

بررسی بُعد زبان‌شناختی ترجمه ماشینی و معرفی اجمالی علوم مرتبط با آن*

علیرضا ولی‌پور**

دانشیار گروه زبان و ادبیات روسی، دانشکده زبان‌های خارجی، دانشگاه تهران، ایران

(تاریخ دریافت: ۸۵/۰۶/۱۳، تاریخ تصویب: ۸۵/۰۷/۰۵)

چکیده

نظام ترجمه ماشینی فعالیتی چند بُعدی است، که در چند حیطه مبتنی بر ابتکار عمل و قدرت حل مسئله است. در این مهارت از برخی دانش‌های بین رشته‌ای نیز استفاده فراوانی می‌شود. ترفندهای هوش مصنوعی با بکارگیری اصول ریاضی، آمار و منطق، قالب مناسبی از زبان را که مناسب برنامه‌نویسی و انتقال به ماشین است، فراهم می‌آورد. در کشور ما این فعالیت دیر شروع شد و اکنون نیز دارای حرکت منسجمی نیست، برای پیشرفت در حیطه ترجمه ماشینی نیازمند احاطه در سه زمینه مختلف است: ۱- داده‌پردازی، توصیف و ساختن مدل‌های مناسب زبان برای کاربرد هوشمند ۲- ساخت و استفاده از نرم‌افزارهای مربوط به این کار ۳- به کار گرفتن فناوری مرتبط. در این مقاله با لحاظ قرار دادن تفکیک این سه زمینه، برخی ابعاد زبان‌شناختی نظام ترجمه ماشینی، براساس محور قرار دادن روش‌های تجربه شده در روسیه تبیین می‌گردد. همچنین به واسطه ارتباط تنگاتنگ ترجمه ماشینی با هوش مصنوعی و برنامه‌نویسی، برخی اطلاعات کلی این علوم، در موارد مرتبط ارائه می‌شود.

واژه‌های کلیدی: ترجمه ماشینی، هوش مصنوعی، داده‌پردازی، لگاریتم، معناشناسی، انفورماتیک.

* این مقاله مستخرج از طرح پژوهشی با عنوان "تجزیه و تحلیل مراحل ساخت و آماده سازی لگاریتم‌ها و کدگذاری لغت‌نامه‌های ویژه ترجمه ماشینی" به شماره ۴۶۰۵۰۰۹/۱/۵ که با حمایت مالی معاونت پژوهشی دانشگاه تهران انجام شده است.

** تلفن: ۰۲۱-۶۱۱۱۹۰۶۹، دورنگار: ۰۲۱-۸۸۶۳۴۵۰۰، E-mail: alreva@ut.ac.ir

مقدمه

ترجمه ماشینی از دستاوردهای مهم هوش مصنوعی است. در این علم از برترین نظریه‌های جدید کاربردی زبان‌شناسی و دانش رایانه‌ای استفاده می‌شود. این پرسش که آیا می‌توان ماشینی اختراع کرد که قادر به انجام رفتارهایی حتی هوشمندانه‌تر از انسان باشد، حتی پیش از پیدایش رایانه‌های قدرتمند و طرح الگوریتم‌های پیشرفته نیز مطرح بود. برخی از اندیشمندان زبان‌شناس مانند وینوگراد (Winograd) بر این باور بودند که این کار شدنی نیست. امروز کسی در پی دادن پاسخ منفی به این سؤالات نیست و شاید پاسخ محتاطانه «نمی‌دانیم» بهترین پاسخ به این نوع سؤالات باشد.

افلاطون و استادش سقراط و شاگردش ارسطو پایه‌های تفکر و فرهنگ غرب را تشکیل می‌دهند. یکی از فیلسوفان معاصر به نام هربرت دریفوس (Herbert Deryfus) می‌گوید که «داستان هوش مصنوعی در اصل از حدود سال ۴۵۰ قبل از میلاد آغاز شده است». زمانی که افلاطون گفت‌وگویی را گزارش می‌کند که در آن سقراط از اپیکور می‌پرسد: «می‌خواهم بدانم ویژگی تقوا چیست که تمامی اعمال را پرهیزگار می‌سازد ... که من ممکن است آن را داشته باشم تا به کار بندم، و از آن به عنوان استاندارد برای قضاوت اعمال شما و دیگر مردمان استفاده کنم». به عبارت دیگر، از سقراط الگوریتمی به منظور تشخیص تقوا از بی‌تقوایی خواسته شده است. ارسطو سعی کرد تا قوانینی را که بر قسمت منطقی ذهن حاکم است، فرموله سازد. او نظامی غیر رسمی از قیاس برای استدلال مناسب ارائه و توسعه داد، که در اساس اجازه می‌دهد بر پایه فرضیات اولیه نتایج به طور مکانیکی تولید شوند. ارسطو به این باور اعتقاد نداشت که تمامی قسمت‌های ذهن توسط پردازش‌های منطقی پوشیده شده و بلکه تصویری از استدلال شهودی داشت. اما ماتریالیسم معتقد است تمامی جهان (شامل مغز و ذهن) مطابق قوانین مادی عمل می‌کنند. ویلهم لایبنیتس (۱۶۴۶-۱۷۱۶) اولین کسی بود که موقعیت ماتریالیستی را به نتایج منطقی‌اش تبدیل کرد و ابزاری مکانیکی برای انجام عملیات منطقی ساخت. اما، فرموله نمودن منطق او چنان ضعیف بود که ماشین تولید مفهوم او نمی‌توانست داده‌های زبانی را پردازش و نتایج جالبی تولید کند (راسل ۸).

با این حال به نظر اندیشمندان در حیطه هوش مصنوعی "از نظر عقلی و طبق روابط آماری و ریاضی و زیرمجموعه‌ای و امکان ترکیبی حروف با یکدیگر، در هر زبانی یک ماشین ساده می‌تواند از ترکیب دو به دو، یا دو به چند و ... کلمات و جملاتی بسازد که بامعنی، بی‌معنا و یا مبهم باشد. مسلماً این ماشین اگر به طور دائم کار کند، در تولیدات خود بهترین

اشعار شاعران نامدار و یا جملات نویسندگان مشهور را نیز بنویسد و حتی ممکن است اشعاری بهتر از اشعار فعلی هم بسراید" (پوپف ۲۰۱). گرچه ماشین این کار را انجام می‌دهد، اما ضایعات آن فراوان است، چون هوش به امانت گذاشته در آن کم است، اما اگر این هوش ارتقاء یابد و حدود ترکیبات برای آن دقیق‌تر تعریف شود، ضایعات کمتر خواهد شد. زیرا فقط انسان است که هوش مبتکرانه دارد و هیچ ابر رایانه‌ای از چنین موهبتی، حتی به صورت جزئی، برخوردار نیست. اما انسان می‌تواند با ایجاد شرایط و سازوکارهایی، شرایط تحقق مصنوعی هوش را فراهم سازد، و بسیاری از پروسه‌های هوشمند را به رایانه بسپارد. استفاده از نرم‌افزار و سخت‌افزار این امکان را دست‌یافتنی ساخته و تنها موضوعی که هنوز در هیچ یک از حیطه‌های هوشمند و از جمله ترجمه ماشینی به تکامل لازم نرسیده، ولی با شتاب فزاینده در حال رشد است، مسئله چه‌گونگی طراحی و انتقال هوش انسان به ماشین است.

بحث و بررسی

الف) تاریخچه مختصر مسئله

ابزارهای محاسباتی قبل از پیدایش رایانه وجود داشت؛ چرتکه قدمتی ۷۰۰۰ هزار ساله دارد. در اواسط قرن هفدهم میلادی، پاسکال ماشینی مکانیکی به نام پاسکالین برای جمع و تفریق اعداد ساخت. لاینیتس این ماشین را در سال ۱۶۹۴ با ساخت ابزار مکانیکی برای ضرب از طریق جمع مکرر توسعه داد. پیشرفت در این زمینه برای بیش از یک قرن متوقف شد، تا آن‌که چارلز بابیج (۱۷۹۲-۱۸۷۱) به این اندیشه رسید که جداول لگاریتمی می‌توانند توسط ماشین محاسبه شوند. او برای این کار ماشینی طراحی کرد که هرگز ساخته نشد. به جای آن او به طراحی موتور آنالیتیکی پرداخت و بدین‌گونه بابیج ایده‌های حافظه قابل آدرس‌دهی، برنامه‌های ذخیره شده و پرش‌های شرطی را مطرح کرد. اگرچه ایده ماشینی قابل برنامه‌ریزی تازه نبود، چرا که در سال ۱۸۰۵ جوزف ماری جاکارد ماشینی بافندگی درست کرد که می‌توانست با استفاده از کارت برنامه‌ریزی شود، اما ماشین بابیج اولین تولید مصنوعی بود که قابلیت محاسبات عمومی را داشت. همکار بابیج، آدا لولیس، دختر شاعر سرشناس لرد بایرون، برای موتور آنالیتیکی برنامه‌ای نوشت و حتی پیش‌بینی کرد که ماشین قادر به بازی کردن شطرنج و خلق موسیقی خواهد بود (راسل ۱۵). اما تا مدت‌ها تحقق ایده‌های مرتبط با هوش مصنوعی در حد بازی و سرگرمی و آرزوهای دست‌نیافتنی باقی ماند.

