌یابند.

۵- تجزیه خودکار محتوا: که گفته می‌شود اطلاعات مبتنی بر حروف در این برنامه هر ده سال یکبار دو برابر می‌شود. حتی در حوزه‌های خاصی همچون علوم طبیعی، حقوق و اقتصاد، جریان ثابت پژوهش‌های جدید مرتبط به قدری است که پژوهش‌گران و متخصصان زمان کافی برای خواندن همه آن را ندارند. تجزیه خودکار محتوا به شکل خلاصه‌های مختصر بسیار مفید است و به عنوان نوعی پیش شرط^۴ برای نمایه‌گذاری مبتنی بر مفهوم به شمار می‌رود که بازبایی دقیق اطلاعات از پایگاه داده‌های متنی را به عهده دارد.

۶- آموزش خودکار: حوزه‌های گوناگونی از آموزش وجود دارد که در آن، زمان بسیاری برای حل تمرین‌های مکانیکی برای زبان خارجی صرف می‌شود. این کار می‌تواند با رایانه هم انجام شود و سرگرمی بیشتری را برای زبان‌آموزان در حین یادگیری فراهم نماید و نیز زمان بیشتری را در اختیار معلم قرار دهد تا بتواند فعالیت‌های علمی‌تری را به آموزش مکالمه اختصاص دهد. افزون بر این، این سیستم‌ها قادر به تولید پروتکل‌های خودکاری می‌باشند که خطاها و زمان مورد نیاز برای انجام تمرین‌های مختلف آموزشی را با جزئیات شامل می‌شوند.

۷- سیستم‌های اطلاعات و گفت و گوی خودکار: که از

نظریه معناشناسی قالبی بود، گردآوری، مطالعه و ترجمه شد و پس از آن، اهم مطالب آن به شیوه‌ای تدوین شد تا مخاطب را با سیمای کلی این میان رشته‌ی نوظهور که اکنون خود رشته‌ای مستقل به شمار می‌رود آشنا نماید.

یافته‌ها و تحلیل

این بخش در برگیرنده ۳ زیربخش است که در ادامه به بررسی هر یک می‌پردازیم.

کاربردهای زبان‌شناسی رایانشی: به اعتقاد (مکینتاش، ۱۹۹۰) به نقل از موفق (الغبر، ۲۰۰۷)، فاصله‌گذاری خودکار^۱ که منظور از آن ایجاد یک خط تیره کوچک است که یا برای اتصال دو سازه از یک کلمه مرکب به کار می‌رود و یا نشانگر جداسازی کلمه‌ای از انتهای جمله ماقبل آن است، یکی از کاربردهای مهم زبان‌شناسی رایانشی محسوب می‌گردد.

بررسی املائی کلمات، بررسی سبک نوشتار، بازبایی اطلاعات، ارجاع‌دهی^۲ و ترکیب کلمات با یکدیگر از دیگر کاربردهای برجسته و مهم زبان‌شناسی رایانشی به شمار می‌رود (موفق الغبر، ۲۰۰۷). علاوه بر آنچه گفته شد، زبان‌شناسی رایانشی نقش بسزایی در فرایند یادگیری و آموزش زبان دارد که این فرایند خود به سه مرحله تقسیم می‌شود (لی، ۲۰۰۰):

الف - رفتاری

ب - ارتباطی

پ - ترکیبی

و هر مرحله خود نشان‌دهنده سطحی از تکنولوژی و نظریه‌های آموزشی مرتبط با رده سنی آن مرحله می‌باشد.

