



## کاربرد پیکره‌های زبانی پویا در ترجمه اینترنتی متون فارسی به انگلیسی؛ بررسی عوامل مؤثر در بروز خطا

سید محمد حسینی معصوم<sup>۱</sup>، مصطفی غضنفری<sup>۲</sup>

### چکیده:

زبان‌شناسی پیکره شاخه‌ای از زبان‌شناسی است که با گسترش فناوری‌های نوین و شبکه جهانی اینترنت شاهد پیشرفت‌های چشم‌گیری بوده است. امروزه کاربرد شبکه جهانی تنها به جست و جوی مطالب و ارائه اطلاعات محدود نمی‌شود. گسترش روزافزون فناوری‌های نوین که در این شبکه ارائه می‌شوند سبب پیدایش زمینه‌های جدید در این عرصه شده است. ترجمه‌ی برخط، تحت شبکه یکی از این زمینه‌های نوین است که در حال حاضر دوران ابتدایی تکامل خود را طی می‌کند. مهمترین ویژگی چنین روشی در ترجمه، استفاده از پیکره‌های زبانی پویاست که دارای اساسی تعاملی و آمار-بنیاد است و به طور لحظه‌ای در حال ارتقاء می‌باشد. این مقاله با تکیه بر یافته‌های زبان‌شناسی پیکره برای نخستین بار تقسیم‌بندی جدیدی از انواع پیکره ارائه می‌دهد و ضمن معرفی شیوه عملکرد ترجمه‌های اینترنتی مبتنی بر پیکره‌های زبانی پویا، به معرفی و آزمون عوامل گوناگون، بالاخص عوامل فرهنگی مؤثر در بروز خطا در ترجمه اینترنتی از فارسی به انگلیسی می‌پردازد. گرچه یافته‌های این تحقیق حاکی از وجود مشکلات جدی در این مسیر است، به نظر می‌رسد الگوریتم‌های پیشرفته نوین در نرم افزارهای معیار ترجمه بتواند گامی امیدوارکننده بسوی تکامل این فناوری و افزایش سرعت و بهبود کیفیت‌ای شیوه از ترجمه ماشینی باشد.

**کلید واژه‌ها:** زبان‌شناسی پیکره؛ زبان‌شناسی رایانه‌ای؛ پیکره زبانی پویا؛ ترجمه اینترنتی؛ ترجمه ماشینی

hosseinimasum@pnu.ac.ir

۱- استادیار گروه زبان‌شناسی و زبانهای خارجی دانشگاه پیام نور (نویسنده مسئول)

aneeshtan@gmail.com

۲- دانشجوی کارشناسی ارشد مطالعات ترجمه دانشگاه آزاد اسلامی، واحد قوچان

**مقدمه:**

با گسترش روزافزون زمینه‌های علمی میان‌رشته‌ای، شاهد پیوندهای جدید میان رشته‌هایی از علوم هستیم که قبلاً کمتر مرتبط به نظر می‌رسیدند. امروزه دانش مهندسی کامپیوتر که خود حاصل پیوند ریاضی و الکترونیک است، با استفاده از یافته‌های علوم اطلاع‌رسانی و نیز دانش زبان‌شناسی، زمینه کاربردی جدیدی را به وجود آورده است که به آن ترجمه اینترنتی می‌گویند. ترجمه اینترنتی نوع خاصی از ترجمه ماشینی<sup>۱</sup> است که به عنوان پدیده‌ای نو در دنیای ارتباطات و اطلاع‌رسانی در حال تحول سریع و تبدیل شدن به یک دانش کاربردی می‌باشد. ترجمه ماشینی که خود شاخه‌ای از زبان‌شناسی رایانه‌ای<sup>۲</sup> محسوب می‌شود، عبارت است از کاربرد یک سیستم نرم‌افزاری رایانه‌ای که از طریق آن می‌توان متون ورودی زبان مبدأ را طی فرایندهایی خاص و پیچیده در سطوح مختلف زبان‌شناختی تجزیه و تحلیل و در نهایت به زبان (یا زبان‌های مقصد) ترجمه کرد و به صورت خروجی در اختیار کاربر قرار داد (خداپرستی، ۱۳۸۰).

