

ابهام در ماشین ترجمه

محمد رضا فلاحتی فومنی^۱

چکیده

هدف از انجام این تحقیق، بررسی گونه‌های مهم ابهام واژگانی و ساختاری و نیز مشکلاتی است که این گونه ابهامها را در امر ترجمه ماشینی ایجاد می‌نمایند. برای انجام این کار پنج نوع مهم ابهام واژگانی با عناوین ابهامهای مقوله‌ای، واژه‌های هم‌آوا - هم‌نویسه، واژه‌های هم‌نویسه، چند معنایی و ابهام انتقالی، با ارائه مثالهایی از فارسی و انگلیسی مورد بررسی قرار گرفت. سپس ابهام ساختاری بررسی شد و دو نوع مهم آن با عناوین ابهامهای ساختاری واقعی و سیستمی تشریح گردید. پس از آن نیز شش روش رایج برای رفع ابهام ساختاری و نیز مشکلاتی که مرجع‌یابی در امر ترجمه ایجاد می‌نماید، بررسی شد. بررسی‌ها نشان داد یکی از موارد ضروری برای عملکرد مثبت ماشینهای ترجمه، پرداختن اصولی به ابهامهای موجود در متن است و اینکه در بسیاری از موارد اطلاعات مربوط به واژه‌ها و حتی جمله (که اغلب به عنوان واحد ترجمه محسوب می‌گردد) برای ابهام‌زدایی کافی نیست و باید از اطلاعات غیر زبانی مانند اطلاعات جهان پیرامون و ... در ترجمه استفاده شود که البته در حال حاضر ماشینهای ترجمه عموماً فاقد این قابلیت مهم هستند. کلیدواژه‌ها: ماشین ترجمه، ابهام، ابهام‌زدایی، ابهام واژگانی، ابهام ساختاری، مرجع‌یابی، ابهام واقعی، ابهام سیستمی.

۱. مقدمه

زبانی که در زندگی روزانه برای ایجاد ارتباط با دیگران به کار می‌بریم، پیچیدگیهای فراوانی دارد. آنچه ابهامهای موجود را بین سخنگویان بومی مرتفع می‌سازد،

۱. عضو هیئت علمی گروه پژوهشی زبان‌شناسی رایانه‌ای کتابخانه منطقه‌ای علوم و تکنولوژی Osic@srlist.com

توانش زبانی آنها، اطلاعات آنها در خصوص جهان پیرامون، طرح پرسش مجدد در صورت وجود یا احساس ابهام و به طور کلی مجموعه‌ای از اطلاعات زبانی و غیر زبانی است که سخنگویان بومی به آن مجهزند. مسلماً در مکالمات بین سخنگویان بومی و افرادی که آن زبان را به عنوان زبان دوم یا خارجی یاد می‌گیرند، موارد ابهام‌زای بیشتری به وجود می‌آید. این نکته نشان می‌دهد استفاده درست از زبان به اطلاعات متنوعی نیاز دارد که متأسفانه تاکنون نتوانسته‌ایم دست کم برخی از آنها (مانند اطلاعات مربوط به جهان پیرامون، قدرت حل مسئله، توانایی پرسش مجدد، استفاده از اطلاعات متنی، بافتی و ...) را به ماشین ترجمه بدهیم. این امر خود سبب کاهش عملکرد سیستمهای ترجمه می‌شود. امروزه برای بسیاری از زبانها هنوز ماشین ترجمه‌ای ساخته نشده است. حتی برای زبان فارسی نیز چند سیستم یکسویه وجود دارد که از انگلیسی به فارسی ترجمه می‌کنند و قادر به حرکت در مسیر عکس (از فارسی به انگلیسی) نیستند. مسلماً نمی‌توان انتظار داشت سیستمهای مناسب در زمانی کوتاه پدید آیند. با این همه، برای دستیابی به سیستمهای ترجمه مناسب نخست باید تحقیقهای زبان شناختی جامعی در حیطه‌های آواشناسی، معناشناسی، نحو، منظورشناسی و ... به انجام برسد و در هر حیطه اجزای مربوط بررسی و به گونه‌ای طرح شود که بتوان از آنها در محیطی رایانه‌ای استفاده کرد.

ابهام از جمله مختصات زبانی است که مشکلات فراوانی را در سیستمهای ترجمه ایجاد می‌کند و سبب نامفهوم شدن یا نامرتب شدن آن می‌شود. رفع ابهام، به مطالعه در حیطه‌های معناشناسی، واژه‌سازی و نحو نیازمند است. امروزه متأسفانه بین زبانشناسی نظری و تحقیقات عملی انجام شده در حیطه ماشین ترجمه شکاف بزرگی وجود دارد. تحقیقات عملی دست کم در ایران اغلب خود را از مباحث مربوط به ماشین ترجمه دور نگه داشته‌اند و تحقیقات نظری نیز نتوانسته‌اند مدلی جامع که به تمامی ابعاد زبان پرداخته باشد، ارائه نمایند.

از سوی دیگر، متأسفانه بسیاری از این نظریه‌ها بر اساس امکانات و ویژگیهای زبان انگلیسی تهیه شده‌اند که علت آن نیز این بوده که زبان مادری زبانشناسی نظری اغلب

انگلیسی بوده است. این نکات ضرورت توجه جدی تر محققان ایرانی را به مباحث نظری و کاربردی زبانشناسی نشان می‌دهد. آنچه به کار ماشین ترجمه و سیستمهای رایانه‌ای می‌آید، زبانشناسی محض نیست؛ بلکه زبانشناسی رایانه‌ای است که متأسفانه هنوز در نظام آموزش عالی کشور جایگاهی ندارد.

۲- تاریخچه تحقیق

در این قسمت، نخست به برخی از مهمترین آثاری که توسط محققان ایرانی و خارجی در خصوص ابهام و موارد مرتبط با آن نوشته شده اشاره خواهد شد. سپس آثاری ارائه می‌گردد که ابهام را در ارتباط با ماشین ترجمه بررسی کرده و بویژه برای ابهام زدایی در سیستمهای ترجمه تدبیری اندیشیده‌اند.