زبان‌شناسی رایانشی از تقطیع‌گرها^۳ برای تحلیل خودکار زبان بهره می‌جوید. کاری که تقطیع‌گر انجام می‌دهد تجزیه یک نشانه پیچیده^۴ به اجزای اولیه آن، طبقه‌بندی مؤلفه‌ها از طریق جست و جوی واژگانی^۵ و ترکیب مؤلفه‌های طبقه‌بندی شده به واسطه قواعد به منظور دسترسی به تحلیل دستوری آن نشانه پیچیده است. از جنبه روش‌شناختی، کاربرد دستور زبان طبیعی به عنوان تقطیع‌گر حائز اهمیت است چرا که این امکان را به افراد می‌دهد تا کفایت توصیفی سیستم‌های دارای قاعده صوری را بیازمایند. از جنبه کاربردی، تقطیع زبان طبیعی کاربردهای متعددی دارد که از آن میان می‌توان به موارد ذیل اشاره نمود (هاوسر، ۱۹۹۹: ۱۲-۱۳):

۱- نمایه‌گذاری و بازبایی پایگاه‌های داده‌های متنی: که متون را به صورت الکترونیکی ذخیره می‌کند. کاربر پایگاه داده‌ها بایستی بتواند

^۴. Complex Sign

^۵. lexical lookup

^۶. Precondition

^۱. Automatic Hyphenation

^۲. Referencing

^۳. Parsers

متن مبدأ و ویرایش دستی متن مقصد ضرورت نخواهد داشت. دستیابی به این هدف دشوار می‌نمود و هنوز به طور رضایت‌بخشی جز در موارد خاص قابل اجرا نیست (همان).

در حال حاضر، نرم‌افزارهای ترجمه بسیار زیادی وجود دارد، از پروژه‌های بسیار بزرگی که شرکت‌ها و مؤسسات تولید کرده‌اند گرفته تا فرهنگ‌های لغت بسیار ساده‌ای که به طور خودکار کار می‌کنند. با این وجود، کیفیت ترجمه‌های این موارد حتی برای سیستم‌های بزرگی که برجسته‌ترین نوابع آن را طراحی و تولید کرده‌اند به طور چشم‌گیری از کیفیت ترجمه دستی انسان پایین‌تر است (همان).

در واقع، کیفیت ترجمه‌ای که ماشین ترجمه انجام می‌دهد در ترجمه از یک زبان به زبان دیگر و بالعکس یکسان نیست؛ مثل ترجمه از انگلیسی به اسپانیایی و بالعکس. از آنجا که تجزیه و تحلیل خودکار متن معمولاً فرایندی به مراتب دشوارتر از تولید متن است، ترجمه از زبانی که مورد توصیف و بررسی با کیفیت بالاتری قرار گرفته بهتر از ترجمه به آن زبان است. بنابراین، توسعه و تولید دستور زبان اسپانیایی و فرهنگ لغت آن می‌تواند کیفیت ترجمه از اسپانیایی به انگلیسی را بهبود بخشد (همان).

یکی از بزرگ‌ترین مشکلات ترجمه خودکار ابهام‌زدایی معنایی کلمات است. در هر فرهنگ لغت دو زبانه، برای بسیاری از واژه‌های زبان مبدأ، ده‌ها معادل در زبان مقصد وجود دارد. به عنوان نمونه، می‌توان به واژه gato در اسپانیایی اشاره کرد که برای آن معادل‌هایی نظیر جک، دله دزد، ماشه، بطری آب گرم، اشتباه و گربه درج شده است (همان).

مشکل دیگری که در ترجمه خودکار وجود دارد بازبازی اطلاعاتی است که در متن مبدأ به صورت تلویحی بیان شده اما در متن مقصد بایستی به صورت تلویحی مطرح گردد. به عنوان مثال، مشخص نیست از جمله‌ی اسپانیایی زیر کدام یک از مفاهیمی که در ادامه آمده است استنباط می‌شود (همان):

Jose le dio a Maria UN libro. Es interesante.

- He is interesting.
- She is interesting.
- It is interesting.
- This is interesting.

تأثیر فیلمور بر زبان‌شناسی رایانشی: فیلمور یکی از پژوهش‌گران برجسته در زمینه معنای واژگانی و ارتباط آن با متن، دستور زبان، پیکره و محاسبات بود و اثر وی تأثیر شگرفی بر زبان‌شناسی رایانشی داشت. نخستین اثر نظری او در سه دهه

سرویس‌های اطلاعاتی خودکار برای برنامه‌های آموزشی از طریق جست و جو و ذخیره‌سازی در پایگاه داده‌های پزشکی گرفته تا مشاوره خودکار در امور مربوط به مالیات.