امروز ما می‌دانیم که ایده هوش مصنوعی، ایده‌ای درست و ضروری برای تحقق آرزوهای اندیشمندان است، اما ابزار تحقق آن مکانیک نیست و نمی‌تواند باشد، زبان‌شناسی، منطق، آمار و ریاضی بر حول مکانیک، از امکان بهره‌برداری از دستاوردهای خود در تحقق ترجمه ماشینی در مانده بودند. اما با ظهور پدیده‌ای به نام دیجیتال این روند تغییر پیدا کرد. این روند کند و ناامید کننده در رابطه با هوش مصنوعی، تا دهه چهارم قرن بیستم ادامه داشت، اما پس از دهه چهارم با اختراع رایانه، با شتاب حیرت‌انگیزی به پیش می‌تازد. هر روز که می‌گذرد، سرعت و ظرفیت سخت‌افزارها افزایش می‌یابد و قیمت‌ها نیز کاهش می‌یابد. پیشرفت در علوم رایانه‌ای به طور فوق‌العاده‌ای موفقیت‌آمیز بوده است و به نظر می‌رسد، این رشد فزاینده پایانی نداشته باشد. رایانه‌های موازی می‌توانند محاسبات بسیار سنگین و پیچیده‌تری را به شرط دادن محتوی درست و استفاده از دستورات محاسباتی برای به دست آوردن خروجی بهتری انجام دهند. دستاوردهای مشترک هوش مصنوعی با علوم دیگر نیز در حال گسترش است و از کشاورزی تا هوا و فضا را متحول ساخته است. زبان و زبان‌شناسی نیز از این جریان مثبت نمی‌بایست جدا بماند، و این دگردیسی و سرمایه‌گذاری و تغییر یا ارتقاء پژوهش‌های زبانی از جایی می‌بایست شروع می‌شد که متضمن سود و پیشرفت در حیطه زبانی باشد.

تردیدی وجود ندارد که پیشرفت‌های علمی، در حد بسیار زیادی، وابسته به ترجمه و انتشار اطلاعات از طریق آن است. در اوایل دهه هفتاد میلادی در مرکز ترجمه روسیه سالیانه حدود ۶ هزار صفحه مطلب ترجمه می‌شد و در اواخر دهه ۷۰ این رقم به ۱۰۰ هزار صفحه رسید که این تقاضا همچنان در حال افزایش است.

در جامعه اقتصادی اروپا هشت زبان رسمی وجود دارد که همه مدارک و اسناد باید به این زبان‌ها ترجمه شوند. طبق آمار ارائه شده، در سال ۱۹۷۸ تعداد خدمه ترجمه و زبان‌شناس در این سازمان ۱۳۰۰ تن بوده‌اند که در یک سال ۶۰۰۰۰۰ صفحه متن ترجمه کرده‌اند، اکنون مترجمان به ۲۵۰۰ تن رسیده‌اند، ولی باز هم کفاف نمی‌کند و کمبود وجود دارد (مارچوک ۵). غیر از نیاز امروز جوامع صنعتی برای دستیابی و تبادل اطلاعات، حذف مانع زبانی در ارتباط بین انسان‌ها آرزوی همگانی بشریت بوده است. بنابراین انگیزه‌های کافی و قوی و حتی دائمی برای تحقق این هدف وجود داشته و دارد. هیچ فرد و یا دولتی بی‌نیاز از کسب اطلاعات مختلف نیست، امروزه داد و ستد اطلاعات، گسترده‌تر و سودمندتر از گذشته شده است و این روند ادامه خواهد داشت. تلاش برای ساخت نظام‌های ترجمه ماشینی در روسیه و غرب،

سابقه‌ای فراتر از نیم قرن دارد، با این حال، روند این تحقیقات و تجربیات دارای فراز و نشیب و تغییر و اتخاذ سازوکارهای مختلف بوده است. از سوی دیگر، علم ترجمه ماشینی که تاریخچه کاربرد عملی و حرفه‌ای آن، مانند موضوع هوش مصنوعی، کمی افزون‌تر از نیم قرن است، محل تلاقی و در هم تنیده شدن علم انفورماتیک و به خصوص مهندسی نرم‌افزار با زبان‌شناسی است. این یکپارچگی به گونه‌ای بوده که در برخی موارد مرز بین این علوم فرو ریخته است و برخی زبان‌شناسان و یا مهندسان نرم‌افزار به تنهایی و مستقل از یکدیگر سعی در تحقق این هدف داشته‌اند. با این حال تجربه‌های موفق این کار، حاکی از مشارکت زبان‌شناسان با مهندسان است، ضمن آن که اطلاع از سازوکارهای طرفین و رعایت تقسیم کار تخصصی، موجب تسهیل و ارتقاء کار خواهد شد. به طور کلی برای تغییرات و تحولات نظام‌های ترجمه ماشینی سه دوره وجود داشته است:

ویژگی‌های اولین دوره: ۱۹۵۷-۱۹۴۶

۱- زبان به عنوان کد مطرح شد.

۲- لغت‌نامه‌هایی مخصوص این کار ساخته شد.

۳- ساختن اولین لگاریتم و برنامه نرم‌افزاری تحقق یافت.

۴- نظام‌های ارائه شده، کارایی حرفه‌ای نداشتند و فقط جنبه تبلیغی و معرفی کار را به صورت کلیشه‌ای ارائه می‌کردند. یعنی معلوم بود توانایی این دستگاه چیست، جمله هم طوری می‌بایست داده می‌شد که ماشین در آنالیز آن به مشکل بر نمی‌خورد، جملات باید به شکل لازم به ماشین داده می‌شد و بعد هم ویراستاری می‌شد.

ویژگی‌های دومین دوره: ۱۹۶۷ - ۱۹۵۷

۱- طرح ضرورت ساختن نظامی که لگاریتم‌ها در آن نقشی اساسی داشته باشند.

۲- از روش‌های آماری استفاده شد.

۳- تحقق شرح و توصیف مبانی دستوری مناسب ماشین، که با دستور سنتی تفاوت داشت و کلمات براساس معیارهای جدیدی تقسیم می‌شدند.

۴- توجه از لغت بیشتر به نحو معطوف شد.

۵- ایده‌هایی برای تقسیم‌بندی‌های معناشناسی ویژه ماشین مطرح شد.

۶- از لگاریتم‌های نحوی و معنایی در نظام ترجمه ماشینی استفاده شد.

۷- روش‌هایی مدنظر قرار می‌گرفت که امکان ترجمه از یک زبان به زبان‌های گوناگون میسر باشد.

۸- استفاده کاربردی، اما محدود، از ترجمه ماشینی در برخی سازمان‌ها شروع شد.
 ۹- نیاز به علوم دیگری مثل ریاضی و برنامه‌نویسی و گسترش لگاریتم‌ها احساس شد.
 ۱۰- لغت‌نامه‌های بسامدی ساخته شد و برنامه‌ریزی براساس آمارهای آن تدارک دیده شد.

۱۱- موفقیت کامل در ترجمه صنعتی به دست آمد.
 ۱۲- با بالا رفتن سرعت رایانه، امکان بهره‌مندی از برنامه‌نویسی‌های هوشمند فراهم می‌شد.

دوره سوم از سال ۱۹۶۷- همراه با شروع پیشرفت در علم انفورماتیک بوده است. رشد سریع فناوری رایانه‌ای، امکانات بهتری را در اختیار برنامه‌نویسان قرار داد. البته روند کار تا سال ۱۹۷۶ کند بود (مارچوک ۱۹۸۳، ۲۰).

از موضوعات مهم بر سر راه ساخته شدن ترجمه ماشینی در سطح واژگان، مشکل چند معنایی بودن واژه‌ها، شناخته شده‌ترین ویژگی زبان‌های طبیعی است (آپرسیان ۱۹۸۷، ۴۷) و این امر زائیده قانونی است که در زبان‌شناسی تحت عنوان اقتصاد کلامی (economy language) اشتهار یافته است. مترجم به هنگام برخورد به واژه‌ای که دارای چند معنا است، با تکیه بر توانمندی زبانی (competence) خود و بافت و موضوع متن، می‌تواند معادل واقعی واژه را تشخیص دهد و برگزیند، اما رایانه فاقد چنین توانمندی است. بنابراین طراح ماشین ترجمه ناگزیر است مقدماتی فراهم سازد و در الگوریتم به تمهیداتی دست یازد که از طریق آن بتواند خلأ موجود را ترمیم کند. این تمهیدات باید از نخستین مراحل پردازش، یعنی از مرحله واژگان و نحو و ساختار آغاز شود و تا آخرین مرحله، یعنی مرحله تجزیه و تحلیل معنانشناختی متن تداوم یابد. اما مشکل کلمات چند معنا^۱ (полисемия) حل نشده بود، افزون بر آن، علی‌رغم تلاش فراوان کدگذاری و ساختن گرامر مناسب که شامل هم صرف و هم نحو بود - در معنانشناسی ماشینی رشد چندانی روی نداده بود و حتی انتخاب معنا در همکاری انسان و ماشین جنبه تخصصی داشت و بهره‌وری عمومی مورد نظر نبود. بنابراین در کنفرانس ماشین ترجمه که در سال ۱۹۷۷ در لوکزامبورگ برگزار شد، به این نتیجه رسیدند که روش

۱- کلمه полисемия در زبان‌شناسی معنای عام و فراگیرتری نسبت به многозначность به معنای چند معنایی دارد، با این حال در این مقاله مراد کلمات و نشانه‌هایی‌اند، که شکلی یکسان، اما معنی و نقشی متفاوت دارند. مانند نشانه "b" در دوم شخص مفرد افعال روسی در مقابل همین نشانه به عنوان تلفظ نرم، و یا مواردی مانند کلمه лук با معانی مختلف آن.