تحقیق در خصوص ابهام، سابقه‌ای طولانی دارد. امپسون (۱۹۳۰) در کتاب جامع خود هفت نوع ابهام را مورد بررسی قرار داد. لاینز (۱۹۷۷) نیز در کتاب معروف خود که با عنوان «معناشناسی» و در دو جلد به چاپ رسید، بحث مفصلی در خصوص ابهام و گونه‌های آن ارائه داد. کتاب معناشناسی سعید (۱۹۹۷) نیز کتابی مناسب برای مطالعه ابهام می‌باشد.

محققان ایرانی نیز در این خصوص آثار خوبی تولید کرده‌اند که از آن جمله می‌توان به کتاب ارزشمند دکتر کوروش صفوی (۱۳۸۰) با عنوان درآمدی بر معناشناسی، کتاب نگاهی تازه به دستور زبان دکتر محمدرضا باطنی (۱۳۷۱) و نیز پایان‌نامه‌های کارشناسی ارشد داوری اردکانی (۱۳۷۵)، معدنی (۱۳۷۶)، صحت (۱۳۷۷) و قنبری (۱۳۷۷) اشاره کرد.

پایان‌نامه علوی (۱۳۷۹) و مقاله‌های کس و هوپ (۱۹۸۵)، دانگ و سایرین (۲۰۰۲)، مگنینی و سایرین (۲۰۰۲) و میهالسی (۲۰۰۲) نمونه‌هایی از مجموعه آثاری است که توسط محققان ایرانی و خارجی در خصوص ابهام نوشته شده و در آنها بخصوص روی روشهای مختلف ابهام‌زدایی به صورت عام و ابهام‌زدایی از مفهوم واژه صحبت شده است. تروزول (۱۹۹۶)، اسپوی و سایرین (۱۹۹۳) و مک دونالد و سایرین (۱۹۹۴) نیز بر روی ابهام نحوی مطالعه کردند و راه‌حلهایی رایانه‌ای را برای ابهام‌زدایی ارائه دادند که بتواند به

عنوان مبنایی برای بهبود عملکرد سیستمهای ترجمه مورد استفاده قرار گیرد. متأسفانه در میان آثار پژوهشگران ایرانی، ابهام اغلب از منظری صرفاً زبانشناختی مورد مطالعه قرار گرفته و در خصوص رفع مشکلات مربوط به آن در ماشین ترجمه و یا ارتباط آن با ماشین ترجمه، کاری صورت نپذیرفته است. مقاله حاضر، بایی است در این خصوص.

۳- انواع ابهام

۳-۱. ابهام واژگانی

این نوع ابهام به طور کلی به پنج دسته تقسیم می‌شود که عبارت است از: ابهام مقوله‌ای، واژه‌های هم‌آوا - هم‌نویسه، چند معنایی و ابهام انتقالی.

ابهام مقوله‌ای وقتی اتفاق می‌افتد که یک واژه در بافتهای مختلف نقشهای نحوی مختلفی چون اسمی، فعلی، صفتی، قیدی و ... را داشته باشد. این نوع ابهام در انگلیسی و فارسی از فراوانی بسیاری برخوردار است و بنابراین برای دستیابی به یک سیستم کارآمد باید به شکلی خاص مورد توجه قرار گیرد. به عنوان مثال، واژه light در انگلیسی می‌تواند اسم، فعل و حتی یک صفت باشد؛ واژه Import نیز می‌تواند نقشهای اسمی و فعلی داشته باشد و ... در فارسی واژه بردار می‌تواند اسم (بردار) و یا فعل امر (بردار) باشد. یکی از تفاوت‌های عمده‌ای که بین این دو زبان وجود دارد، ظهور مصوت کوتاه در واژگان انگلیسی و عدم ظهور این گونه مصوتها در واژگان فارسی است. این گونه ابهامها در بسیاری از موارد از سوی سخنگویان بومی از طریق مراجعه به اطلاعات بافتی قابل رفع است، ولی برای اینکه ماشین ترجمه بتواند این نوع مشکلات را مرتفع سازد باید اطلاعات مربوط در بخش ساختوازی و نحو ماشین ترجمه وارد شود (بخش ساختوازی اطلاعات مربوط به واژه‌ها، ریشه‌یابی، وندها و ... و بخش نحوی اطلاعات مربوط به ساختار نحوی جمله را در خود داراست.

وقتی در جمله تنها یک واژه مبهم وجود داشته باشد رفع ابهام آسان‌تر است، ولی گاه ممکن است بیش از یک کلمه مبهم در جمله وجود داشته باشد که درک جمله را

دشواری می‌سازد. در چنین مواردی، مراجعه به اطلاعات نحوی و بافتی برای رفع ابهام ضروری است، مانند:

(۱). آن مرد سرخورد و مرد.

در این مثال «مرد» اول به «انسان بالغ مذکر» و «مرد» پایانی به «فعل گذشته از مصدر مُردن» برمی‌گردد. ضمناً واژه «سرخورد» نیز دارای چهار نوع ابهام به شرح زیر است:

سَر ← (با فتنه) قسمتی از بدن که بالای گردن قرار می‌گیرد.
 سُر ← (با ضمه) لیز خوردن و از دست دادن تعادل بدن.
 خُورد ← جویدن و قورت دادن غذا (مربوط به سَر)
 خُورد ← جزئی از فعل (سُر خورد) به معنی لیز خوردن.