به عقیده هاوسر، این فهرست به هیچ وجه کامل نیست چرا که کاربردهای احتمالی زبان‌شناسی رایانشی می‌تواند مشتمل بر تمامی حوزه‌هایی باشد که بشر امروز و در آینده از طریق رایانه و سایر ماشین‌های مشابه با آن ارتباط برقرار می‌کند.

دستاوردهای زبان‌شناسی رایانشی: تا به امروز ماحصل

زبان‌شناسی رایانشی محدود به مواردی می‌شود که در بخش پیشین به آن اشاره شد. مواردی از جمله بازبازی اطلاعات، ترجمه خودکار، تعامل زبان طبیعی، استخراج داده‌های حقیقی، تولید متن، درک زبان طبیعی، تشخیص گفتار و آماده‌سازی متن که خود شامل فاصله‌گذاری خودکار، تشخیص گفتار، بررسی املا، کلمات، بررسی دستور زبان، بررسی سبک و سیاق نوشتار و ارجاع‌دهی می‌باشد (بولشاکوف و گلبوخ، ۲۰۰۴).

از میان موارد فوق، یکی از برجسته‌ترین دستاوردهای زبان‌شناسی رایانشی که امروزه کاربرد فراوانی در میان شرکت‌ها، اقشار دانشگاهی و افراد غیر متخصص در زمینه مطالعات ترجمه دارد، ترجمه ماشینی است که به علت بسامد فراوان استفاده از آن و همچنین اهمیت آن در زمینه‌های گوناگون، در ادامه همراه با ساز و کار و مشکلاتی که دارد به آن پرداخته شده است.

ترجمه ماشینی که از آن به عنوان ترجمه خودکار نیز یاد می‌شود فرایند ترجمه از یک زبان طبیعی به یک زبان طبیعی دیگر است. مطابق با اظهارات (بولشاکوف و گلبوخ، ۲۰۰۴)، تجارت و متون علمی جهان به سرعت رو به افزایش است و بسیاری از کشورها در حوزه‌های علمی و بازرگانی بسیار فعال بوده و کتاب‌ها و مقالات زیادی را به زبان خود منتشر می‌کنند. با رشد روابط بین‌المللی و افزایش میزان همکاری و تعامل میان کشورها، نیاز به ترجمه قراردادهای قانونی، اسناد فنی، دستورالعمل‌ها، تبلیغات و دیگر متون مورد استفاده در زندگی روزانه میلیون‌ها نفر از اهمیت بنیادی و حیاتی برخوردار گشته است. اولین برنامه‌های مربوط به ترجمه خودکار یا همان ترجمه ماشینی چهل سال قبل تولید شد. در آغاز، امیدوار بودند تا از هر متنی ترجمه کلمه به کلمه صورت بگیرد تا جایی که به نظر می‌رسید تنها مشکل موجود فقدان فرهنگ لغت است. با وجود این، پس از اولین آزمایش‌ها، این امید از میان رفت. از آن پس، هدف ایده‌آل ایجاد برنامه‌هایی شد که بتوانند عمیقاً معنای متن دلخواه را از زبان مبدأ استنباط کنند، آن را در یک زبان میانجی جهانی ضبط نمایند و سپس به زبان مقصد با حداکثر میزان دقت و صحت ترجمه نمایند. فرض بر این بود که استنباط دستی