هر چند ایده ترجمه ماشینی به چندین دهه قبل باز می‌گردد و در مورد آن تحقیقات متعددی انجام شده است، ترجمه اینترنتی به عنوان نوعی خاص از این روش ترجمه پدیده‌ای جدید و در حال گسترش است. جالب این است که علی‌رغم این که ترجمه ماشینی در پروژه‌های بسیار پرهزینه دنبال شده و هنوز نتیجه رضایت بخشی از آن حاصل نشده است، ترجمه اینترنتی به سرعت در حال کسب جایگاهی برجسته در دنیای ارتباطات است و با وجود تمام کاستی‌های آن به عنوان راه حلی سریع و ارزان برای نیازهای کاربران در سراسر دنیا پذیرفته شده است. هنگامی که از ترجمه اینترنتی سخن می‌گوییم باید دیدگاهی واقع‌گرایانه داشته باشیم و انتظارات خود را با وضع موجود هماهنگ کنیم. هنوز بسیار زود است که از ترجمه اینترنتی انتظار عملکردی در حد مترجم انسانی داشته باشیم. پژوهش حاضر تلاشی است در جهت بررسی نقش زبان‌شناسی پیکره در پیشرفت این شیوه خاص از ترجمه و تحلیل عوامل مؤثر در بروز خطا در برون‌داد چنین سامانه‌هایی به ویژه با توجه به داده‌های زبان فارسی و انگلیسی. بدین منظور ضمن ارائه توصیفی از عملکرد پیکره‌های زبانی پویا به عنوان اساس کار ترجمه اینترنتی، به مطالعه نمونه‌هایی عملی از برون‌داد پایگاه ترجمه اینترنتی گوگل در ترجمه متون فارسی به انگلیسی می‌پردازیم.

<sup>۱</sup>. Machine Translation

<sup>۲</sup>. computational linguistics

## پیشینه ترجمه در بسترهای الکترونیک

شاید در کنار تمام ویرانی‌هایی که جنگ جهانی و در پی آن جنگ سرد به بار آورد، باید فهرستی از نوآوری‌های علمی و فنی را هم اضافه کرد که مستقیم یا غیر مستقیم حاصل رقابت‌های دو ابرقدرت شرق و غرب در کسب موقعیت برتر در جهان جدید بود. زبان و ترجمه هم از این امر مستثنی نیست. همانگونه که یکی از موفق‌ترین روش‌های آموزش زبان به نام روش شنیداری-گفتاری<sup>۱</sup> نخستین بار در ارتش آمریکا پایه‌گذاری شد، نخستین تلاشها در ترجمه ماشینی هم در دهه ۱۹۵۰ و با ترجمه جملات روسی به انگلیسی در کشور آمریکا صورت گرفت. هر چند در آغاز نتیجه کار شگفت‌انگیز می‌نمود و مشکلات موجود به سادگی قابل حل به نظر می‌رسید، عملاً پیشرفت پروژه‌های ماشین ترجمه بسیار کند بود و پس از مدتی سرمایه‌گذاری در این زمینه به شدت کاهش یافت.

با گسترش کاربرد رایانه و ارزان‌تر شدن فن‌آوری‌های دیجیتال، بار دیگر امیدهایی در این زمینه پیدا شد و از دهه ۱۹۸۰ حرکت جدیدی در این عرصه آغاز گردید. در ابتدایی‌ترین سطح، ترجمه رایانه‌ای شامل جایگزینی واژه‌های زبان مبدأ با معادل‌هایی از زبان مقصد است. اما این کار به تنهایی نمی‌تواند منجر به تولید یک ترجمه مناسب شود، چرا که مترجم نیاز به تشخیص عبارات‌ها و سازه‌های بزرگ‌تر از واژه و یافتن معادل‌های مناسب در زبان مقصد دارد. حل این مشکل با کاربرد پیکره‌ها و تکنیک‌های آماری میسر است که این زمینه به شدت در حال پیشرفت است و حرکتی پیوسته در راستای بهبود ترجمه‌ها را باعث شده است. از دیدگاه سامرز (۲۰۰۱: ۷۰) ترجمه ماشینی در سطوح گوناگون از توانایی‌های یک رایانه در جستجو و مقایسه سود می‌برد و در کنار ترجمه انسانی قرار می‌گیرد و به همین دلیل به آن، «ترجمه به کمک رایانه»<sup>۲</sup> نیز می‌گویند. سامرز همچنین در اهمیت ترجمه ماشینی عنوان می‌کند، طی ۱۰ سال، تنها در آمریکا بیش از ۲۰ میلیون دلار صرف فعالیت در زمینه ترجمه ماشینی شد؛ کشوری که در حوزه نظامی و بخاطر جنگ سرد پیشرو در رمزگشایی و برگردان مکالمات دشمن بود. این حجم گسترده سرمایه‌گذاری در زمینه ترجمه ماشینی و اهمیت آن از جنبه سیاسی، نظامی و علمی همگی ناظر به اهمیت مطالعه نقاط ضعف ترجمه ماشینی است. در این راه، کاربرد علم زبانشناسی به ویژه زبانشناسی

<sup>۱</sup>. audio-lingual method

<sup>۲</sup>. Computer Aided translation

پیکره<sup>۱</sup> و زبان‌شناسی رایانه‌ای به کمک متخصصان علم مهندسی آمده است تا روز به روز شاهد ارتقاء این فن آوری باشیم.