مشاهده می‌شود که در جمله (۱) در مجموع ۶ مورد ابهام وجود دارد که سخنگویان بومی با استفاده از اطلاعات نحوی و نیز بافتی، جمله را ابهام‌زدایی می‌کنند. البته، این نکته بدان معنا نیست که سخنگویان بومی همواره می‌توانند جمله را درست ابهام‌زدایی کنند. به عنوان مثال، برای جمله (۱) می‌توان تعبیر (۲) را در نظر گرفت. (هرچند تعبیر (۲)، نمونه‌ای بعید و تا حد زیادی دور از انتظار است، نمی‌توان آن را به کل ناممکن خواند):

(۲). آن مرد سر خورد و مرد. (علت مرگ، خوردن سر فردی بوده است)

نکته دیگر در خصوص ابهام مقوله‌ای آن است که تجزیه‌گر ماشین ترجمه، برخلاف سخنگویان بومی صرفاً در خصوص مقوله‌ها و ترکیبهای مجاز مقوله‌ای حاوی اطلاعاتی است و بنابراین موارد ابهام‌زا در ماشین ترجمه بسیار فراوان خواهد بود، بویژه در عناوین خبری موجود در روزنامه‌ها که در آنها واژه‌های دارای نقشهای دستوری اغلب حذف می‌شود و مثلاً به یک عنوان خبری که در قالب جمله (۳) ارائه گردیده است توجه کنید:

(۳) Foot Heads Arms Body

در این مورد خواننده باید در خصوص واقعیت‌های موجود در جهان پیرامون نیز اطلاعات لازم را داشته باشد. او باید بداند که یک سیاستمدار انگلیسی به نام Foot به ریاست کمیته‌ای منصوب شده که مسئولیت آن تحقیق و بررسی در مورد برنامه کنترل تسلیحات است. در جمله بالا، تقریباً تمامی واژه‌ها دارای ابهام‌اند و بویژه دو واژه Heads و Arms ابهام مقوله‌ای دارند چون هر دو می‌توانند فعل سوم شخص مفرد و یا اسم جمع

باشند. یا در فارسی وقتی گزارشگر ورزشی در ورزشگاه آزادی تهران از جوانی جمله (۲) را می‌پرسد:

(۴). قرمز یا آبی؟

سیستم ترجمه به دلیل ناآشنایی با فرهنگ ایران و اطلاعات جهان پیرامون، این جمله را به صورت Red Or Blue? ترجمه خواهد کرد، در صورتی که منظور از قرمز یا آبی تیمهای فوتبال پرسپولیس و استقلال تهران است.

۳-۱- هم‌آوا - هم‌نویسه، هم‌نویسه و چند معنایی

نوع دوم ابهام واژگانی زمانی که یک واژه دارای بیش از یک معنا باشد. زبان‌شناسان معمولاً بین واژه‌های هم‌آوا - هم‌نویسه، هم‌آوا و چند معنایی تمایز قائل می‌شوند (در این بخش، از بحث در خصوص واژه‌های هم‌آوا (مانند خار و خوار) صرف نظر خواهد شد زیرا در مبحث ماشین ترجمه با متن مکتوب سر و کار داریم و نه گفتار شفاهی).

واژه‌های هم‌آوا - هم‌نویسه به واژه‌هایی اطلاق می‌گردد که دارای نگارشی یکسان باشند و به یک شکل هم تلفظ شوند، ولی معانی کاملاً متفاوتی داشته باشند. مثلاً در انگلیسی واژه Light دارای دو معنی روشن و سبک است. در فارسی نیز واژه «سیر» دست کم دارای دو معنای نامرتب است: ۱. نوعی گیاه با طعم و بویی تند و ۲. مقابل گرسنه. یا در جمله‌های (۵) و (۶) واژه‌های «تار» و «چنگ» دارای ابهام‌اند.

(۵). در شب تار تار می‌زند.

(۶). همه دلخوشند که مطرب بزند به تار چنگی

من از آن خوشم که چنگی بزیم به تار مویی

گاه در زبان با واژه‌هایی برخوردار می‌کنیم که دارای املاهای کاملاً یکسان ولی تلفظ متفاوتند، مثل گل (گل و گُل) و یا دیر (دیر و دیر). این گونه واژه‌ها را «واژه‌های هم‌نویسه» می‌نامیم. علت فراوانی این نوع واژه‌ها در زبان فارسی آن است که در این زبان مصوت‌های کوتاه در خط نموده نمی‌شوند، مانند مثال (۷).

(۷). پانصد گرم آتش گرم خریدم.

گاه تغییر جایگاه تکیه نیز سبب تغییر معنای واژه می‌شود.

(۸). ماهی یک ماهی می‌خورم.

(۹). خدا روزی روزی را از تو می‌گیرد.

در مثال ۸ ماهی اول به «یک ماه» و ماهی دوم به «جاندار آبی» اشاره دارد. در

مثال ۹ نیز روزی اول به معنی «یک روز» و روزی دوم به معنی «رزق» است.

سرانجام به واژه‌های چند معنایی می‌رسیم که مجموعه‌ای از واژه‌های دارای صورت و تلفظ یکسان را توصیف می‌کند که بر خلاف واژه‌های هم‌آوا - هم‌نویسه بین معانی آنها نوعی ارتباط وجود دارد، مانند واژه «سر» که در ترکیبات «سر آدم، سر کلاس، سر بطری، سر کوه و ...»

نگاهی به مثالهای بالا نشان می‌دهد که سیستم ترجمه در صورتی می‌تواند به تعبیر و ترجمه‌ای درست از واژگان و در نهایت جمله دست پیدا کند که بتواند تعبیر درست را از طریق بررسی و مراجعه به ساخت نحوی، اطلاعات ساختوازی و اطلاعات بافتی انتخاب نماید. نکته حایز اهمیت دیگر این است که بطور طبیعی بین مقوله‌های بالا (هم‌آوا - هم‌نویسه، چند معنایی و هم‌نویسه) نمی‌توان بطور کامل تمایز قایل شد، چون با گذشت زمان ممکن است واژه‌ای هم‌آوا - هم‌نویسه به واژه‌ای چندمعنایی و ... تبدیل شود. به عنوان مثال، در بیشتر کتابهای زبانشناسی واژه Bank نمونه‌ای از واژه‌های هم‌آوا - هم‌نویسه محسوب گردیده که این به معنای بی‌ارتباط بودن معانی مختلف این واژه است؛ حال آنکه از نگاهی تاریخی و در زمانی، می‌توان بین دو مفهوم این واژه ارتباط پیدا کرد و بنابراین می‌توان این واژه را نمونه‌ای از چند معنایی دانست. علت این امر هم آن است که در سالیان گذشته جویندگان طلا در حاشیه رودخانه اطراق می‌کردند و به جستجوی طلا می‌پرداختند. پس بر اساس اصل بسط معنایی واژه Bank می‌تواند به مرکز مالی نیز اشاره کند.