مثالی از مفهوم قالب، قالب معامله تجاری است که عناصر آن شامل خریدار، فروشنده، کالا و پول است (باید توجه داشت که عناصر این قالب از حیث نقش‌های موقعیتی طراحی شده‌اند که خود متفاوت با مفهوم نقش‌های معنایی در دستور حالت است). از میان مجموعه عظیمی از فعل‌هایی که از حیث معنایی به یکدیگر مرتبطند و به این قالب پیوند می‌یابند، می‌توان به فعالیت از قبیل خریدن، فروختن، پرداختن، خرج کردن، هزینه کردن و مطالبه کردن اشاره نمود. فعل خریدن بر خریدار و کالا تأکید دارد و فروشنده و پول را در پس‌زمینه قرار می‌دهد. فعل فروختن بر فروشنده و کالا تأکید دارد و خریدار و پول را در پس‌زمینه قرار می‌دهد. فعل پرداختن بر خریدار، فروشنده و پول تأکید دارد و کالا را در پس‌زمینه قرار می‌دهد و به همین ترتیب است سایر عناصر این قالب. دانش و تجربه‌ای که به واسطه‌ی قالب معامله تجاری ساخته می‌شود، پس‌زمینه و انگیزه برای مقوله‌های بازنمایی شده توسط کلمات را ارائه می‌کند. این کلمات که مواد زبانی هستند، این قالب را از ذهن گوینده و شنونده استخراج می‌کنند (همان).

فرض اساسی در معناشناسی قالبی در ارتباط با معنای کلمات این است که نه تنها کلمه‌های حوزه‌های فنی بلکه تمام کلمات قاموسی^۴ که باید درک شوند خود نیازمند قالب‌های پس‌زمینه‌ای هستند که در آن معنای انتقال یافته انگیزه یا تعبیر می‌شود (فیلمور و بیکر، ۲۰۰۹: ۳۱۷).

تمام مواردی که پیرامون قالب‌های معنایی به آن اشاره شد در پروژه شبکه قالبی فیلمور به نشانی اینترنتی (framenet.icsi.berkeley.edu) برای آن دسته واژه‌های قاموسی که همان مقوله‌های دستوری اصلی فعل، اسم و صفت به شمار می‌روند درج شده است. پروژه مذکور که دستاورد مهم معناشناسی قالبی به شمار می‌رود، ارتباط تنگاتنگی با زبان‌شناسی رایانشی دارد و با درج هر مقوله دستوری تحت عنوان واحد واژگانی^۵ و قالب(های) مرتبط با آن، بافت معنایی هر واحد واژگانی را در اختیار کاربران قرار می‌دهد. به عنوان نمونه، به منظور دستیابی به قالب‌های معنایی یک واحد واژگانی از مقوله دستوری اسم همچون Motion و درج این واژه در کادر جست و جوی شبکه قالبی، مرتبط‌ترین قالب‌ها با واژه مذکور ظاهر می‌شوند: Motion-directional, Motion-noise, Motion-scenario که با کلیک بر روی هر یک از چهار واژه، صفحه جدیدی باز می‌شود که در آن ابتدا تعریفی از قالب مربوطه ارائه

شست، هفتاد و هشتاد میلادی پیرامون دستور حالت^۱ و سپس معناشناسی قالبی^۲ (فیلمور، ۱۹۷۷، ۱۹۸۲، ۱۹۸۵) تأثیر قابل توجهی بر زبان‌شناسی رایانشی، هوش مصنوعی و بازنمود دانش گذاشت. اخیرترین پژوهش وی در دو دهه اخیر شبکه قالبی است که پیکره‌ای حاشیه‌نویسی شده و واژگان رایانه‌ای است که منجر به ایجاد امورات مربوط به درک زبان طبیعی از جمله برجسبدهی نقش معنایی شد (ژورافسکی، ۲۰۱۴).

پروژه شبکه قالبی در حال آماده‌سازی یک منبع واژگانی است که هدف آن، برای بخش عظیمی از واژه‌های انگلیسی معاصر، ایجاد جملاتی با حاشیه‌نویسی‌های نحوی و معنایی است که از آن اطلاعات معتبری پیرامون قابلیت‌های ترکیبی و ظرفیتی هر واژه به دست داده شود. جنبه‌های مهم و محوری کار این پروژه تولید یک شبکه توصیفی براساس قالب‌های معنایی^۳ دربرگیرنده‌ی عناصر قالبی (نقش‌های معنایی) و همچنین مستندسازی مشاهداتش براساس گواهی‌های به دقت حاشیه‌نویسی شده‌ای است که از پیکره‌های الکترونیکی بزرگ گرفته شده است (فیلمور و بیکر، ۱۹۴۷).