انسانی و ماشینی به طور توأم مورد توجه قرار گیرد.

سرانجام از سال ۱۹۷۷، دست‌اندرکاران نظام‌های ترجمه ماشینی رویه محدودیت را برگزیدند، که طبق آن، همان‌طور که شعر و رمان را در مراحل قبلی برای ترجمه ماشینی ثقیل تلقی کرده بودند، موضوع مترجم کلی هم ایده مناسبی در حیطه این نظام تشخیص داده نشد و تحقق آن به آینده نه چندان نزدیک موکول شد.^۱ در روز ۲۶ ماه می ۱۹۷۸ در نیویورک، سازمان ملل متحد، گردهمایی کارشناسان امور مربوط به ترجمه ماشینی بود و نظام‌های کارآمد معرفی و به نمایش گذاشته شدند. کمپانی کانادایی ورد ترانسلیشن (World Translation of Canada) نظام ترجمه ماشینی خود را تحت عنوان سیستم ارائه داد. این نظام ابتدا در دانشگاه جورج تاون در سال ۱۹۷۴ طراحی شد و سپس با جلب مشارکت مؤسسات تحقیقاتی دیگر تکمیل شد. بسیاری از زبان‌شناسان، خوش‌بینانه استفاده از نظریات دستور گشتاری و ساخت‌های نحوی اندیشمندانی مانند چامسکی را بسیار مهم و کارگشا معرفی کردند، زیرا به اعتقاد ایشان "ترجمه واژه به واژه برای طراحی رایانه‌ای که قادر باشد متنی را از یک زبان به زبان دیگر ترجمه کند، غیر عملی و غیر ممکن است". (دبیر مقدم ۱۲۰).

شوروی‌ها از پیشگامان این عرصه در دهه پنجاه میلادی بودند، و پژوهش‌های خود را در زمینه بررسی واژگان و معناشناسی در انستیتوی مرکز اطلاعات علمی وینیتی (Винити) و آزمایشگاه مدل‌سازی الکترونیک انجام می‌دادند. در همین سال‌ها بلسکایا (Белская) الگوریتمی جامع برای ترجمه ماشینی دوسویه انگلیسی - روسی ارائه داد. در طی دهه شصت، زبان‌شناسان فرمالیست روسی در زمینه‌های متعدد موضوع مورد بحث، تحقیقات ارزنده‌ای را سامان دادند. غنا و عمق پژوهش‌های گسترده آنان در زمینه نشانه‌شناسی و تجزیه و تحلیل نشانه‌ها، رابطه متن و معنا، تجزیه و تحلیل‌های مورفولوژیکی، نحوی و معناشناختی، بیانگر آن بود که روس‌ها مسائل و مشکلات مربوط به ابهام‌های واژگانی و معنایی را حل کرده بودند. همزمان در دانشگاه ایالتی مسکو نیز گروهی به سرپرستی کولاگینا (Кулагина) و مولوشنایا (Молошная) الگوریتمی قوی در زمینه تجزیه و تحلیل مورفولوژیکی و ساختاری ابداع کردند. ملچوک (Мельчук) زبان‌شناس نامدار روسی در انستیتوی پژوهش زبان‌های خارجی

۱- از نظر عقل و منطق بدون تردید با تقویت دستاوردهای هوش مصنوعی، افزایش سرعت و حافظه رایانه و برخی عوامل دیگر، ترجمه ماشینی هم مانند نان ماشینی، یا خدمات جراحی ماشینی و... جایگزین ترجمه انسانی می‌شوند. این امکان دست نیافتنی نخواهد بود که تصورات ژول ورن گونه پوپف تحقق یابد و در کنار ترجمه ماشینی، ادبیات ماشینی نیز بوجود آید، ماشین‌هایی که قادر خواهند بود، داستان و شعر تولید کنند.

وینتی، اولین کسی بود که مؤلفه‌های معنایی را در ترجمه ماشینی به کار بست، تا سرانجام نخستین ماشین ترجمه کاربردی شوروی‌ها در سال ۱۹۷۶ به ثمر نشست و مورد استفاده همگان قرار گرفت (خداپرستی ۳) نکته مهم آن که راز و رمز نظام ترجمه ماشینی پوشیده ماند تا زمانی که ملچوک^۱ و یکی از همکارانش به غرب پناهنده شدند. پس از ملچوک موضوع ترجمه ماشینی در روسیه برای مدتی با رکود مواجه شد، اما با تلاش و همت پروفیسور یوری نیکلایویچ مارچوک تحقیقات قبلی ادامه یافت و تکمیل شد. پروفیسور یوری نیکلایویچ مارچوک روش کار را تا حد فراوانی تصحیح و ارتقاء داد، او با اصل قرار دادن رویه معنانشناختی ملچوک (متن - معنا) از نظریه‌های فرمالیست‌هایی مانند چامسکی نیز استفاده کرد. تحقیقات تیم ترجمه ماشینی «آپار»^۲ از اوایل دهه هفتاد تا اواخر آن (۱۹۷۹) منجر به آماده کردن و شروع بهره‌برداری دولتی از این نظام ترجمه ماشینی در برخی مراکز شوروی شد. این تیم تحقیقاتی^۳ دارای چند تیم زیرمجموعه‌ای مرتبط با هم بودند که به شرح ذیل بوده است:

- ۱- تیم برنامه‌نویسی رایانه‌ای در محیط پرولوگ^۴ به سرپرستی دکتر یو. ا. ماتورین (Ю. А. Моторин)، کارشناس انفورماتیک.
- ۲- تیم مسئول ساخت لغت‌نامه‌های متنی به سرپرستی دکتر و. ای. شربینین (В. И. Щербинин).
- ۳- تیم مسئول ساخت لگاریتم و پارسر (تحلیل‌گر نحوی)، به سرپرستی دکتر آ. ن.

۱- اگر فرمالیست‌های جهان را به دو دسته تقسیم کنیم، چامسکی مقابل ملچوک قرار می‌گیرد. مارچوک معتقد است، آنقدر که در عمل، ایده‌های معنانشناختی ملچوک به ترجمه ماشینی کمک کرده، ایده دستور همگانی و جهانی چامسکی کاربرد چندانی نداشته است. نگاه شود به کتاب تحلیل نحوی-معنایی مارچوک:

Марчук Ю. Н. Синтактико-семантический анализ в системе машинного перевода АМПАР.-
Международный форум по информации и документации. М.: ВИНТИ, 1980, том 5, № 2,
с. 15-17.

2- Автоматизированный Машинный Перевод с Английского на Русский

۳- این گروه پس از اتمام طرح تحقیقاتی «آپار»، طرح مشابه‌ای را به نام «نریا» (НЕРПА) (نظام ترجمه ماشینی روسی - آلمانی) را نیز به نتیجه رساندند که با استفاده از تجربیات «آپار»، ساخته شده است. این تیم در حال حاضر با استفاده از تجارب به دست آمده، نظام‌های دیگر ترجمه ماشینی را نیز به وجود آورده‌اند که نظام فرانسه به روسی آن تحت عنوان «فراپ» (ФРАП)، تجربه نوین این تیم است و روش ساخت آن با موارد قبلی تا حدی متفاوت است.

۴- به گفته پروفیسور مارچوک، پرولوگ مورد استفاده در این تحقیق با پرولوگ موجود در بازار تفاوت دارد و با تغییرات لازم مناسب و تنظیم برای محاسبات و برنامه‌نویسی زبان طبیعی شده است.

کیسیلف (А. Н. Киселевым).

متأسفانه ارزیابی‌های زبان‌شناسان ایرانی از پدیده ترجمه ماشینی منفی و گاهی نیز مبهم بوده است. در یکی از آثار سه دهه قبل زبان‌شناسی ایران آمده است که "ترجمه ماشینی به مفهوم و مقیاس تجارتي فعلاً قابل عرضه نیست، زیرا نه کیفیت آن رضایت‌بخش است و نه از نظر اقتصادی مقرون به صرفه است." (باطنی ۱۰۷) در صورتی که در همان زمان نظام‌های ترجمه ماشینی مانند آمپار در روسیه و سیستران در کانادا به طور حرفه‌ای مورد استفاده قرار گرفته بودند و با شتاب رشد مثبت خود را دنبال می‌کردند.

اما ناراحت‌کننده و عجیب آن که، مدت‌ها بعد در اولین چاپ فرهنگ اصطلاحات زبان‌شناسی دوباره همین ارزیابی مبهم، کلی و غلط از ترجمه ماشینی ارائه شده است! خواننده، منفی و مبهم بودن را در پاراگراف زیر می‌تواند مشاهده کند. ساغرونیان در فرهنگ اصطلاحات زبان‌شناسی نوشته است، "ماشین ترجمه یا به عبارت دیگر حسابگر الکترونی! ماشینی است که یک رشته علائم را طبق دستوراتی که قبلاً دریافت داشته به یک رشته علائم دیگر تبدیل می‌کند،" پس از این توضیحات مبهم و کلی اضافه کرده که "پیش‌ویرایش و پس‌ویرایش به کیفیت ماشین ترجمه کمک می‌کند" (۷۷). در صورتی که این موضوع تا حدی در نظام‌های دوره اول، در برخی مدل‌ها ابداع شده بود و به علت عدم کارایی، به خصوص موضوع پیش‌ویرایش آن، به کلی منسوخ شد. وی به نقل و استناد از باطنی تکرار کرده است "ترجمه ماشینی به مفهوم و مقیاس تجارتي فعلاً قابل عرضه نیست، زیرا نه کیفیت آن رضایت‌بخش است و نه از نظر اقتصادی مقرون به صرفه است!" (همان ۷۶).