### زبان‌شناسی پیکره

زبان‌شناسی پیکره عبارت است از مطالعه زبان در قالب نمونه‌های طبیعی گفتار و متون مکتوب. هدف از تهیه این پیکره‌های زبانی استخراج قواعد زبان طبیعی برای اهداف مختلف آموزشی و کاربردی است. یک پیکره زبانی عبارت است از مجموعه‌ای از نمونه‌های طبیعی زبانی که به صورت الکترونیکی جمع‌آوری و ذخیره شده اند (هانستون، ۲۰۰۶، نقل از براون، ۲۰۰۶، ج ۳: ۲۳۴). اغلب پیکره‌های امروزی حداقل یک میلیون واژه دارند و شامل متن‌های کامل یا برگزیده‌ای از متون طولانی می‌شوند. می‌توان زبان‌شناسی پیکره را روشی پیشرفته برای پیدا کردن پاسخ سوالات متداول در زبان‌شناسی دانست. به دیگر سخن، پیکره‌ها بستری برای آزمودن فرضیه‌های زبانی و دادن جنبه کمی و آماری به پژوهش‌های زبانی فراهم می‌کند. پیکره‌ها در اندازه‌های مختلف و برای اهداف متفاوت تدوین می‌شوند.

برای استفاده از پیکره‌ها معمولاً نرم‌افزارهای خاصی طراحی می‌شوند که امکان جست و جوی متنوعی را به کاربران و محققان می‌دهند. این نرم افزارها دارای برخی توانایی‌های پایه هستند که از جمله می‌توان به موارد زیر اشاره کرد:

- امکان جست و جوی یک عنصر خاص (واژه، عبارت، جمله ..)
- شمارش و محاسبه بسامد نسبی عنصر مربوطه
- نمایش موارد کاربرد آن عنصر به منظور بررسی بیشتر توسط پژوهشگر

این رویکرد آمار-بنیان زبان‌شناسان پیکره و عدم توجه به جنبه‌های تبیینی و استدلالی در پژوهش‌های زبانی، گاه از سوی دیگر زبان‌شناسان مورد انتقاد قرار می‌گیرد (ر.ک. میسر، ۲۰۰۲: ۲-۵)، هر چند که اگر تعریف نظریه زبانی را مورد بازنگری قرار دهیم و با دیدی نو به یافته‌های زبان‌شناسان پیکره نگاه کنیم، درمی‌یابیم که این یافته‌ها در پیشبرد جنبه‌های تبیینی نظریه زبان‌شناسی نیز مفید هستند. به علاوه، پیکره‌ها کاربردهای عملی مختلفی دارند که از جمله می‌توان

<sup>۱</sup> . corpus linguistics

به استفاده از آنها در زبانشناسی قانونی<sup>۱</sup>، سبک‌شناسی، و تحلیل گفتمان انتقادی<sup>۲</sup>، اشاره کرد. اما مهم‌ترین کاربرد پیکره‌های زبانی در آموزش زبان و ترجمه، به ویژه ترجمه‌ی ماشینی است (هانستون، ۲۰۰۶، نقل از براون، ۲۰۰۶، ج ۳: ۲۴۵).

می‌توان از پیکره‌های زبانی برای آموزش مترجمان، استفاده کرد. همچنین مترجمان حرفه‌ای هم می‌توانند در کار ترجمه از آنها بهره بگیرند. از پیکره‌ها می‌توان در مطالعه فرایند ترجمه و بررسی نحوه انتخاب عناصر توسط مترجمان نیز یاری جست. معمولاً در چنین مواردی، از پیکره‌های موازی<sup>۳</sup> استفاده می‌کنند که دو پیکره را به نحوی در کنار هم قرار می‌دهند که ترجمه هر جمله در متن اصلی به راحتی قابل بازیابی باشد.