همان‌گونه که پیشتر ذکر شد، گاه چند نوع ابهام در یک جمله ظاهر می‌شود، به طوری که بدون مراجعه به اطلاعات بافتی متن (بندهای قبل و بعد، صفحات و فصلهای قبل و بعد) حتی سخنگوی بومی نیز قادر به درک معنای موردنظر نویسنده نخواهد بود. ضمناً گاه یک واژه ابهام‌زا در جمله وجود دارد، ولی واژگان دیگر موجود در جمله به شکلی است که همان واژه ابهام‌زای منفرد نیز حتی برای سخنگویان بومی با تجربه، ایجاد مشکل می‌کند، مانند مثالهای (۱۰) و (۱۱) زیر:

(۱۰). او آن داستان را قشنگ خواند.

تعبیر اول: او آن داستان را به شکلی زیبا قرائت کرد.

تعبیر دوم: او آن داستان را زیبا ارزیابی / قلمداد کرد.

(۱۱). این کارت زشت است.

تعبیر اول: این کار تو زشت است.

تعبیر دوم: این کارت (برگه تبریک، عروسی، مشخصات و ...) زشت است. در حوزه معناشناسی پاره‌ای از ابهامها از طریق مراجعه به مشخصه‌ها و مؤلفه‌های معنایی واژه‌ها قابل رفع است. بر این اساس، جمله (۱۲) جمله‌ای مورد قبول و پذیرفته نیست زیرا فعل خوردن به فاعلی نیاز دارد که کنش‌گر باشد؛ یعنی توانایی انجام فعل (در این مورد عمل خوردن) را داشته باشد ضمناً این فعل به مفعولی نیاز دارد که قابل خوردن باشد، مانند سیب.

(۱۲). سیب علی را خورد.

اگرچه استفاده از مشخصه‌های معنایی در مثال (۱۲) ابهام را مرتفع می‌سازد، گاه با مواردی مواجه می‌شویم که مشخصه‌های معنایی نیز قادر به رفع ابهام نیستند. به عنوان مثال، در جمله‌های (۱۳) و (۱۴) واژه‌های «چنگال» و «همراه» دارای ابهام‌اند و با مراجعه به مشخصه‌های معنایی نیز نمی‌توان ابهام موجود را برطرف نمود. تنها باید امیدوار باشیم که در جایی از متن این معما حل شده باشد.

(۱۳). برای خوردن سوسیس از چنگالش استفاده کرد.

(۱۴). همراه علی گم شد.

در مثال (۱۳) «چنگال» می‌تواند به دو معنی باشد: (۱). وسیله‌ای برای خوردن که با قاشق به کار می‌رود و (۲). دست. در مثال (۱۴) نیز «همراه» می‌تواند به معنای فرد «همراه» یا «تلفن همراه» باشد. در هیچ یک از این دو مورد مشخصه‌های معنایی قادر به رفع ابهام نیستند.

۳-۱-۲. ابهام انتقالی

تاکنون از ابهام مقوله‌ای، واژه‌های هم‌آوا - هم‌نویسه، هم‌معنایی و هم‌نویسه به عنوان نمونه‌هایی از ابهام واژگانی صحبت کردیم. تمامی این ابهامها از یک جهت به هم شباهت دارند و آن این است که همگی ابهامهایی تک‌زبانانه‌اند؛ یعنی در یک زبان خاص اتفاق می‌افتند. اما در کار ترجمه ماشینی با نوع دیگری از ابهام مواجه‌ایم که بین دو زبان اتفاق می‌افتد. این نوع ابهام که ابهام انتقالی یا ابهام در ترجمه نامیده می‌شود، زمانی اتفاق می‌افتد که یک واژه زبان مبدأ بتواند بالقوه به بیش از یک واژه در زبان مقصد ترجمه شود. نمونه بارز این نوع ابهام را می‌توان در واژه‌های «واگرا» و «اصل واگرایی» مشاهده کرد. بر اساس این اصل، یک واژه زبان مبدأ با حفظ صورت به معانی متعددی اشاره می‌کند، ولی در زبان

مقصد برای بیان آن معانی مختلف، واژه‌ها و صورت‌های متفاوتی وجود دارد، مانند واژه Cousin در انگلیسی که دارای ۸ معنای مختلف است ولی در فارسی برای بیان آن ۸ معنی از ۸ واژه پسرعمو، دخترعمو، پسردایی، دخترعمو، پسرعمو، پسرعمو، دخترخاله و دخترخاله استفاده می‌شود. این گونه واژه‌ها اغلب در زبان مقصد ایجاد مشکل می‌کند و باید به واژه‌های مختلف شکسته شوند. بنابراین، این نوع ابهام به امر ترجمه مربوط است و نه به تجزیه و تحلیل یک زبان خاص.

۴- ابهام ساختاری

در بخش ۳ ریشه بیشتر ابهام‌های موجود را در واژه جستجو کردیم، اما نوع دیگری از ابهام نیز وجود دارد که ریشه آن (به جای واژه) ساختار دستوری جمله زبان مبدأ است. این نوع ابهام زمانی اتفاق می‌افتد که بر اساس قواعد موجود در دستور بتوانیم یک جمله روساختی بیش از یک ژرف ساخت قائل شویم. بنابراین، در این نوع ابهام دو نکته از اهمیت بسیاری برخوردار است: ۱- سطح یا واحد جمله و ۲- دستور مورد استفاده در سیستم ترجمه.

بیشتر سیستم‌های ترجمه ماشینی تجزیه و تحلیل را صرفاً در واحد جمله به انجام می‌رسانند و به واحدهای بزرگتر از آن مثل بند و ... و نیز نحوه ارتباط جملات نمی‌پردازند. ضمناً تجزیه گر ماشین ترجمه در امر تجزیه و تحلیل نمی‌تواند خارج از دستور موجود عمل کند. بنابراین، عدم کفایت دستور نیز سبب بروز مشکل در عملکرد تجزیه گر ماشین ترجمه خواهد شد. در حقیقت، با استفاده از دستور زبان است که ماشین ترجمه قرائت‌های مختلف را شناسایی می‌کند. ابهام ساختاری دارای انواع مختلفی است که در ادامه به اختصار به دو مورد از مهمترین آنها با عناوین «ابهام‌های ساختاری واقعی و سیستمی (تصادفی)» اشاره می‌شود.