ب) ساختار زبان‌شناختی نظام ترجمه ماشینی

در ساختار زبان‌شناختی نظام‌های ترجمه ماشینی، سه مؤلفه اصلی وجود دارد و هر مؤلفه نیز به نوبه خود از مجموعه اجزایی تشکیل می‌شود که در راستای تحقق نقش یا نقش‌های آن سه مؤلفه اصلی عمل می‌کنند. مؤلفه‌های فوق‌الذکر در تناظر با سه سطح که علم زبان‌شناسی نوین برای هر زبان طبیعی و زنده قائل است، قرار دارند. این سه سطح عبارتند از:

۱- واژگان ۲- نحو و ساختار ۳- معناشناسی

باید همواره این نکته را در نظر داشت که تفکیک این سطوح مبین عدم ارتباط بین مختصات واژگان، نحو و ساختار و معناشناسی متن نیست، بلکه هر سه سطح در سلسله مراتب چند بعدی خود، در حکم تار و پود یکدیگرند (خدایپرستی ۳).

در این بخش روش‌های به کار گرفته شده برای این سه سطح در حد حوصله این مقاله تشریح می‌شوند. لازم به یادآوری است که برخی سطوح بر دیگر سطوح مقدم بوده و پیش‌نیاز سطوحی است که در بسیاری کاربردهای پردازش زبان طبیعی، مورد استفاده قرار می‌گیرند. از جمله این سطوح پایه‌ای می‌توان به ایجاد منابع زبانی، تحلیل گره‌های لغوی، ساخت واژگی و نحوی اشاره داشت، به همین دلایل برای دستیابی به فناوری ترجمه ماشینی احتیاج به تأمین پیش‌نیازهای این فناوری در دو بعد زبانی و برنامه‌نویسی است، آنچه که به بحث ما، یعنی طرف زبانی این پیش‌نیازها مربوط می‌شود، تأمین کمبودها در زمینه لغت‌نامه‌های گرامری، بسامدی، متنی و نشانه‌گذاری شده است، موضوع مهمی که در کشور ما سابقه پرداختن به آن بسیار ناچیز و ضعیف است. واضح است که اگر ما درصدد ارتقاء سطح استفاده از زبان در ابعاد جدیدتر و امروزی باشیم، چاره‌ای نداریم که زیرساخت‌های محکم و نوینی به وجود آوریم، تا اطلاعات اولیه و ضروری برای پردازش‌های مورد نیاز زبانی و یا برای استفاده در ارتباط دوسویه با ماشین فراهم آید.

همان‌طور که گفته شد، ایجاد نظام ترجمه ماشینی فعالیتی چند بعدی در حیطه هوش مصنوعی است، که از دستاوردهای مجموعه گسترده و بعضاً متفاوتی در بالاترین سطوح هوش مصنوعی بهره‌بردار می‌کند. در ساخت نظام ترجمه ماشینی ترندهای هوش مصنوعی با بکارگیری اصول ریاضی، آمار و منطق، قالب مناسبی از زبان را که مناسب برنامه‌نویسی و انتقال به ماشین است فراهم، مرتبط و تحقق می‌بخشد. به عبارت دیگر، ترجمه ماشینی که تاریخیچه کاربرد عملی و حرفه‌ای آن، مانند خود موضوع هوش مصنوعی کمی افزون‌تر از نیم قرن است، محل تلاقی و در هم تنیده شدن علم انفورماتیک و به خصوص مهندسی نرم‌افزار با زبان‌شناسی است. لازم به یادآوری است، در تجربیات موفق نظام‌های ترجمه ماشینی در روسیه و اروپای غربی و حتی آمریکا، غیر از وظایف تخصصی، موضوع مدیریت کل پروژه‌ها با افراد زبان‌شناس بوده است. این موضوع دلایل متقنی دارد، که ارائه جزئیات آن در حوصله این مقاله نیست. با این حال در ساختن نظام ترجمه ماشینی، تیم‌های زبان‌شناس وظایف مربوط به خود را دارند، که شامل تقریر و نگارش انواع لغت‌نامه‌های مورد نیاز و دستورات تبیین‌کننده صرفی، نحوی و معناشناسی و تهیه لگاریتم‌های دستوری در موضوعات ذکر شده است. لغت‌نامه‌های نظام‌های ترجمه ماشینی با لغت‌نامه‌های عادی و عمومی متفاوت است، اگرچه در تألیف برخی از آنها از اطلاعات موجود در لغت‌نامه‌های عادی استفاده می‌شود، اما برخی دیگر صرفاً برای کاربرد ماشینی استخراج و تهیه می‌شوند. مانند لغت‌نامه‌های متنی، بسامدی، ریشه - بن،

پیشوندها - میانوندها و پسوندهای معناساز یا دستوری و همانند آنها.

۱- واژگان نظام ترجمه ماشینی

لغت‌نامه‌ها در نظام ترجمه ماشینی نقش مهمی را ایفا می‌کنند، همان‌طور که دستور زبان مورد نیاز در این نظام با دستور زبان معمولی متفاوت بوده و دارای گستردگی و پیچیدگی‌های فراوانی است. لغت‌نامه نظام ترجمه ماشینی نیز با لغت‌نامه‌های عادی در ابعاد مختلف متفاوت‌اند، این تفاوت هم در استخراج مدخل لغات، پردازش، کدگذاری و هم در کاربرد آنها وجود دارد. لغت‌نامه‌های موجود در نظام ترجمه ماشینی دارای تفاوت‌های اساسی با لغت‌نامه‌های مورد استفاده عام و غیر ماشینی است، اولاً در این لغت‌نامه‌ها، لغات کدگذاری می‌شوند و براساس کد ذخیره شده، بازخوانی می‌شوند و ثانیاً، دارای این ویژگی‌اند که در مراحل مختلف عملکرد نظام ترجمه ماشینی، به ترتیب و تفاوت مورد استفاده قرار می‌گیرند. کدگذاری و تدوین لغت‌نامه‌های ماشینی باید به گونه‌ای باشد که به علت مراجعات متعدد نظام به آن، به موضوع برتری ذاتی آن که امکان بهره‌برداری با سرعت بالا است، صدمه‌ای وارد نشود.

روش‌های دسته‌بندی و ثبت لغات در لغت‌نامه‌های نظام ماشینی (با اقتباس از شیوه به‌کار رفته در نظام ماشینی آمپار روسی)

در تقسیم بندی اولیه، کلمات به هفت دسته و یک گروه مجزا تقسیم می‌شوند. این تقسیم‌بندی‌ها کلیشه‌ای و با توجه به روش اتخاذی در نظام ترجمه ماشینی برگزیده می‌شوند، بنابراین در نظام‌های مختلف دسته بندی‌های متفاوتی وجود دارند. این تفاوت‌ها به علل مختلف‌اند اما شباهت‌هایی نیز دارند، که همگی به منظور تسهیل تجزیه و تحلیل نظام ترجمه ماشینی است. سبک تشریح شده توسط ما روش اتخاذی پروفیسور مارچوک مجری طرح ترجمه ماشینی آمپار است.

دسته هفتگانه عبارت است از: ۱- اسم ۲- صفت ۳- حرف اضافه ۴- فعل ۵- قید ۶- حروف ربط ۷- علائم نقطه‌گذاری. و گروه مجزا کلماتی، که اصطلاحاً به آنها

آموگراف^۱ می‌گویند (کلماتی که دارای دیکته‌ای واحدند، اما دارای معنی و یا نقش دستوری متفاوتی‌اند). برای هر کدام از این دسته‌ها یک عدد قراردادی درج می‌شود و برای هر دسته نیز زیرمجموعه‌ای از گروه‌های مرتبط آن لحاظ و شماره‌گذاری می‌شود. در داخل هر دسته اسم، برای مثال دسته‌های معنایی، دستوری مشخص می‌شود و قید می‌گردد که آیا روش استخراج معنای این اسم منوط به لحاظ کردن متن گنجانده شده در آن است و یا وابستگی به متن ندارد. اسم در نظام ترجمه ماشینی، می‌تواند یکی از گروه‌های زیر باشد:

۱- ضمیر ۲- عدد ۳- جغرافیایی ۴- اسم ارگان یا سازمان ۵- انتشاراتی ۶- اسم جاندار
۷- اسم مدرک ۸- واحد پول و ...
با لحاظ کردن همه نشانه‌های متصور، نشانه‌های جداکننده آن از دیگر دسته‌ها به دست می‌آید.

فعل را براساس ویژگی‌های ریخت‌شناسی (صرفی دستوری) آن به سه دسته تقسیم می‌کنند: نقش نحوی افعال در مرحله تحلیل نحوی استخراج می‌شود.

۱- افعالی که ریشه حال و یا آینده دارند.

۲- افعالی که ریشه زمان گذشته را دارند.

۳- وجود نشانه صفت فعلی زمان گذشته (причастие прошед).

نشانه‌های نقطه‌گذاری به نشانه‌هایی که عبارت را جدا می‌کنند، مانند ویرگول، نقطه، پرانتز، گیومه و غیره تقسیم و نقش دستوری آن‌ها کدگذاری می‌شود.