### انواع پیکره های زبانی

با توجه به جدید بودن رشته زبانشناسی پیکره، به ویژه در مفهوم الکترونیکی آن، تقسیم‌بندی‌های متفاوتی از انواع پیکره توسط زبانشناسان و محققین پیکره ارائه شده است. به طور کلی می‌توان گفت که مهم‌ترین ملاک تقسیم‌بندی پیکره‌ها، یک یا چند زبانه بودن آن‌هاست. پیکره‌های تک‌زبانه<sup>۴</sup> شامل نمونه‌های جمع‌آوری شده تنها از یک زبان هستند. چنین پیکره‌هایی عمدتاً در پژوهش‌های زبانی برای شناخت ساختار درونی معنایی، نحوی یا کاربردشناختی<sup>۵</sup> زبان مورد نظر مورد استفاده قرار می‌گیرند. این چنین پیکره‌هایی همگی از نوع مقایسه‌ای هستند و دارای دو گونه می‌باشند:

الف. پیکره‌هایی که متون اصیل و متون ترجمه شده به یک زبان خاص را مقایسه می‌کنند (مثلاً متون اصلی انگلیسی و متونی که از زبان‌های دیگر به انگلیسی ترجمه شده‌اند).

ب. پیکره‌هایی که متون تولید شده توسط گویشوران بومی و زبان آموزان را مقایسه می‌کنند.

پیکره‌های چند زبانه<sup>۶</sup> شامل داده‌هایی از دو یا چند زبان هستند. دو نوع عمده از این پیکره‌ها عبارتند از:

<sup>۱</sup>. Forensic Linguistics (کاربرد زبانشناسی در تشخیص جرم)

<sup>۲</sup>. Critical Discourse Analysis

<sup>۳</sup>. parallel corpora

<sup>۴</sup>. monolingual corpora

<sup>۵</sup>. pragmatic

<sup>۶</sup>. multilingual corpora

الف. پیکره‌های ترجمه‌ای<sup>۱</sup> یا موازی که از متون زبان مبدأ و ترجمه آنها تشکیل شده است

این ترجمه‌ها ممکن است یکسویه<sup>۲</sup> (تنها از زبان الف به زبان ب) یا دو سویه<sup>۳</sup> باشند.

ب. پیکره‌های مقایسه‌ای<sup>۴</sup> که مشتمل بر متون ترجمه نشده یا ترجمه شده از یک گونه‌ی

سبکی خاص در زبان‌های مورد نظر است (گرنجر و دیگران، ۲۰۰۸: ۲۰).

نکته‌ای که در این جا باید مورد توجه قرار گیرد این است که در منابع مختلف، ملاک تقسیم‌بندی انواع پیکره‌ها محتوای آنها و منبع گردآوری آنهاست. اما در این میان ماهیت عملکردی پیکره‌ها و تغییرپذیری یا ثبات آنها کمتر مورد توجه قرار گرفته است. البته شاید علت این امر، جدید بودن پدیده‌ی ارائه‌ی پیکره‌ها در شبکه جهانی است. بنابراین در این پژوهش یک تقسیم‌بندی جدید برای نخستین بار ارائه می‌شود و اصطلاح خاصی برای آن پیشنهاد می‌شود که به نظر نگارندگان می‌تواند برای شناسایی انواع پیکره‌های زبانی مناسب باشد. بر این اساس، بسته به این که پایگاه داده‌های پیکره پس از جمع‌آوری به شکل یک بسته‌ی ثابت مورد استفاده قرار بگیرد یا به طور آنی و مداوم در حال به‌روزرسانی باشد، می‌توان پیکره‌های زبانی را به دو نوع **ایستا<sup>۵</sup> و پویا<sup>۶</sup>** تقسیم بندی کرد.

پیکره‌های زبانی ایستا در گذشته به صورت دستی توسط پژوهشگران جمع‌آوری می‌شدند، اما امروزه این داده‌ها به کمک رایانه و همچنین با استفاده از اینترنت جمع‌آوری، برچسب‌گذاری<sup>۷</sup> و تدوین می‌شوند. ایستایی این پیکره‌ها ممکن است در نگاه نخست، یک محدودیت به نظر آید اما این مزیت را به پیکره می‌دهد که داده‌های ورودی کاملاً کنترل شده و منطبق با اهداف پژوهشگر انتخاب و جمع‌آوری می‌شوند و همگی دارای استانداردهای تعریف شده‌ی پژوهش هستند. ایستایی این پیکره‌ها به معنی ثبات ابدی آنها نیست بلکه پژوهشگر می‌تواند در فواصل زمانی معین با ارائه داده‌های جدید، پیکره خود را روزآمد کند. اما باید توجه داشت این کار هیچ گاه به

<sup>۱</sup>. translation corpora

<sup>۲</sup>. unidirectional

<sup>۳</sup>. bidirectional

<sup>۴</sup>. comparable corpora

<sup>۵</sup>. static corpora

<sup>۶</sup>. dynamic corpora

<sup>۷</sup>. tagging

صورت خودکار انجام نمی‌شود و تنها زمانی که پژوهشگر بخواهد، این امر اتفاق می‌افتد و البته مستلزم صرف وقت زیاد است.