۴-۱. ابهام ساختاری واقعی

این نوع ابهام بسیار مورد توجه زبان‌شناسان بوده است تا بتوانند از میان تعبیرهای نحوی مختلف، تعبیر مناسب را برگزینند. این نوع ابهام‌ها را واقعی می‌نامیم چون سخنگویان بومی نیز جمله را دارای ابهام ارزیابی می‌کنند. به مثال زیر توجه کنید:

(۱۵). The man saw the girl with the telescope.

این جمله در انگلیسی دارای ابهام ساختاری واقعی است و می‌تواند دارای یکی از تعابیر زیر باشد:

(۱۵a) The man saw the girl who had the telescope.

(۱۵b) The man saw the girl with the aid of the telescope.

نکته مهم این است که این جمله در یک بافت خاص عاری از هر گونه ابهامی باشد. مثلاً اگر این جمله در داستانی به کار رود، دنبال کردن جریان داستان مشخص می‌کند که کدام یک از دو تعبیر فوق مد نظر بوده است. این در حالی است که سیستم ترجمه قادر به دنبال کردن خط سیر داستان نیست. بنابراین، در این مورد ابهام واقعی با نوع دیگری از ابهام که سیستمی نامیده می‌شود، ترکیب شده است. این مورد در فارسی نیز به فراوانی اتفاق می‌افتد. به مثالهای زیر توجه کنید:

(۱۶). او به مدیر تور گفت که نمی‌تواند بدود.

(۱۶الف) او به مدیر تور گفت که مدیر تور نمی‌تواند بدود.

(۱۶ب) او به مدیر تور گفت که او (فاعل) نمی‌تواند بدود.

(۱۷). برخی از مردان و زنان ایران.

(۱۷الف) برخی از مردان ایران و برخی از زنان ایران.

(۱۷ب) برخی از مردان ایران و تمام زنان ایران.

همان‌گونه که در مثالهای بالا مشاهده می‌شود، این جملات حتی برای سخنگویان بومی نیز ابهام تولید می‌کند. معمولاً در بسیاری از دستورها برای رفع این مشکل از اصل قرابت یا نزدیکی استفاده می‌شود. بدین ترتیب «نمی‌تواند» به «مدیر تور» و «برخی» به «مردان» برمی‌گردد. اما باید توجه داشته باشیم که این اصل همواره صادق نیست و ممکن است منظور نویسنده تعبیر دیگری باشد؛ مثلاً منظور او واقعاً برخی از مردان و برخی از زنان ایران باشد.

نوع دیگری از ابهام ساختاری واقعی به عبارت حرف اضافه‌ای مربوط می‌گردد. مثلاً در جمله (۱۸).

(۱۸). دانشمندی از ایران همراه علی بود.

مشخص نیست که «از ایران» صفت نسبی برای دانشمند است (دانشمند ایرانی) و یا نشان‌دهنده نقطه شروع است (از ایران آن دانشمند همراه علی بود). گاه رعایت نکردن علائم نشانه‌گذاری نیز به ابهامی ساختاری منجر می‌شود، مانند مثال (۱۹).

(۱۹). آن کتاب تاریخ را توصیف کرد.

(الف) موضوع آن کتاب، توصیف تاریخ است. (در این صورت بعد از واژه کتاب باید کاما وجود داشته باشد)
(ب) فردی آن کتاب تاریخ را توصیف کرد.

بدیهی است، سخنگویان بومی با مشاهده علایم نشانه‌گذاری صحیح و با استفاده از اطلاعات متنی نسبت به انتخاب تعبیر درست اقدام به عمل می‌آورند. حال آنکه انتخاب تعبیر درست از سوی ماشین ترجمه بسیار دشوارتر خواهد بود.

نوع دیگری از ابهام به ساختار واژه در جمله مربوط می‌گردد. این مورد اغلب در مورد اسامی غیر بسیط نیز اتفاق می‌افتد. در زبان فارسی اغلب اسامی غیر بسیط به شکلهای مختلفی نوشته می‌شوند که این تنوع برای سیستم ترجمه و حتی فرد مترجم، مشکلاتی را ایجاد می‌کند. به عنوان مثال، در جمله (۲۰) چنانچه واژه «همکلاس» به صورت پیوسته نوشته نشود، می‌تواند ابهام‌زا باشد.

(۲۰). من هم کلاس را دیدم.

(الف) من هم، کلاس را دیدم.

(ب) من همکلاس را دیدم.

متأسفانه در زبان فارسی برای رفع این مشکل راه حلی با ضمانت اجرایی مناسب که بتواند سبب یکدست‌سازی نوشتار فارسی گردد ارائه نگردیده و یا اگر هم مطرح شده از ضمانت اجرایی لازم برخوردار نیست. مسلماً این گونه اسامی چنانچه سر هم نوشته شوند مشکل‌ساز نخواهند بود و حتی در صورت استفاده از صورت ناپیوسته آن، می‌توان از طریق ایجاد یک فاصله بین هم و کلاس (مفهوم ۲۰ الف) و ایجاد فاصله کمتر نیم فاصله بین هم و کلاس (مفهوم ۲۰ ب) تمایز لازم را ایجاد نمود.

۴-۲. ابهام ساختاری سیستمی

گاه علت بروز وجود نقصان در سیستم است که نمی‌تواند از جمله ابهام‌زدایی کند. به بیان دیگر، در این گونه موارد سخنگویان بومی به راحتی جمله را درک و ابهام‌زدایی می‌کنند ولی ماشین ترجمه قادر به انجام چنین کاری نیست. این نوع ابهامها را سیستمی می‌نامیم و به انواع مختلفی تقسیم می‌شوند که در ادامه به چند مورد در زبان فارسی و انگلیسی اشاره می‌شود.