پس از تعیین مراتب درجه‌بندی کلمات هفتگانه ذکر شده، می‌بایست تکلیف کلمات آموگراف (تک‌دیکته‌ای) معلوم شود. این‌ها (آموگراف‌ها) را به شرح ذیل می‌توان تقسیم‌بندی کرد:

الف) ابهام دوتایی

۱- دسته مبهم بین اسم و فعل

۲- فعل و صفت

۳- کلمه ربط و اسم

ب) ابهام سه‌تایی

۱- در دسته آموگراف‌ها، انواع دیگری از آمونیم‌ها نیز قرار می‌گیرند، مانند آموفون‌ها (омофоны) و آموفرم‌ها (омоформы)، اما از آنجایی که این دو دسته بواسطه ابهام آوایی و یا پادژی (حالتی) بروز می‌کنند، در این تقسیم‌بندی‌ها مورد توجه قرار نمی‌گیرند.

۱- دسته فعل - حرف ربط - حرف اضافه

۲- فعل - صفت - حرف ربط

ج) ابهام چهارتایی

۱- فعل - اسم - صفت - حرف اضافه

باید یادآور شویم که دسته‌های مبهم بالای سه عضوی، محدودند و تعداد آن‌ها زیاد نیست، ولی باید در محاسبات آورده و برنامه‌ریزی شوند. حتی برخی حروف دارای ابهام معنایی و دستوریند که باید مشخص شوند. برای مثال، حرف *a*، که هم جزو حروف الفبا است و هم دارای معانی مختلف دستوری و حتی معنایی است. پس از این تقسیم‌بندی، همه کلمات از نظر ساختار ظاهری و صرفی دسته‌بندی می‌شوند و زیرمجموعه‌های آن‌ها مشخص می‌شود. برخی از این دسته‌ها ممکن است عضو بسیار محدودی داشته باشند و برخی حتی در شروع راه‌اندازی نظام عضوی نداشته باشد، با این حال پیش‌بینی تفکیک در مدل به تسهیل کار و ادامه تکمیل آن کمک خواهد کرد.

تقسیم بعد، تقسیم کلمات به کلماتی است که دارای معنای واحدی (однозначность) بوده و یا دارای معنای چندگانه (многозначность) می‌باشند. برای مثال، اکثر اسامی خاص و یا اصطلاحات علمی معنی واحدی دارند. این تقسیم‌بندی در نظام ترجمه ماشینی اهمیت فراوانی دارد و باید همه کلمات از این حیث کدگذاری شوند که آیا چند معنایی‌اند یا خیر. فقط برای آموگراف‌ها دیگر ضرورت ندارد، زیرا وقتی لگاریتم تشخیص کلمات روسی замо́к - му́ка - мука و یا کلمات فارسی مُو و مُو تعیین شد، ضرورتی به تعیین چند معنایی در آن‌ها وجود نخواهد داشت. این تقسیم‌بندی‌های جهانی نیست و ممکن است از زبانی به زبان دیگر متفاوت باشند، مثلاً ممکن است در یک زبان ابهام بین اسم و صفت وجود داشته باشد، ولی ابهام بین فعل و اسم نباشد.

بعد از تمام این تقسیم‌بندی‌ها و توصیفات، می‌بایست این بانک داده‌ها به زبان رایانه درآید، زبان ماشین و رایانه عدد است، بنابراین همه آن‌ها در لغت‌نامه‌های متنی و دستوری زبان مبدأ، به اعداد تبدیل می‌شوند و در مجموعه داده‌های زبان مقصد هم همین روال وجود دارد. آدرس و ارتباط با اعدادی است که نقش رابط را بین اعداد کدگذاری شده ایفا می‌کنند. در این باره به طور مفصل‌تر در مبحث «برنامه‌نویسی» این مقاله توضیح داده می‌شود.

۲- ریخت‌شناسی و نحو در نظام ترجمه ماشینی

دستور در نظام ترجمه ماشینی شامل صرف، نحو و معناشناسی است. در دستور زبان غیرماشینی برخی روابط زبر زنجیری نیز مورد مطالعه و تعریف قرار می‌گیرند، اما از آنجایی که تا نسل سوم نظام‌های ترجمه ماشینی، منبع ورودی متن کتبی است و نه گفتار، بنابراین به مقولات زبر زنجیری مانند لحن، آهنگ، ریتم و غیره توجهی نمی‌شود.

در دستور زبان ماشینی لازم است نشانه‌های صرفی، که شامل پایانه‌های صرفی و پیشوندها است، تعریف و کدگذاری شوند. همچنین لازم است ویژگی‌های تمام اجزاء کلام، که شامل اعداد، زمان، گونه فعل، صرف افعال و اسامی و صفات است، به شکل مناسب برای ساخت الگوریتم تنظیم و ارائه شوند. مجموع چرخه تجزیه و تحلیل در هر نظام متفاوت است. در نظام ترجمه ماشینی انگلیسی به روسی آمپار این مراحل به یازده فقره مستقل بالغ می‌شوند، که این مجموعه نظام پارسینگ آن را تشکیل می‌دهند.

1. Первичная обработка	۱- مرحله نخست بازشناسی علائم
2. Предварительный этап	۲- سنتز مقدماتی
3. Словарь оборотов	۳- لغت‌نامه ترکیبات
4. Анализ глагольных конструкций	۴- تجزیه ساختار افعال
5. Разрешение омографии	۵- رفع ابهام آموگراف‌ها
6. Грамматический анализ до перевода	۶- تحلیل دستوری قبل از ترجمه
7. Перевод однозначных слов	۷- ترجمه کلمات تک معنا
8. Перевод многозначных слов	۸- ترجمه کلمات چند معنا
9. Контрольный перевод многозначных слов	۹- کنترل ترجمه چند معنا
10. Грамматический анализ после перевода	۱۰- تحلیل بعد از ترجمه
11. Синтез русского текста	۱۱- سنتز متن

ترتیب و مراحل نظام پارسینگ در نظام ترجمه ماشینی انگلیسی به روسی آمپار

در لغت‌نامه گرامری نظام ترجمه ماشینی، ثبت ریشه کلمات با روال معمولی مقداری متفاوت است، ریشه کلمه طوری انتخاب می‌شود که خود آن دچار تغییر نگردد و فقط با ارائه

دقت آن کامل خواهد بود.

در نظام ترجمه ماشینی لغت‌نامه‌های دستوری دیگری نیز مفروض است، از مهم‌ترین آن‌ها می‌توان از لغت‌نامه ریشه کلمات، پسوندها، پیشوندها و در مجموع همه حالت‌های احتمالی نام برد که شامل لغت‌نامه حالت در اسامی و لغت‌نامه جهت در افعال می‌باشد. اما استفاده اصلی و اولیه از لغت‌نامه بن واژه‌ها است.

در کنار لغت‌نامه‌های دستوری، در نظام ترجمه ماشینی، جداولی نیز وجود دارند. این جداول‌ها به تسهیل در کار نظام ترجمه کمک می‌کنند. از مهم‌ترین آن‌ها، جداول‌های واژه‌سازی است. این جداول‌ها به خصوص از آن جنبه نیز اهمیت دارند که گاهی یک کلمه دارای دو بن است و طبق تقریر از پیش آماده شده، تعیین شده که چه بنی و یا چه پایانه‌ای، چه نقشی را در واژه‌سازی بر عهده دارند.

در دسته‌بندی کلمات، برای هر کلمه یک جدول حالات آن وجود دارد، اما برای افعال دو جدول؛ به خاطر آن که گونه فعل در روسی پدیده ویژه‌ای است. در دسته‌بندی گونه افعال، افعال از این حیث که دارای یک گونه و یا دو گونه‌اند، نیز متمایز می‌شوند. در نوع دوم، خود این دسته افعال باز به گروه‌های دیگر تقسیم می‌شوند:

افعالی که گونه مطلق یا استمراری آن‌ها با پیشوند ساخته می‌شوند، مانند:

писать → написать

افعالی که گونه مطلق یا استمراری آن‌ها با پسوند ساخته می‌شوند، مانند:

кричать → крикнуть

و یا افعالی که دارای اشکال کاملاً متفاوتند، مانند:

говорить → сказать

افعالی که تقابل گونه آن‌ها به وسیله پسوند و یا کلمه دیگری است، در لغت‌نامه دارای دو بن خواهند بود. افعالی که تقابل گونه آن‌ها به وسیله پیشوند است، بن مربوط به آن‌ها از افعال غیر کامل (HCB) انتخاب و درج می‌گردد.

۳- معنا شناسی در نظام ترجمه ماشینی

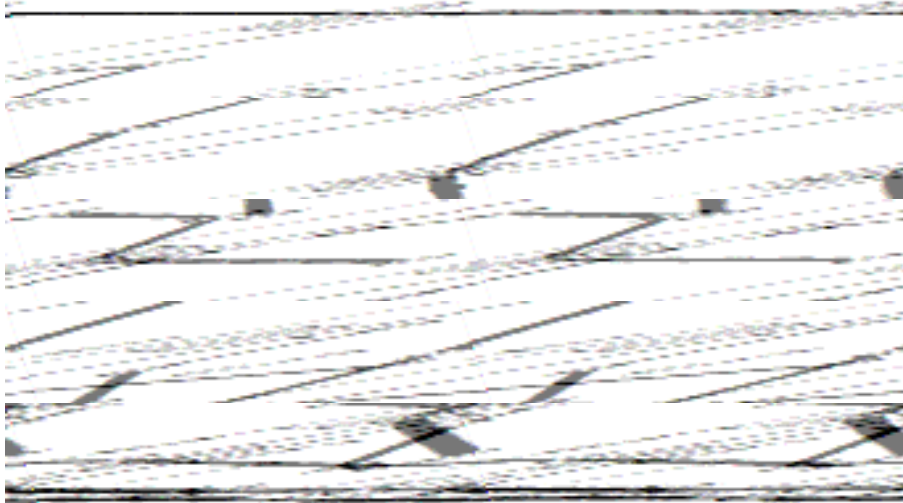
در هنگام ترجمه لازم است چند معنایی دستوری و لغوی حل و ابهام آن برطرف شود. چند معنایی لغوی از نظر فراوانی آن دارای اهمیت فراوانی است. برای حل این مشکل از نظریه‌های دترمینانت (determinant) استفاده می‌شود. این نظریه‌ها شامل توضیحات و

توصیفاتی‌اند که معنا و مفهوم خاصی را تحت شرایط معینی به معانی دیگر مقدم داشته و قرینه مفهوم مشخصی می‌شوند. لازم به یادآوری است که این مشکل در هنگام ترجمه انسانی هم روی می‌دهد و مترجم گاه در می‌ماند که کلمه مربوطه چند معنا، براساس چه پدیده‌ای ایجاد ابهام می‌کند. به عقیده زبان‌شناس مشهور روسی "چند معنایی ویژگی بیشتر کلمات عادی همه زبان‌ها است" (رفورماتسکی ۱۹۹۷، ۸۱).