نوع دوم، پیکره‌های زبانی پویا هستند که اساساً تنها در محیط اینترنت و به صورت برخط عمل می‌کنند. این پیکره‌ها با نرم‌افزارهای پیچیده‌ای که از آنها پشتیبانی می‌کند، به طور آنی داده‌های زبانی از زبان‌هایی که برایشان تعریف شده در سراسر صفحات وب جمع‌آوری می‌کنند و معادل‌های زبانی را از زبان‌های مختلف با هم مقابله<sup>۱</sup> میکنند. ترجمه مبتنی بر پیکره‌های زبانی پویا، از روش ترجمه ماشینی آمار-بنیاد<sup>۲</sup> استفاده می‌کند. «در ترجمه آماری سعی می‌شود تا ترجمه مورد نظر با استفاده از روش‌های آماری بر اساس متون دوزبانه به وجود آید. وقتی مقدار زیادی متون دوزبانه در دسترس باشد، نتایج کار ترجمه بسیار شگفت‌انگیز خواهد بود. در این روش‌ها با توجه به هم‌نشینی کلمات در متون مبدأ و مقصد، کلمات مناسب در ترجمه انتخاب می‌شود» (دبیرخانه شورای عالی اطلاع رسانی، ۱۳۸۸: ۸).

نرم‌افزارهای مترجم اینترنتی به دلیل توانایی‌های بالاتر در به‌روز رسانی و افزایش لحظه‌ای دایره پیکره زبانی خود از اهمیت و محبوبیت بیشتری در مقابل نرم‌افزارهای رومیزی برخوردارند. نرم‌افزارهای مترجم برای بالا بردن میزان دقت خود دسته‌بندی‌هایی کلی از شاخه‌های تخصصی و عمومی ارائه می‌دهند تا خروجی دقیق‌تری داشته باشند. ترجمه مقالات فنی و مهندسی، معماری، تاریخی و شاخه عمومی از دسته‌بندی‌های رایج این نوع نرم‌افزارها هستند. نکته‌ای که در اینجا قابل ذکر است این که قرارگرفتن یک پیکره بر روی شبکه لزوماً به معنای پویایی آن نیست. بسیاری از پیکره‌ها که به صورت پراکنده توسط محققان مختلف تهیه شده‌اند ممکن است به صورت برخط ارائه شوند، اما تنها شرکت‌های خدمات اینترنتی بسیار بزرگ با قابلیت‌های نرم‌افزاری بسیار پیچیده می‌توانند پیکره‌های پویا تولید کنند. به ویژه استفاده از پیکره‌های پویا در امر ترجمه، مسأله‌ای است به مراتب پیچیده‌تر. تهیه چنین پیکره‌ای برای مقابله تنها دو زبان، مستلزم فرایندهای آماری، نرم‌افزاری و زبانشناختی پیچیده است. حال در نظر بگیرید، نرم‌افزار ترجمه اینترنتی گوگل با قابلیت ترجمه سریع ده‌ها زبان به یکدیگر، متکی بر چه پایگاه داده‌های عظیمی است.

<sup>۱</sup>. Aligning

<sup>۲</sup>. Statistical Machine Translation

## فرآیند ترجمه رایانه‌ای

روش‌های مختلفی برای ترجمه رایانه‌ای وجود دارد که از جمله می‌توان به روش قاعده-بنیاد<sup>۱</sup>، روش آماری و روش مبتنی بر مثال اشاره کرد. همان گونه که پیشتر اشاره شد، در ترجمه اینترنتی به ویژه ترجمه مبتنی بر پیکره‌های پویا، از روش آماری استفاده می‌شود. اما فرآیند کلی نرم-افزارهای ترجمه، همچون ترجمه انسانی بر طبق اصول بنیادی زیر بنا شده‌اند:

الف- رمزگشایی معنای متن زبان مبدأ

ب- رمزگذاری مجدد معنا در زبان مقصد

در پشت این فرآیند به ظاهر ساده عملیات شناختی بسیار سنگینی نهفته است. البته شرکت‌های معتبر جهانی هیچ گاه فرآیند اصلی و عملکردهای محاسباتی خود را آشکار نخواهند کرد. در اینجا مناسب است از محاسبات «بلو»<sup>۲</sup> به عنوان یک روش ارزیابی خودکار برای ترجمه رایانه‌ای یاد کنیم. شرکت گوگل اعلام کرده است که از روش ترجمه آمار-بنیاد که بر پایه تکنیک «بلو» از شرکت آی. بی. ام. است استفاده می‌کند. گیر (۲۰۰۹) از محققین شرکت گوگل، ترجمه ماشینی گوگل را محصولی قابل اعتماد و استاندارد می‌داند و با بیان آمارهایی در این زمینه امیدهای خود را نسبت به آینده این موتور ترجمه ابراز می‌نماید. در حال حاضر گوگل به دلیل دسترسی به حجم زیادی از پیکره‌های پویا از زبان طبیعی توانسته است مجموعه‌ای از بهترین سیستم‌های ترجمه اینترنتی را با استفاده از روش آماری ارائه دهد (دبیرخانه شورای عالی اطلاع رسانی، ۱۳۸۸: ۱۱).