گاه ابهام ساختاری تصادفی زمانی اتفاق می‌افتد که واژه‌های دارای ابهام مقوله‌ای، به صورت تصادفی با هم ترکیب می‌شوند.

(۲۱). Fruit flies like grape.

در جمله (۲۱) Flies می‌تواند فعل سوم شخص مفرد یا اسم جمع باشد؛ Like می‌تواند نقش فعلی یا اسمی داشته باشد و ... در حقیقت در جمله بالا Fruit Flies اسم مرکب است و چنانچه به صورت توالی فاعل و فعل تجزیه شود، معنای جمله عوض خواهد شد. بنابراین، به جای «مگسهای میوه انگور را دوست دارند» جمله ممکن است به صورت «میوه دوست داشتن را انگور پرواز می‌کند» ترجمه شود. سخنگویان بومی این جمله را به شکلی درست تقطیع خواهند کرد. چون از اطلاعات ژرف ساختی استفاده می‌کنند و تعبیر دیگر را منطقاً نامعقول می‌دانند.

نوع دیگری از ابهام ساختاری تصادفی زمانی اتفاق می‌افتد که با فعل مرکب و یا توالی‌ای از یک سری اسامی مثل یک اسم مرکب، سر و کار داشته باشیم.

(۲۲). گل یخ یخ زد.

اسم مرکب فعل مرکب

در مثال (۲۲) یک اسم مرکب (گل یخ) و یک فعل مرکب (یخ زد) وجود دارد و هر گونه تقطیع دیگری مثلاً به صورت «گل + یخ + یخ + زد» یا «گل + یخ + یخ زد»، مانع دستیابی ما به تعبیر درست خواهد شد. تعبیر درست «گل یخ + یخ زد» است. با توجه به فراوانی اسامی مرکب در فارسی و انگلیسی و نیز فراوانی ابهامهای مقوله‌ای در این دو زبان، با نمونه‌های زیادی از ابهام ساختاری مواجه خواهیم شد.

نوع دیگری از ابهام ساختاری سیستمی به عدم تطابق حروف اضافه در زبانهای مختلف مثل فارسی و انگلیسی مربوط می‌گردد. بنابراین، سیستم باید در خصوص نحوه کاربرد حروف اضافه و ادات (که مورد اخیر به همراه فعل می‌آید) اطلاعات درستی داشته باشد.

(۲۳). علی در شیراز در به در شد.

برای ترجمه جمله (۲۳) به زبان انگلیسی سیستم باید بتواند بین سه مورد وقوع واژه «در»، تمایز قایل شود. ضمناً در حین ترجمه باید به شکلی بین دو حرف اضافه at و in یکی را برگزیند. سیستم باید بداند که شیراز شهر بزرگی است و بنابراین حرف اضافه in مناسب است و نه at.

۳-۴. رفع ابهام ساختاری

- برای رفع ابهام، راه‌حلهای مختلفی وجود دارد که از مهمترین آنها می‌توان به موارد زیر اشاره کرد:
- استفاده از اطلاعات مربوط به مؤلفه‌های معنایی و دانش زبانی
- استفاده از کلیدهای بافتی
- استفاده از دانش غیر زبانی (اطلاعات مربوط به جهان پیرامون)
- مشورت با افراد متخصص
- استفاده از سیاست گمانه زنی یا بهترین حدس
- چشم‌پوشی کردن از موارد ابهام‌زا

۳-۴-۱. استفاده از اطلاعات مربوط به مؤلفه‌های معنایی و دانش زبانی

اغلب جملات مبهم را می‌توان از طریق مراجعه به دانش زبانی و مؤلفه‌های معنایی ابهام زدایی کرد. دانش زبانی دارای انواع مختلفی است. نکته مهم این است که در تمام آنها از اطلاعات مربوط به واژه‌ها و نحوه ترکیب آنها استفاده می‌شود و به وقایع جهان پیرامون توجهی نمی‌شود. این گونه قالبها نوع عناصری را که می‌توانند پیش یا پس از فعل ظاهر شوند، مشخص می‌کنند. به عنوان مثال، فعلی مانند «دادن» حداقل به سه گروه اسمی نیاز دارد که دو مورد دارای نقشهای دهنده (فاعلی) و گیرنده (مفعولی) است و مورد دیگر به چیز یا شیئی اشاره دارد که داده یا گرفته می‌شود.

(۲۴). **علی کتاب را به محمد داد.**

محمد = گیرنده کتاب، علی = دهنده کتاب، کتاب = شیء داده شده یا گرفته شده. برای انتخاب گروه‌های اسمی مناسب نیز از مؤلفه‌های معنایی مثل [+/-جاندار] [+/-انسان] و ... استفاده می‌شود. این مؤلفه‌ها در مثال بالا مانع ایفای نقش فاعلی از سوی واژه کتاب می‌شوند.

۲-۳-۴. **کلیدهای بافتی**

در زبان تعداد جملاتی که حاوی نمونه‌هایی از ابهام واقعی باشند، بسیار کم است. چنانچه از طریق مؤلفه‌های معنایی و دانش زبانی ابهام‌زدایی نشود، از اطلاعات بافتی استفاده می‌گردد. در این گونه موارد کلیدهای بافتی مشخص می‌کند که کدام تعبیر، درست یا مورد نظر است. به عنوان نمونه، در مثالهای ۱۴ و ۱۵ تنها راه درک جمله این است که فرد از جملات قبل، بندهای قبل و یا حتی فصلهای قبل مربوط به آن مطلع باشد. سیستمهای ترجمه اغلب نمی‌توانند از این نوع اطلاعات بافتی استفاده کنند چون هیچ قانون مشخصی وجود ندارد که بر اساس آن سیستم بتواند محل اطلاعات مورد نیاز برای ابهام‌زدایی را شناسایی نماید.

۳-۳-۴. **استفاده از دانش غیر زبانی (اطلاعات مربوط به جهان پیرامون)**

گاه پیش می‌آید که کلیدهای بافتی و مؤلفه‌های معنایی هیچ یک نمی‌توانند ابهام موجود را مرتفع سازند. در چنین مواردی، گاه با استفاده از اطلاعات مربوط به جهان پیرامون، می‌توان ابهام‌زدایی کرد.