موضوع مشکل کلمات چند معنا، از همان پیدایش اولیه ترجمه ماشینی خود را نشان داد و باعث بروز مشکل در روند ترجمه شد، حتی برخی زبان‌شناسان مانند ت. وینوگراد این مشکل را قابل حل ندانسته و نقطه ضعف بزرگ ترجمه ماشینی تلقی کردند^۱. با این حال زبان‌شناسان علاقه‌مند به ادامه کار در طرح‌های ترجمه ماشینی، به زودی فهمیدند این مشکل از طریق خود متنی که کلمات در آن قرار می‌گیرند، قابل حل است. اما خود متن و درک از آن، تلقی واحدی نداشت، در مجموع دو روش اصلی برای حل محاسباتی (الگوریتمی) این مشکل ارائه شد. سرانجام این‌گونه بود که راه‌حلی نحوی را ارائه داد، طبق نظریه وی "می‌توان با لحاظ کردن نقش‌های نحوی کلمات چند معنا، این مشکل را رفع کرد" (۲۶) این روش بعد از آن طراحی شد که روش توصیفی (Дистрибутивный)، قبل از این‌گونه توسط زبان‌شناس دیگری به نام هاریس ابداع شده بود، در عمل ناکارآمد بود و موضوع مبنا قرار گرفتن معنا در آن، با ابهام روبه‌رو می‌شد. در این روش با در نظر گرفتن نقش نحوی کلمات، دیگر ضرورتی ندارد که نقش کلمه در ارتباط با کل جمله در نظر گرفته شود، زیرا می‌توان نقش نحوی کلمه مورد نظر را از اجزاء کوچک‌تر جمله نیز به دست آورد. در مرکز تحقیقات انفورماتیک کمپانی لاک‌هید آمریکا، از همین روش برای رفع ابهام چند معنایی کلمات برای زبان انگلیسی استفاده می‌شود. برای مثال، کلمه elevation دارای معنی ارتفاع و ارتقا است، این معانی در داخل جمله کاملاً از یکدیگر متمایزند. اما اگر نقش نحوی کلمه مورد نظر را در ارتباط با هر یک از معانی در نظر بگیریم، الگوی مجزایی وجود دارد و دقیقاً هر الگو بیانگر معنی معینی از کلمه چند معنایی مورد نظر است. بنابراین می‌توان به یک فرمول‌بندی معین دست یافت و آن این‌که "هر معنای پنهان در کلمات چند معنا، دارای الگوی نحوی خاصی است، بنابراین هر کلمه چند

۱- وینوگراد با ارائه جملاتی مانند *the fish was bought by the cook, the fish was bought by the river* عدم درک ماشین را در چنین موقعیت‌هایی متذکر می‌شود، در صورتی که می‌توان با نگارش لگاریتم مناسب به نظام ترجمه ماشینی، دستور داد کجا کلمه *by* را قید و کجا نشانه مجهول را به کار برد. نظام‌های رده بالای ترجمه ماشینی مانند آمپار و یا سیستران در چنین مواردی دچار خطا نمی‌شوند.

معنا، دارای چند الگوی نحوی است“ (رتسکر ۲۷۳).
 بنابر آمارهای استخراج شده از لغت‌نامه‌های مختلف، پراکندگی درصد تعداد کلمات چند
 معنا در اجزای مختلف گفتار، یکسان نیست. برای نمونه، براساس محاسبه انجام شده توسط
 مارچوک (۱۹۸۳، ۱۴۳)، اسم و قید کمتر دچار ابهام در چند معنایی می‌شوند، ولی فعل بیشتر
 از همه دچار این ابهام می‌شود. به درصد استخراجی مربوط به اجزای کلام در نظام ترجمه
 ماشینی آمپار توجه کنید.
 فعل - ۳۲٪، اسم - ۴۰٪، صفت - ۲۲٪، قید - ۳٪، حرف اضافه و ربط - ۱/۵٪.



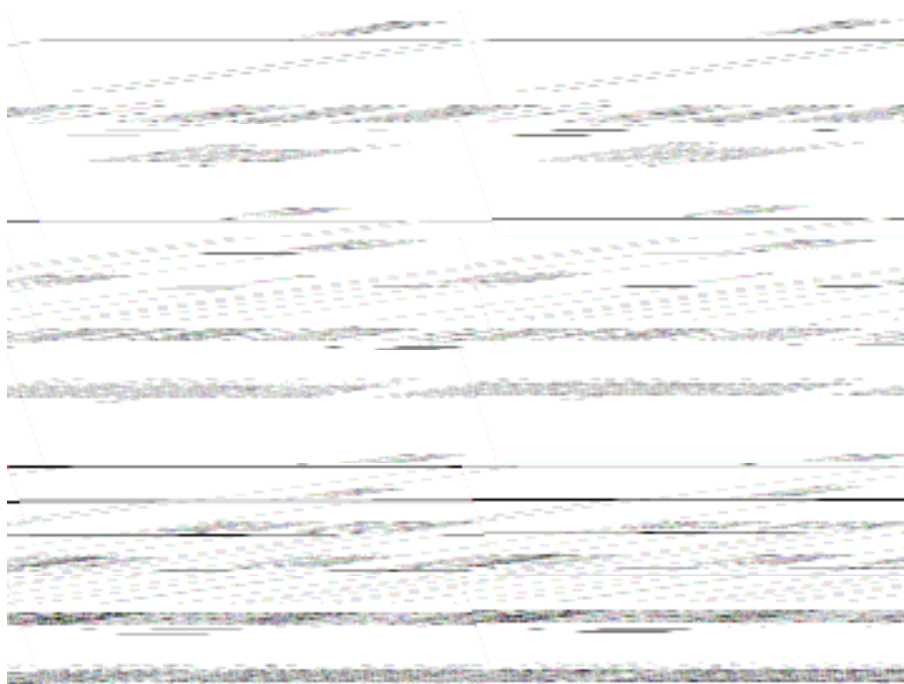
نمودار پراکندگی اجزای کلام در لغت‌نامه بسامدی نظام ترجمه ماشینی آمپار: ۱- اسم ۲- فعل
 ۳- صفت ۴- آموگراف‌ها ۵- قید ۶- حروف اضافه ۷- حروف ربط ۸- علائم نقطه‌گذاری.

این درصدها خام است و باید پردازش شود. نمودار وزنی این موضوع را بهتر نشان
 می‌دهد، و طبق آن درصد ضریب ابهام در چند معنایی فعل ۵۵٪ در برابر ۱۸٪ برای اسم است.
 تحلیل‌گر نحوی آمپار^۱ در مرکز ملی ترجمه متون علمی - ادبی آکادمی علوم شوروی^۲

۱- از این تحلیل‌گر نحوی در حال حاضر در نظام‌های ترجمه ماشینی، فرانسوی به روسی (фрап) و آلمانی به
 روسی (нерап) استفاده می‌شود.

۲- ВЦП - Всесоюзный центр перево дов научно-технической литературы и документации
 ГКНТ и АН СССР

(ВЦП) طراحی شد، از این تحلیل‌گر برای ساخت نظام‌های ترجمه ماشینی استفاده فراوانی شده است. در تهیه لگاریتم‌های این تجزیه‌گر استفاده زیادی از دستور گشتاری و ژرف‌ساخت‌های ارائه شده در آن، برای ساخت درخت‌های متناظر ترجمه نشده است. در واقع نظام طوری طراحی شده است، که عملکرد آن در جملات ساده چندان به کار گرفته نمی‌شود و در حد تغییرات صرفی الگوی متناظر، برای ترجمه ساخته می‌شود، ولی در جملات مرکب، از گشتارهای متناظر آن استفاده می‌شود. مثال‌های زیر و ترجمه انجام شده توسط آن، بیانگر این موضوع‌اند.



تحلیل‌گر نحوی در این‌گونه نظام‌ها، معمولاً در مرحله معنانشناسی (Семантика)، جملات متن را در چهار مرحله پردازش می‌کند؛ در مرحله اول احتمال ابهام در کلمات مشابه مشخص می‌شود، کلماتی مانند мукá, мýка و замóк, зámок که دارای تلفظ مشابه نیستند، اما دیکته آن‌ها یکی است و کلماتی که هم دیکته و هم تلفظ یکسانی دارند، مانند лук شناسایی و پس از مشخص شدن معنا در جایگاه نحوی خود قرار می‌گیرند.