واضح است که فرآیند اولیه در ترجمه رایانه‌ای بسیار ساده و کلی نوشته شده است، اما شرکت‌های معتبر سالانه هزینه‌های بسیار سنگینی بابت دستیابی به واحد اندازه‌گیری و فرمول دقیق‌تر محاسبه معادل ترجمه‌ای متقبل می‌شوند و این‌ها همه در راستای دستیابی به ترجمه‌ای نزدیک به عملکرد انسان است تا نرم‌افزاری نوشته شود که بتواند همانند یک انسان، اطلاعات ورودی را تحلیل و در نهایت خروجی مناسب تحویل دهد.

<sup>۱</sup>. Rule-based

<sup>۲</sup>. Bleu



نانی (۲۰۰۲) از مهندسان شرکت آی.بی.ام. بیان می کند: «از آنجا که ارزیابی ترجمه ماشینی توسط انسان، جامع اما پرهزینه است و ماهها نیاز به بررسی و ارزیابی یک پروژه تحقیقاتی است، وجود یک سیستم محاسباتی خودکار در ارزیابی ترجمه ماشینی بسیار پر اهمیت است». به اعتقاد محققان شرکت آی.بی.ام. روش بلو، دایره ارزیابی ترجمه ماشینی را گسترش می دهد و برای محقق دسترسی بیشتری به اطلاعات تحلیلی و آماری از ترجمه انجام شده فراهم می کند.

### اهمیت مطالعه ترجمه رایانه ای

مطالعه ترجمه رایانه ای اصولاً یک رویکرد نظری به فرآیندهای دخیل در ترجمه است، که حتی می شود آن را همگام با ترجمه کلاسیک دانست. همانطور که در ترجمه انسانی عواملی در بروز خطا در ترجمه نقش دارند، به شکلی دیگر می توان رد پای همان عوامل دخیل را در ترجمه ماشینی نیز مشاهده کرد. در این پژوهش در پی آنیم تا مزایا و معایب کاربرد پیکره های پویا در ترجمه اینترنتی را بررسی کنیم و برون داد این روش را در ترجمه نمونه هایی از متون فارسی به انگلیسی ارزیابی نماییم.

«از دهه ۹۰ میلادی روش های ترجمه در دو دسته متفاوت طبقه بندی می شده اند: قاعده-بنیان<sup>۱</sup> و پیکره-بنیان<sup>۲</sup>. ترجمه در روش قاعده-بنیان بر پایه واژه شناسی و معناشناسی متن مبدأ صورت می گیرد. اما در شیوه پیکره-بنیان از تحلیل زبان شناختی عمیق خودداری می شود و اطلاعات در قالب متون ترجمه شده و واژه های محتوایی ذخیره می شوند. هرچقدر دایره این پیکره زبانی افزایش یابد، یافتن معادل های ترجمه ای نیز راحت تر خواهد بود و ترجمه بهتری انجام خواهد شد» (وانگ جینگ یی، ۲۰۰۶).

از دیدگاه بیکر (۱۹۹۲: ۲۰)، خطاها در ترجمه بیشتر به علت نبود معادل ها میان زبان مبدأ و مقصد بوجود می آیند، هرچند که مترجم ماهر قادر به یافتن و یا بازآفرینی معادل های زبانی است. این نوع تحلیل در مورد نرم افزارهای مترجم نیز مصداق دارد، چراکه نبود پیکره زبانی به روز در زبان مقصد می تواند یک خروجی کاملاً اشتباه به همراه داشته باشد و بخاطر فقدان ارزیابی مترجم

<sup>۱</sup>. rule-based

<sup>۲</sup>. corpus-based

انسانی، این احتمال قوی‌تر و عملکرد ترجمه ماشینی در اینگونه موارد بسیار نامناسب‌تر خواهد بود. هاجنیز و سامرز (۱۹۹۲: ۸۰) در تأکید اهمیت ارزیابی ترجمه، تحلیل‌های مختلف را بیان می‌کنند و به شیوه‌های گوناگون ارزیابی در ترجمه رایانه‌ای می‌پردازند تا نبود تحلیل و ارزیابی انسانی را در ترجمه ماشینی کم‌اثرتر کنند.