(۲۵). **علی اسب را با دوربین دید.**

در جمله بالا، ریشه ابهام در عبارت حرف اضافه‌ای با «دوربین» است. در حقیقت، می‌توانیم دو تعبیر زیر را تصور کنیم:

(الف) علی از دوربین برای دیدن استفاده کرده و (ب) اسب دوربین داشته و علی هر دو را با هم دیده است.

در این مورد، با استفاده از اطلاعات جهان بیرون می‌توانیم ابهام‌زدایی کنیم؛ بدین ترتیب که بر اساس تجربیات ما از جهان پیرامون، تعبیر (الف) منطقی‌تر به نظر می‌رسد.

(۲۶). The man saw the girl with red hair.

در مثال (۲۶) نیز برای درک عبارت حرف اضافه‌ای از اطلاعات جهان پیرامون استفاده می‌شود و تعبیر «آن مرد دخترِ مو قرمز را دید» مناسب و مرتبط تشخیص داده می‌شود و تعبیر «آن مرد دختر را با موی قرمز دید» تعبیری نادرست محسوب می‌گردد، چون بر اساس تجربیات جهان خارج می‌دانیم نمی‌توان با موی قرمز (به عنوان یک وسیله) شیء یا فردی را دید، بلکه برای دیدن از ابزاری چون دوربین، تلسکوپ و ... استفاده می‌شود. همان‌گونه که پیشتر نیز ذکر شد، وارد نمودن تمامی اطلاعات جهان پیرامون به شکلی که بتواند ابهامهای احتمالی را مرتفع سازد، کاری غیرممکن و مدام در حال تغییر است. پیشرفتهای به دست آمده در هوش مصنوعی نیز نتوانسته بهبود چندانی پدید آورد، بنابراین، از پیچیدگی و گستردگی اطلاعات جهان پیرامون می‌توان به عنوان دو عامل مهم که مانع دستیابی به راه حلی فوری می‌شوند نام برد.

۴-۳-۴. مشورت با افراد متخصص

یکی دیگر از راهکارهای ابهام‌زدایی، به تماس با نویسنده و یا مشورت با افراد متخصص مربوط می‌گردد؛ بدین ترتیب که افراد متخصص در امر ترجمه وارد شده و بر اساس درک خود، یکی از تجزیه و تحلیلها را برمی‌گزینند. البته این تدبیر نیز قادر نیست همواره ما را به تعبیر درست برساند.

۴-۳-۵. استفاده از سیاست گمانه‌زنی یا بهترین حدس

چنانچه با استفاده از راهکارهای بالا ابهام مرتفع نشود، جمله دارای ساختار رایج‌تر به عنوان تعبیر مورد نظر برگزیده می‌شود. مثلاً عبارتهای حرف اضافه‌ای اغلب اسم پیش از خود را توصیف می‌کنند؛ عبارتهای نزدیکترین فعل به خود را توصیف می‌کنند و ... البته از آنجا که این رویکرد بر پایه حدس و گمان بنا شده، از ضریب خطای بالایی برخوردار است.

(۲۷). Investigation of techniques of stress analysis by thermal emission.

در مثال (۲۷) حدس ما این خواهد بود که عبارت by thermal emission به اسم

پیش از خود یعنی analysis برمی‌گردد و نه به techniques یا investigation.

۴-۳-۶. چشم‌پوشی از مورد ابهام‌زا

در این رویکرد، ابهام موجود در زبان مبدأ به همان صورت و بدون ابهام‌زدایی وارد زبان مقصد می‌شود؛ بدین ترتیب که جمله زبان مقصد نیز دارای ابهامی مانند ابهام موجود در جمله زبان مبدأ می‌شود. از این روش صرفاً در مواردی می‌توان استفاده کرد که زبانهای مبدأ و مقصد از نظر ساختار و واژگان بسیار به هم شبیه و نزدیک باشند، مثل آلمانی و فرانسه. البته، حتی در این دو زبان نیز همیشه امکان انجام چنین کاری وجود ندارد. نکته‌ای که در خصوص ابهام‌های شش‌گانه فوق باید ذکر شود این است که از نظر خود سیستم این‌گونه تمایزها نامربوط است. برای سیستم نوع ابهام از اهمیت چندانی برخوردار نیست. نکته حائز اهمیت این است که آیا سیستم برای رفع ابهام، اطلاعات و امکانات لازم را در اختیار دارد یا خیر؟

۴-۳-۷. مرجع‌یابی

مرجع‌یابی به بررسی ارتباط بین اسم و مرجع آن می‌پردازد که این دو ممکن است در یک جمله یا در جملات، بندها و حتی صفحات مختلف وجود داشته باشند. بحث در خصوص مرجع‌یابی خود به مقاله‌ای جدا نیاز دارد، ولی از آنجا که آن نیز نوعی ابهام به شمار می‌آید، در قسمت پایانی مقاله اشاره کوتاهی به آن خواهیم داشت. مرجع‌یابی نیز می‌تواند تولید ابهام کند، چون ممکن است مرجع یک ضمیر نامشخص باشد یا کاملاً مشخص نباشد. به مثالهای زیر توجه کنید:

(۲۸). میمون سیب را خورد چون آن گرسنه بود.

(۲۹). میمون سیب را خورد چون آن رسیده بود.

(۳۰). Me Killed her.