معناشناسی قالبی یک برنامه پژوهشی است که به پیوستگی میان زبان و تجربه تأکید می‌کند و چهارچوبی برای ارائه‌ی نتایج این پژوهش در اختیار قرار می‌دهد. منظور از قالب یعنی هر نظامی از مفاهیم که به یکدیگر مرتبطند به شیوه‌ای که درک هر مفهوم مستلزم درک کل نظام است. در معناشناسی قالبی، هر کلمه یک مقوله از تجربه را بازنمایی می‌کند (پتراک، ۱۹۹۵: ۱).

مفهوم قالب در معناشناسی قالبی اشاره مستقیم به قالب‌های حالت فیلمور دارد. در دستور حالت، نقش‌های معنایی موضوعات محمول‌ها نقش مهمی در تعیین مشخصه‌های فعل و بند ایفا می‌کند. قالب‌های حالت به صورت یک صحنه یا موقعیت انتزاعی درک می‌شوند تا جایی که برای درک ساخت معنایی فعل لازم است ویژگی‌های چنین صحنه‌های طحوراه‌ای شده‌ای تحقق یابد. در نخستین پژوهش‌های پیرامون معناشناسی قالبی، میان صحنه و قالب تمایز ایجاد شد که منظور از صحنه یک موجودیت مفهومی و شناختی و قابل تجربه بود و منظور از قالب یک موجودیت زبان‌شناختی. در پژوهش‌های پس از آن، صحنه به کنار نهاده شد و قالب، یک دستگاه ساختاری شناختی تلقی گشت که بخش‌هایی از آن به واسطه‌ی کلماتی که مرتبط با آن بود و در فرایند ادراک به کار می‌رفت نمایه می‌شد (همان).

۴. lexical Words

۵. Lexical Unit

۱. Case Grammar

۲. Frame Semantic

۳. Semantic Frames

شده و سپس عناصر قالبی اصلی^۱ و فرعی^۲ مرتبط با آن قالب به همراه مجموعه عناصر قالبی اصلی^۳ ظاهر می‌گردد. لازم به ذکر است که الگوریتم شبکه قالبی برای بازنمود قالب‌های هر یک از عناصر قاموسی زبان انگلیسی به همین شیوه و ترتیب می‌باشد و تا کنون مشابه این شبکه برای زبان‌هایی همچون ژاپنی، کره‌ای و چینی و چند زبان دیگر ایجاد شده است. جهت آشنایی بیشتر با کارکرد شبکه قالبی فیلمور، شبکه قالبی سایر زبان‌های دنیا و پروژه‌های مرتبط با آن می‌توان به پایگاه این شبکه که پیشتر به آن اشاره شد، مراجعه نمود.

در پایان اشاره‌ای خواهیم داشت به پژوهش‌هایی که تا کنون پیرامون زبان‌شناسی رایانشی در ایران انجام شده است. تا آنجا که نگارنده اطلاع دارد تا کنون پژوهشی این‌چنین که صرفاً به معرفی زبان‌شناسی رایانشی و توصیف ماهیت و کارکرد آن بپردازد منتشر نشده است و تنها منابع مرتبط با این رشته دو کتاب (خسروی‌زاده، ۱۳۹۳؛ شهریاری فرد، ۱۳۹۶) است که نوعی گردآوری مشتمل بر مقالات ارائه شده در همایش محسوب می‌گردد. لذا در بخش منابع فارسی از بخش منابع و مأخذ این مقاله، مقالات انتشار یافته‌ی متمرکز بر موضوعات مرتبط با زبان‌شناسی رایانشی و سیستم‌های هوشمند درج شده است.

نتیجه‌گیری

از بررسی منابع مورد مطالعه پژوهش حاضر چنین نتیجه شد که زبان‌شناسی رایانشی که در ابتدا یک میان رشته تلقی می‌شد هم‌اکنون در میان رشته‌های دانشگاهی و بالأخص زیر شاخه‌های رشته زبان‌شناسی جایگاهی مستقل را به خود اختصاص داده و با بهره‌گیری از ارکان اصلی و بنیادی دو رشته زبان‌شناسی و علوم رایانه‌ای به تحلیل و تبیین پردازش زبان طبیعی پرداخته و نمونه‌ای برجسته از دستاورد آن استفاده از هوش مصنوعی در تولید ماشین‌های خودکار ترجمه از جمله گوگل ترنسلیت (Google Translate) می‌باشد.