حال می‌توان براساس تحلیل بالا عنوان کرد که ترجمه ماشینی در مواقعی که ارزیابی و تحلیل انسانی لازمه یافتن معادل باشد، همچون ترجمه عناصر فرهنگی، رویکرد و خروجی ضعیفی خواهد داشت. در این مقاله تمرکز اصلی ما بر روی عوامل فرهنگی و ورودی‌های فرهنگی زبان فارسی در ترجمه به زبان انگلیسی است و برای مطالعه این عوامل، داده‌هایی از زبان فارسی به نرم‌افزار معیار می‌دهیم تا بتوانیم با تحلیل خروجی‌ها نقاط ضعف اینگونه نرم‌افزارهای ترجمه اینترنتی را بیشتر بشناسیم.

### تحلیل عملکرد نرم‌افزار مترجم اینترنتی گوگل در ترجمه



نرم‌افزار مترجم گوگل به عنوان یک نرم‌افزار معیار در ترجمه اینترنتی شناخته می‌شود. این نرم‌افزار، بنیادی برخط<sup>۱</sup> دارد که امکان استفاده از آن فعلاً تنها در محیط اینترنت فعال است. برخط بودن مترجم گوگل در راستای به‌روز رسانی لحظه‌ای و گسترش بانک اطلاعاتی پیکره زبانی آن می‌باشد. اوچ محقق عالی رتبه گوگل (۲۰۰۶) در مقاله‌ی خود وضعیت کنونی مترجم ماشینی گوگل را بررسی می‌کند. وی با بیان این که در گذشته موتور ترجمه ماشینی<sup>۲</sup> این شرکت در معادل‌های ترجمه‌ای از زبان عربی به انگلیسی با مشکلات بسیاری روبرو بوده است، جمله‌ی ترجمه شده زیر از زبان عربی به انگلیسی را به عنوان مثال می‌آورد.

<sup>۱</sup>. Online-Based Software

<sup>۲</sup>. MT. Engine

"Alpine White new presented tape registered for coffee confirms"  
"laden"

جمله بالا بسیار نامفهوم و ناامیدکننده بود، اما اکنون مترجم آزمایشی جدید گوگل می‌تواند معادل زیر را ترجمه کند.

"The White House confirmed the existence of a new Bin Laden tape"

کاملاً مشخص است که ترجمه گوگل در حال حاضر بسیار پیشرفت کرده است و توانسته است در زمینه‌های جایگزینی دستوری<sup>۱</sup>، لغات عام و اسامی خاص تشخیص بهتری داشته باشد و یک جمله قابل فهم و معتبر در متون خبری<sup>۲</sup> و علمی تولید کند. اما چگونه این نرم‌افزار توانسته است به این سطح از هوش مصنوعی برسد؟ از آنجا که نوشتن چنین سیستم پیچیده‌ای بسیار دشوار است، پاسخ دقیق به این سؤال بسیار مشکل است، اما نظام کلی آن بر پایه دو اصل بنیادی رمزگشایی معنا در زبان مبدأ و رمزگذاری آن در زبان مقصد با استفاده از روش آماری مبتنی بر پیکره‌های پویاست که در بالا به آن اشاره شد.

اما سؤال دیگری که مطرح می‌شود این است که آیا گستره پیشرفت نرم‌افزار گوگل در زمینه‌های خاص‌تر مانند زمینه‌های فرهنگی، ادبی نیز نتیجه‌ای امیدوارکننده دارد؟ پاسخ به این سؤال تنها با مطالعه دقیق‌تر عوامل فرهنگی ممکن می‌شود.

مترجم گوگل اکنون می‌تواند به بیش از ۵۰ زبان زنده دنیا ترجمه کند. در این نرم‌افزار روش اختصاصی ترجمه برای هر زبان نوشته نشده است و ساختار آن بی‌نیاز به محققین بومی<sup>۳</sup> در زبان مبدأ و مقصد است. الزاماً نیازی نیست پژوهشگران شرکت گوگل برای افزودن زبان فارسی آن را به خوبی بدانند، بلکه نیاز به پیکره واژگانی و اصطلاحات زبان فارسی است. گوگل با جستجوهای لحظه‌ای‌اش قادر است میلیون‌ها واژه و جمله را به زبان فارسی ردیابی کند و دایره پیکره‌های زبانی خود را تقویت کند. شاید بتوان امیدوار بود که با گسترش و بهبود این پیکره پویا، در آینده شاهد ترجمه‌های قابل اعتمادتری به زبان فارسی باشیم. در واقع هر چقدر که تعداد صفحات وب

<sup>۱</sup>. grammatical transposition

<sup>۲</sup>. Informative Text

<sup>۳</sup>. Native Researchers

به یک زبان خاص بیشتر باشد، پیکره زبانی حاصل غنی‌تر و فرایند ترجمه اینترنتی موفق‌تر خواهد بود.