در مثال (۲۸) ضمیر «آن» به «میمون» و در مثال (۲۹) به «سیب» برمی‌گردد. بنابراین، در مثال (۲۸) «میمون گرسنه بود» و در مثال (۲۹) «سیب رسیده بود»، چرا که محدودیت‌های هم‌آوایی مشخص می‌سازد که سیب نمی‌تواند گرسنه باشد (برخلاف انسان و حیوان) و میمون نمی‌تواند (برخلاف میوه) رسیده باشد. ولی در جمله (۳۰) با مشکل دیگری مواجه

هستیم؛ بدین ترتیب که ترجمه ماشینی این جمله به صورت «او او را کشت» خواهد شد و می‌بینیم که ابهام این جمله از جمله زبان مبدأ بیشتر است. علت این امر آن است که در جمله انگلیسی برای مذکر و مؤنث از ضمیری جدا استفاده می‌کنیم، ولی در زبان فارسی برای هر دو جنس از ضمیر «او» استفاده می‌شود. بنابراین، جمله ترجمه شده ابهام بیشتری را در خود دارد. برای رفع این مشکل، باید ضمن یافتن مرجع این دو ضمیر، دست کم یکی از آنها را رمزگشایی کنیم؛ مثلاً «او مری را کشت»، «جک مری را کشت»، «او آن زن را کشت» و ... طبیعی است با خروج مرجع‌یابی از سطح جمله و گسترش کار به بند و صفحه و ... کار مرجع‌یابی دشوارتر و برای ماشین تا حد زیادی ناممکن می‌گردد.

خلاصه

در این مقاله، ضمن بررسی ابهام‌های واژگانی و ساختاری، سعی شد مشکلاتی که این‌گونه ابهام‌ها برای ماشین‌های ترجمه ایجاد می‌کنند، تشریح گردد. در پایان نیز شش روش مختلف ابهام‌زدایی معرفی گردید. از آنچه گفته شد می‌توان چنین استنباط نمود که سیستم‌های ترجمه برای کارایی مطلوب باید از نظریه‌ای مناسب و جامع استفاده نمایند. ضمناً سیستم ترجمه در صورتی موفق خواهد بود که بتواند مانند ذهن یک انسان در هنگام مواجهه با پیچیدگی‌های زبانی از اطلاعات نحوی، واژگانی، معنایی، بافتی، دانش جهان پیرامون و ... استفاده نماید. نبود ماشین ترجمه فارسی به انگلیسی، نشان می‌دهد که تاکنون محققان کشور در این خصوص تحقیقات منسجم و دنباله‌داری را به انجام نرسانده‌اند. با توجه به اینکه بیشتر نظریه‌های زبانی از سوی محققان انگلیسی ارائه گردیده (حتی دستور زبان چامسکی نیز پس از گذشت حدود نیم قرن نتوانسته جای خود را در مدارس باز کند) حتی برای زبان انگلیسی و زبانهای مشابه آن نیز با مشکلات فراوانی مواجه است. این موارد نشان می‌دهد برای رسیدن به ماشین‌های ترجمه مناسب هنوز باید کار بسیاری در حیطه زبان‌شناسی رایانه‌ای به انجام برسد که همان‌گونه که پیشتر نیز ذکر شد، در حال حاضر چنین رشته‌ای در دانشگاه‌های ایران تدریس نمی‌شود.

منابع

- اعلم، هوشنگ (۱۳۶۸)، *ابهام و ابهام‌زدایی در زبان فارسی*، کیهان فرهنگی، ش ۴.
باطنی، محمدرضا (۱۳۷۱)، *نگاهی تازه به دستور زبان*، تهران، آگاه.
داوری اردکانی، نگار (۱۳۷۵)، *بررسی ابهام*، *ابهام نامفهومی در زبان فارسی*، رساله
کارشناسی ارشد، زبان‌شناسی، تهران، دانشگاه تربیت مدرس.
صحت، نازنین (۱۳۷۷)، *بررسی ابهام ساختاری در زبان فارسی*، رساله کارشناسی
ارشد زبان‌شناسی، تهران، دانشگاه آزاد اسلامی.
صفوی، کورش (۱۳۸۰)، *درآمدی بر معنی‌شناسی*، تهران، حوزه هنری سازمان
تبلیغات اسلامی.
علوی، نگار (۱۳۷۹)، *بررسی و طبقه‌بندی گونه‌های رفع ابهام در زبان فارسی*،
رساله کارشناسی ارشد زبان‌شناسی، تهران، دانشگاه آزاد اسلامی.
قنبری، عاطفه (۱۳۷۷)، *بررسی مسئله چند معنایی در زبان فارسی*، رساله کارشناسی
ارشد زبان‌شناسی، تهران، دانشگاه علامه طباطبایی.
معدنی، میترا (۱۳۷۶)، *بررسی و طبقه‌بندی گونه‌های ابهام در زبان فارسی*، رساله
کارشناسی ارشد زبان‌شناسی، تهران، دانشگاه آزاد اسلامی.
Dang, yi Chia, Palmer, Chiou Dang et al. (۲۰۰۲). "Simple Features for Chinese Word sense disambiguation." In Chen, H-H. Lin, C.-Y.-, Proceeding of the ۱۹th International Conference on Computational Linguistics (COLING ۲۰۰۲).
Empson, W. (۱۹۳۰). *Seven Types of Ambiguity*. London: Chatto.
Kess, J.F. and Hoppe R.A. (۱۹۸۵). *Bias, Individual References & Shared knowledge in Ambiguity*. *Journal of Programatics*, ۹, ۲۱-۳۹.
Lyons, J. (۱۹۷۰) *Semantics*. ۲ Vols. Cambridge: Cambridge University Press.

MacDonald, M.C., Pearlmutter, N.J. and Seidenberg, M.S. (۱۹۹۴). Syntactic Ambiguity Resolution as Lexical Ambiguity Resolution. In C. Clifton, L. Frazier, and K. Rayner (Eds.), *Perspectives on Sentence Processing*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Magnini, B., et al. (۲۰۰۲). The Role of Domain Information in word sense Disambiguation. *Natural Language Engineering*, ۸ (۴), ۳۵۹-۳۷۳.

Mihalcea, R. (۲۰۰۲). Instance based Learning with Automatic Feature Selection Applied to word sense Disambiguation. In Chen, H-H. Lin, C.-Y., *Proceedings of the ۱۹th International Conference on Computational Linguistics (COLING ۲۰۰۲)*.

Saeed, J.I. (۱۹۹۷). *Semantics*. Oxford: Blackwell.

Spivey-Knowlton, M.J., Trueswell, J.C., and Tanenhaus, M.K. (۱۹۹۳). Context and Syntactic Ambiguity Resolution. *Canadian Journal of psychology*, ۴۷, ۲۷۶-۳۰۹.