مرحله دوم از سمت راست به چپ، هر یک از کلمات جمله را شروع کرده و احتمال بالقوه روابط نحوی را براساس کلمات موجود ارزیابی می‌کند. روابط کلمات در جمله از نظر نوع ارتباط نحوی و ساختارهایی که می‌توانند داشته باشند، مورد بررسی قرار می‌گیرند، کلماتی مانند فعل + مفعول، حرف اضافه + مفعول و غیره با جداول و منابع دسته‌بندی شده و فرموله شده در ماشین مقایسه می‌شوند.

در مرحله سوم تحلیل خود را از چپ به راست جملات شروع می‌کند و دوباره همان روابط را مورد کنترل قرار می‌دهد. برای مثال؛ معنی فعل مرکب put طبق الگوریتم از قبل ساخته شده در لغت‌نامه گرامری نظام، به معادل روسی درست آن تبدیل می‌شود.

- 1) 'примириться': P+-put+-up+-with+C. He never put up with injustice.
- 2) 'положить': P+put+up+C1+on+C2. She put up the book on the shelf.
- 3) 'отложить': P+put+off+C. They put off their visit for a week.

در مرحله چهارم با استفاده از داده‌های به دست آمده از مراحل قبلی، و با مشخص شدن مرز جملات، نوع جملات اصلی و تبعی، مبتدا و خبر یا نهاد و گزاره را تعیین می‌کند. در نظام آمپار صداها لگاریتم درباره نحوه ترکیب کلمات با یکدیگر و ترکیبات متناظر آن در زبان مقصد تعبیه شده است. و در پایان این مرحله، بن مورد نظر کلمه با پیشوند و یا پسوند یا پایانه در زبان مقصد ترکیب می‌شود و جزئیات نحوی دیگر مانند قرار گرفتن قید در مکان خاص اعمال می‌شود.

ج) برنامه نویسی

در فعالیت‌های مربوطه به داده‌پردازی و ترجمه ماشینی، لازم نیست زبان‌شناس، برنامه‌نویس هم باشد. اما دانستن اصول اولیه برنامه‌نویسی، ارتباط زبان‌شناس و برنامه‌نویس را تقویت کرده و کار را تسهیل می‌کند. فرض کنیم کسی می‌خواهد خانه بسازد، او اگر فقط قصد خود را بگوید و اعلام نکند که خانه وی چه مشخصاتی باید داشته باشد، مهندس و معمار ساختمان یا به سلیقه خود عمل می‌کنند، که این سلیقه همیشه مورد پسند واقع نمی‌شود و یا مجبور خواهند بود سؤال کنند تا موضوع روشن‌تر تداعی و تحقق یابد. با این حال برنامه‌نویسی کار بسیار پیچیده‌ای نیست و اگر زبان‌شناس وقت و علاقه نشان دهد زود با اصول آن آشنا خواهد شد. البته هر روز که می‌گذرد برنامه‌نویسی نیز دچار تحول شده و ساده‌تر می‌شود و

قطعات برنامه‌ها، مانند قطعات خانه‌های پیش‌ساخته، آماده قرار دادن در محل مورد نظر می‌شوند. آنچه بیش از برنامه‌نویسی و قطعاً پیش از آن برای زبان‌شناس اهمیت دارد، داشتن اطلاعاتی کاربردی از آمار، منطق و تا حدی ریاضیات گسسته است.^۱ یوری نیکلای‌یوویچ مارچوک (Ю. Н. Марчук) یک زبان‌شناس است، اما به گفته خودش، چون مسئولیت گروه تحقیقاتی ترجمه ماشینی «آمپار» را بر عهده داشت، با علاقه‌مندی مجبور شد تا حد ضرورت، علوم مرتبط در این تحقیق را یاد بگیرد، علومی مانند تبدیل علائم زبان‌های طبیعی به علائم رایانه‌ای، نظریه‌های ترجمه و متن و برنامه‌نویسی.

واضح است که برنامه‌نویس از شرط‌های فراوانی است که با کلمه پر بسامد *if*، تکلیف انتخاب را مشخص می‌کند. البته آنچنان که گفته شد، روش‌های برنامه‌نویسی می‌توانند متعدد باشند و اگر تعداد نمودارهای درختی محدود و دارای تناظر منطقی باشند از این روش هم می‌توان برای تعیین ساختار کامل جملات استفاده کرد. زبان اصلی هر رایانه، زبان ماشین است. هر یک از دستورهای زبان ماشین، رشته‌ای دودویی از ۰ و ۱ است که یک عمل و صفر مکان خانه‌های حافظه مربوط به آن عمل را مشخص می‌کند. برای مثال، اگر خواسته باشیم فرمول جبری زیر را که محاسبه قیمت (Cost) یک کالا است به یک برنامه زبان ماشین بنویسیم:

$$\text{Cost} = \text{Price} + \text{Tax};$$

با این فرض که قیمت (Price) و مالیات (Tax) یعنی داده‌های فرمول بالا، قبلاً در خانه‌های حافظه ذخیره شده‌اند، لازم است عملیات زیر را انجام دهیم:

- ۱- قیمت کالا (Price) را از حافظه به دست آوریم.
- ۲- مالیات (Tax) را به آن اضافه کنیم.
- ۳- حاصل جمع را در خانه حافظه دیگری ذخیره کنیم.

سه دستور زیر چه‌گونگی ظاهر شدن عملیات بالا را در یک زبان ماشین فرضی نشان می‌دهد:

0010 0000 0000 0011

0100 0000 0000 0101

0011 0000 0000 0110

در هر یک از دستورهای زبان ماشین بالا، چهار رقم سمت چپ، عمل مورد نظر را نشان

۱- برای تبدیل علائم زبان‌های طبیعی به علائم رایانه‌ای از ریاضیات گسسته استفاده می‌شود.

می‌دهند. این قسمت از دستور، کد عملیاتی نام دارد. ارقام باقیمانده، آدرسی را در حافظه اصلی نشان می‌دهند. اولین کد عملیاتی یعنی 0010، به رایانه می‌گوید داده را از خانه حافظه 0011 بازیابی کند و آن را در یک خانه حافظه خاصی ذخیره کند. کد عملیاتی 0100 به رایانه فرمان می‌دهد محتوای خانه حافظه 0101 را بازیابی کند و آن را با محتوای خانه حافظه خاصی جمع کند و حاصل جمع را در آن خانه حافظه خاص ذخیره کند و آن را روی محتوای قبلی حافظه بنویسد. کد عملیاتی 0011 به رایانه می‌گوید محتوای جدید آن خانه حافظه خاص را در آدرس 0110 ذخیره کند. توجه دارید که هر دستور حاوی یک عمل و یک خانه حافظه (آدرس) است. این نقش دوگانه عمل و داده‌ها، ویژگی عمومی زبان‌های برنامه‌نویسی است. اگرچه درک سه دستور زبان ماشین بالا، برای رایانه کاری بسیار ساده است، با این وجود درک و فهم آن برای اکثر مردم، حتی برنامه‌نویس‌های حرفه‌ای و قوی بسیار مشکل است. بنابر این خیلی زود این باور ایجاد و تقویت شد که برای تسهیل کار برنامه‌نویسی از برنامه‌های رابط بین انسان و ماشین استفاده شود. این زبان‌ها را زبان‌های سطح بالا نام نهاده‌اند. هنگام نوشتن برنامه‌ها به یک زبان سطح بالا، از دستورهایی استفاده می‌شود، که به زبان روزمره انگلیسی بسیار شبیه‌اند. در زبان جاوا، از دستورهایی مانند دستور زیر استفاده می‌شود:

$$\text{Cost} = \text{Price} + \text{Tax};$$

که شباهت بسیار زیادی با فرمول اصلی دارد. با وجود این، معنی این دستور، معادل یک فرمول جبری نیست. در زبان جاوا، این دستور دارای مفهوم بسیار خاص زیر می‌باشد:

«مقدار Price را با مقدار Tax جمع کنید و حاصل جمع را با نام Cost در حافظه ذخیره کنید». هنگام نوشتن برنامه به یک زبان سطح بالا، با استفاده از نام‌های توصیفی مانند Price، Cost و Tax به جای آدرس عددی حافظه، می‌توانیم به داده‌های ذخیره شده در آن خانه حافظه دسترسی پیدا کنیم. علاوه بر این، برای بیان عملیاتی که می‌خواهیم انجام دهیم از علامت‌های آشنا (نظیر +) استفاده می‌شود. هر محاسبه، چه به زبان ماشین و چه به یک زبان سطح بالا بیان شود، معمولاً شامل خواندن داده، انجام یک عمل روی آن داده، و آنگاه انجام کارهایی روی نتیجه است.

هر زبان سطح بالا دارای یک استاندارد زبان است که ساختار دستوری آن زبان را شرح می‌دهد. هر یک از دستورهای زبان سطح باید با قوانین دستوری‌ای سازگار باشد که در استاندارد زبان مشخص شده است. این قوانین، بسیار دقیق‌اند، یعنی برای دستورهایی که به صورت تقریبی درست‌اند جایی وجود ندارد. برنامه‌هایی که با این قوانین سازگارند دارای

قابلیت حمل‌اند. منظور از قابلیت حمل این است که بدون ایجاد هیچ تغییری، بتوان آن‌ها را روی هر نوع رایانه دیگر اجرا کرد. از سوی دیگر، یک برنامه زبان ماشین را تنها می‌توان روی یک نوع رایانه خاص مورد استفاده قرار داد، زیرا برای خارج شدن از برنامه، وابسته به سخت‌افزار CPU آن رایانه خاص‌ایم. زبان‌های برنامه‌نویسی C، C++، بیسیک، کوبول، فورترن، جاوا، لیسپ و پاسکال و دلفی از جمله زبان‌های سطح بالای متداول‌اند (دیتل ۱۱). اگرچه هر یک از این زبان‌ها برای اهداف خاصی طراحی شده‌اند، با این وجود از تمام این زبان‌ها برای نوشتن نرم‌افزارهای کاربردی مختلف استفاده می‌شود. در ترجمه ماشینی نیز فقط از یک نرم‌افزار برنامه‌نویسی استفاده نمی‌شود، و با توجه به ساز و کار مورد نظر، از کارایی نرم‌افزارهایی که امکان بهره‌مندی دارند، استفاده می‌شود. در ترجمه ماشینی جمله زیر که از انگلیسی به روسی انجام شده، تبدیل واژه‌ها به اعداد و سپس استخراج واژه مناسب با آدرس مشخص شده به چشم می‌خورد.^۱

Throughout one nine six six, the Greek working class maintained with honor its vanguard role in developing the struggles of the people.