دستاورد برجسته معناشناسی قالبی فیلمور (۱۹۷۷، ۱۹۸۲، ۱۹۸۵) که با نام پروژه شبکه قالبی از آن یاد می‌شود و تأثیر آن بر زبان‌شناسی رایانشی قابل اغماض نیست آخرین موضوعی بود که در این مقاله به بحث گذارده شد. واضح است که پژوهش حاضر که یک مطالعه توصیفی به شمار می‌رود، می‌تواند مقدمه‌ای بر پژوهش‌های آتی در حوزه زبان‌شناسی رایانه‌ای و دستاوردهای بعدی و رو به پیشرفت آن قرار گیرد.

^۲. FE Core Set

^۱. Core Frame Elements (CFE)

^۲. Non-Core Frame Elements (NCFE)

فهرست منابع

- Berkeley, California.
- Fillmore, Ch. & Baker, C. (2009). "A frames approach to semantic analysis",
- Fillmore, Charles. J. (1977 a). "The case for case reopened", In P. Cole, & J. Sadock, *Syntax and Semantics 8: Grammatical Relations*, PP. 59-81. New York: Academic Press.
- Fillmore, Charles. J. (1977 b). "Scenes and frames semantics. In A. Zampolli, *Linguistic Structures Processing*", *Fundamental Studies in Computer Science* Vol. 5, PP. 55-81. Amsterdam: North Holland.
- Fillmore, Charles. J. (1977 c). "Topics in lexical semantics", In R. W. Cole, *Current Issues in Linguistic Theory*. Bloomington: Indiana University Press.
- Fillmore, Charles. J. (1982). "Frame semantics", *The Linguistic Society of Korea: Linguistic in the Morning Calm*, PP. 65-173.
- Fillmore, Charles. J. (1985). "Frames and the semantics of understanding", *Quaderni di Semantica*, Vol. 6, PP. 222-254.
- Finch, Geoffrey. (2000). "Key concepts in language and linguistics", Second edition. Palgrave Publishing.
- Hausser, Ronald. (1999). "Foundations of computational linguistics: Human-computer communication in natural language", Springer.
- Jurafsky, Dan (2014). "Obituary: Charles J. Fillmore", In *Association for Computational Linguistics*. Vol. 40(3), PP. 725-732.
- Kay, Martin. (2003). "In Ruslan Mitkov", the oxford handbook of computational linguistics. U.S.A. Oxford University Press.
- Lee, Kuwang-wu. (2000). "English teacher's barriers to the use of Computer-Assisted Language Learning", In the *Internet TESL Journal*. Vol. VI (12).
- McIntosh, Ronald C. (1990). "Hyphenation", Retrived from the World Wide Web: <http://www.hyphenologist.co.uk/book/BOOK-ED3.HTM>
- Muwafaq Al-Ghabra Iman Muwafaq Muslim. (2007). "An introduction to computational linguistics: Advantages & آیت، سیدسعید. (۱۳۸۹). "طراحی و پیاده‌سازی دایفون زبان فارسی برای کاربرد زبان‌شناسی رایانه‌ای"، پژوهش‌های زبان‌شناسی دانشگاه اصفهان، شماره ۲، پیاپی ۳.
- حسابی، اکبر. (۱۳۹۴). "مشکلات انطباق دسته‌های هم‌معنایی فارسی‌نت با دسته‌های هم‌معنای وردنت پرنستون"، پژوهش‌های زبان‌شناسی، شماره ۲، پیاپی ۱۳.
- خسروی‌زاده، پروان. (۱۳۹۳). "مجموعه مقالات دومین هم‌اندیشی زبان‌شناسی رایانشی"، نشر نویسه پارسی.
- شهریاری فرد، آذین. (۱۳۹۶). "مجموعه مقالات چهارمین همایش ملی زبان‌شناسی رایانشی"، نشر نویسه پارسی.
- عاصی، مصطفی. (۱۳۸۰). "نقش رایانه در ایجاد استانداردهای زبانی"، مجله فرهنگ. شماره ۳۷ و ۳۸.
- عاصی، مصطفی. (۱۳۸۳). "پردازش دستوری زبان فارسی با رایانه"، ویژه‌نامه فرهنگستان (دستور)، شماره ۱.
- فلاحی، محمدرضا. و احمدی‌نسب، فاطمه. (۱۳۹۷). "زبان‌شناسی رایانشی: مدل‌ها، منابع و کاربردها"، انتشارات مرکز منطقه‌ای اطلاع‌رسانی علوم و فناوری.
- مرادی، مهدی،، مواجی، وحید، خسروی‌زاده، پروانه. و شیخ‌زاده، مارال. (۱۳۹۱). "فرهنگ توصیفی زبان‌شناسی رایانشی"، انتشارات پردیس دانش.
- ولی‌پور، علیرضا. (۱۳۸۶). "مقایسه و بررسی برخی نظام‌های ترجمه ماشینی روسی و غربی"، پژوهش‌های زبان خارجی، شماره ۴۲.
- ولی‌پور، علیرضا. (۱۳۸۵). "بررسی بعد زبان‌شناختی ترجمه ماشینی و معرفی اجمالی علوم مرتبط با آن"، پژوهش‌های زبان خارجی، شماره ۳۶.
- Bolshakov, Igor. & Gelbukh, Alexander. (2004). "Computational linguistics: Models, resources", applications. Direccion de Publicaciones. Printed in Mexico.
- Clark, Alexander. Fox, Chris. & Shalom, Lappin. (2010). "The handbook of computational linguistics and natural language processing", Wiley-Blackwell Publishing.
- Crystal, David. (2003). "A dictionary of linguistics & phonetics", Fifth edition. Blackwell Publishing.
- Fillmore, Ch. & Baker, C. (1947). "Frame semantics for text understanding", International Computer Science Institute,