همان‌طور که ملاحظه می‌شود، تحلیل‌گر نحوی در مرحله آنالیز (تجزیه) در خط فوقانی خود عدد ۲۳ را ثبت کرده است، بنابراین در مرحله سنتز (تحلیل)، به همان تعداد آدرس مورد جستجو قرار می‌گیرد.

26 01 17 00000000001010 02 23 23	Throughout = BCE ВРЕМЯ
20 00 21 00000010010110 00 01 09	One =1
00 02 12 00000010010111 14 21 23	Nine =1
00 02 20 00000010010111 17 11 18	Six =6
00 02 20 00000010011101 17 11 18	Six =6

۱- در نماد ارائه شده، به غیر از اعدادی که در مرکز قرار دارند، اعداد دوتایی در راست و چپ توصیف‌گر نشانه‌های دستوری و آدرس‌های جستجو و وصل لگاریتم‌هایند. (این نمونه پارسینگ برگرفته از کتاب ترجمه رایانه‌ای پروفیسور مارچوک است)

19 24 00 00000000000000 00 03 06	,	= ,
12 00 02 00000000000110 00 00 01	The	=.
12 02 05 00000010100110 05 08 01	Greek	= ГРЕЧЕСКОГО
24 15 05 00000010100110 07 28 31	Working	= РАБОЧЕГО
22 04 21 00000010101101 15 04 30	Class	= КЛАССА
26 26 26 01010010100111 01 13 24	Maintain	= ПОДДЕРЖИВАЛАСЬ
26 01 26 00000000001011 11 29 00	With	= С
22 15 29 0000010001110127 18 24	Honor	= С ЧЕСТЬЮ
12 00 22 00000100010113 07 15 27	Its	= ЕГО
00 09 04 00000100010110 04 14 00	Vanguard	= АВАНГАРДНАЯ
10 14 30 00000100010011 27 03 01	Role	= РОЛЬ
26 01 14 00000000001011 11 20 00	In	= В
11 30 30 00000111101102 06 21 26	Developing	= РАЗВИТИИ
12 00 02 00000000000110 00 00 01	The	=.
10 20 15 0000010110 10 25 06 16	Struggles	= БОРЬБЕ
26 00 00 00000000001010 00 00 01	Of	=.
12 00 02 00000000000110 00 00 01	The	=.
22 24 11 00000010101101 15 21 08	People	= НАРОДА

19 24 16 00000000010000 00 03 26 .=.
 ВСЕ ВРЕМЯ 1966, ГРЕЧЕСКОГО РАБОЧЕГО КЛАССА ПОДДЕРЖИВАЛАСЬ С
 ЧЕСТЬЮ ЕГО АВАНГАРДНАЯ РОЛЬ В РАЗВИТИИ БОРЬБЕ НАРОДА.



دربارهٔ تبدیل کلمات به اعداد، توجه داشته باشید، که برخی کلمات، معادل واژگانی

ندارند، کدگذاری به گونه‌ای انجام می‌گیرد، که هیچ کلمه‌ای انتخاب و جایگزین نمی‌شود. برای نمونه در این جمله مرکب کلمه of و حرف تعریف the دارای معادل واژگانی تھی است، اما نقطه و ویرگول نشانه معادل خود را دارند.

ترجمه ماشینی، یکی از نتایج فعالیت‌های کاربردی پردازش زبان است که ممکن است به روش‌های مختلفی انجام گیرد، با این حال روش تجزیه و انتقال و ترکیب و تحلیل داده‌ها در اکثر روش‌ها کم و بیش مشابه است. آنچنان که در این مقاله به آن اشاره شد، آنقدر که در عمل، ایده‌های معناشناختی ملچوک به ترجمه ماشینی کمک کرده، ایده دستور همگانی و جهانی چامسکی کاربرد چندانی نداشته است. با این حال اگر قرار باشد در نظام پارسینگ ترجمه ماشینی، ایجاد درخت‌های متناظر لازم و در عملکرد نظام قرار گیرد، پارسینگ نظام تا حدی با روش بکار گرفته در نظام‌های روسی متفاوت خواهد بود. پارسینگ نظام ترجمه ماشینی با لحاظ ایجاد درخت‌های متناظر را می‌توان به صورت خلاصه طبق نمودار زیر بیان نمود:



نتیجه‌گیری

بُعد زبان‌شناختی نظام ترجمه ماشینی شامل تحقیق و بررسی درباره مقولات واژگان، صرف - نحو، و معناشناسی است. مؤلفه‌های سه‌گانه اشاره شده در تناظر با سه سطح که علم زبان‌شناسی نوین برای هر زبان طبیعی و زنده قائل است، قرار دارند. نحوه بررسی و ارائه دستاوردهای پژوهش در این ابعاد با روش‌های سنتی و آموزشی مرسوم متفاوت است. ما در این مقاله سعی کردیم دسته‌بندی و حدود و ثغور این مقولات را در ارتباط با ترجمه ماشینی ارائه نماییم. با توجه به مطالب ارائه شده در این مقاله، می‌توان چنین نتیجه گرفت، که در ساخت ترجمه ماشینی، محدودیتی در استفاده از روش‌ها و نظریه‌های زبانی وجود ندارد. برای

ساخت لغت‌نامه‌های متنی و دستوری و تحلیل‌گر نحوی از دستاوردهای دستور زبان تطبیقی (Comparative grammar)، مقابله‌ای (Contrastive grammar)، احتمالی (Expectance grammar)، محوری (Pivot grammar)، ساختاری (Structural grammar)، ساخت‌گروهی (Phrase structure grammar)، گشتاری (ransformational - Generative grammar) و هر روش و نظریه جدیدی که مناسب ساز و کار آن باشد، استفاده شده است. برای نمونه، نحوه مشخص کردن بن کلمه، و یا ریشه فعل در نظام ترجمه ماشینی با تعاریف سنتی و آموزشی رایج در رابطه با زبان روسی متفاوت است، از سوی دیگر در نظام ترجمه ماشینی موضوع کلمات چند معنا را می‌توان با لغت‌نامه‌های متنی حل کرد، اما همیشه این روش نتیجه‌بخش نیست. مواردی وجود دارد که باید رویه دیگری برگزید. همه این مطالب دال بر آن است که میدان پژوهش و تحقیق در این حیطه گسترده است و نیاز به توجه در ایجاد زیر بنای لازم در ابعاد مختلف آن، از ضروریات زبان‌شناسی معاصر ایران است.

کتابشناسی

- باطنی، محمدرضا. (۱۳۵۴). *مسائل زبان‌شناسی نوین*. تهران: نشر آگاه.
- دبیر مقدم، محمد. (۱۳۸۳). *زبان‌شناسی نظری*. تهران: نشر سمت.
- دیتل، هاروی. (۱۳۸۵). *راهنمای جامع برنامه‌نویسان*. مترجم، بهرام پاشایی. چاپ یازدهم. تهران: نشر جهان نو.
- راسل، جی. اس. (۱۳۸۳). *هوش مصنوعی، رهیافتی مدرن*. مترجم، رامین رهنمون و آناهیتا هماوندی. چاپ سوم. تهران: نشر ناقوس.
- خداپرستی، فرج الله. (۱۳۶۵). *ابداع نخستین تجزیه و تحلیل‌گر نحوی*. شیراز: نشر دانشگاه شیراز.
- ساغرونیان، سید جلیل. (۱۳۶۹). *فرهنگ اصطلاحات زبان‌شناسی*. مشهد: نشر نما.
- Апресян, Ю. Д. (1987). *Лингвистическое обеспечение в системе автоматического перевода третьего поколения*. АН СССР, «Кибернетика». М.
- Ингве, В. (1965). *Значение исследований в области машинного перевода*. М.: ИТИ.
- Марчук, Ю.Н. (1983). *Проблемы машинного перевода*. М.: Наука.
- . (1976). *Вычислительная лексикография*. М.: ВЦП.

- . (1980). *Синтактико-семантический анализ в системе машинного перевода АМПАР*. Международный форум по информации и документации. М.: ВИНТИ.
- Мельчук, И.А. (1989). *К построению действующей модели языка "Смысл – Текст"*. М.
- Попов, Э.В. (1990). *искусственный интеллект книга 1 радио и связь*. М.
- Рецкер, Я.И. (1979). *Введение в общую филологию*. М.
- Фреге, Г.В. (1983). *Вычислительные машины*. М.: Прогресс.
- Winograd, T. (1977). *Five Lectures on Artificial Intelligenc., Linguistic Structure Processing*. New York: North-Holland Pub. and Co. Amsterdam and Oxford. V. 5. 491-572.