- disadvantages”, University of Baghdad. No. 51, Journal of COL OF B. ED, PP. 29-40.
- Petruck, M.R.L. (1995). “Frame semantics and the lexicon: Nouns and verbs in the body frame”, *Essays in Semantics and Pragmatics*. PP. 279-296.
- Richards, Jack C. & Schmidt, Richard. (1985). “Dictionary of language teaching & applied linguistics”, Third edition. Longman Publishing.
- Singh, Gyanendra. Mishra, Ajitanshu. & Dheeraj, Sagar. (2013). “An overview of artificial intelligence”, In *SBIT Journal of Sciences and Technology*. Vol. 2(1), PP. 1-4.
- Zakaria Kurdi, Mohamed. (2016). “Natural language processing and computational linguistics 1: Speech, morphology, syntax”, Wiley Publishing.

Computational Linguistics: A Descriptive Approach towards its Uses and Techniques

Zolfa Imani

Abstract

Man's requirements to the use of Internet and the development of an interdisciplinary field called computational linguistics, which has allocated a separate position to itself by combining two majors such as linguistics and computational sciences on the one hand as well as the interested and the students' needs to getting familiar with and being informed of the function of computational linguistics along with the lack of Persian resources in line with describing the field on the other hand, made the author to write down the results of her research on this major in which it has been introduced and to provide the students particularly the general linguistics researchers and the scholars of the relevant areas with it. Computational linguistics, artificial intelligence as well as natural language processing are considered as the key concepts the description and function of each of which have been elaborated on. Additionally, pointing out the machine translation which is considered as one of the most important achievements of the field and has allocated a high frequency of use among the students and professors especially those ones whose major is not the English Language field of study as well as referring to the frame semantics theory proposed by Fillmore (1975, 1982, 1985) and his project which has had an outstanding impact on computational linguistics, the current article will end.

Keywords: Computational linguistics; Artificial intelligence; Machine translation; Natural Language Processing; Frame Semantics.