این به‌روزرسانی با جستجو در میلیون‌ها صفحه اینترنتی ممکن می‌شود و این نرم‌افزار اینترنتی با الگوریتم‌هایی که برای معادل‌یابی برخط عناصر زبانی همچون کلمات، اصطلاحات و حتی اسامی خاص در خود دارد، می‌تواند به سیستم آماری خود روزانه صدها هزار عنصر زبانی جدید و یا به‌روز شده را اضافه کند. اما آیا اعتماد به پیکره زبانی گردآوری شده از محیط مجازی اینترنت توسط گوگل رویکرد صحیحی است و آیا خروجی‌های آن برحسب دانش زبانی استاندارد از زبان‌ها همچون زبان فارسی خواهد بود؟ اوچ<sup>۱</sup> در پاسخ به این سؤال، اساس توسعه یک ماشین ترجمه آماری برای زبان‌های جدید را در داشتن پیکره متنی دوزبانه<sup>۲</sup> از میلیون‌ها واژه و دو پیکره تک‌زبانه<sup>۳</sup> که هر یک شامل میلیاردها واژه هستند می‌داند. گوگل برای گردآوری چنین پیکره زبانی عظیمی از منابع مختلفی همچون مستندات سازمان ملل که به شش زبان رسمی منتشر می‌شوند استفاده می‌کند.

### ضعف پیکره‌ها و عوامل موثر در خطای ترجمه اینترنتی

همان‌گونه که ترجمه انسانی بر اثر عوامل متعددی ممکن است دچار خطا شود، ترجمه هوشمند رایانه‌ای هم مصون از خطا نیست. این عوامل ممکن است ناشی از تفاوت‌های ساختاری و معنایی دو زبان و یا اختلافات فرهنگی و یا حتی ایدئولوژیک اهالی دو زبان باشد. هاجینز (۲۰۰۷) در بررسی جامع خطاهای ترجمه ماشینی از دیدگاه زبان‌شناسی، مشکلات ترجمه ماشینی را ناشی از تفاوت‌های زبان‌شناختی در سطح واژگانی و ساختاری میان دو زبان می‌داند. وی با ارائه نمونه‌هایی، مشکلات ذاتی (زبان‌شناختی) را در تفاوت‌های ساختاری دو زبان تبیین می‌کند.

(1) Young people like this music.

– Cette musique plaît aux jeunes gens.

<sup>1</sup>. Franz Josef Och

<sup>2</sup>. bilingual text corpus

<sup>3</sup>. monolingual corpora

(2) The boy likes to play tennis.

– Der Junge spielt gern Tennis.

(3) He happened to arrive in time.

– Er ist zufällig zur rechten Zeit angekommen.

(4) Le moment arrivé je serais prêt.

– When the time comes, I shall be ready.

در مثال‌های فوق، تغییر در بیان جملات و ساختارهای مختلف امکان تشخیص کامل بافت و عناصر زبانی مرتبط به آن را برای نرم‌افزارهای ترجمه رایانه‌ای دشوار می‌کند. در مورد نمونه‌هایی مانند شماره ۴، می‌توان گفت راه حلی وجود ندارد و نرم‌افزار مترجم قادر به ترجمه صحیح جمله با حفظ ساختار صحیح نمی‌باشد. هاجینز در تحلیل جملات بالا بیان می‌کند که ترجمه ماشینی با رویکرد آمار-بنیاد و ترجمه مبتنی بر نمونه‌های زبانی می‌تواند خروجی هوشمندانه‌تری داشته باشند.

در بیان جان هاجینز (۱۹۹۲ : ۵۱) عوامل دیگری نیز در خطاهای ترجمه نقش دارند و آن تفاوت‌های غیر زبانشناختی میان دو زبان است که می‌تواند یک نرم‌افزار ترجمه را دچار خطا کند. برای مثال، او یک واقعیت<sup>۱</sup> را در دو جمله و قالب می‌آورد و با مقایسه آن دو جمله تفاوت‌ها را در ترجمه متذکر می‌شود.

The soldiers shot at the women and some of them fell.

The soldiers shot at the women and some of them missed.

قبل از ترجمه‌ی جمله‌های بالا می‌بایست بدانیم ضمیر "them" به چه مرجعی باز می‌گردد. برای مثال در ترجمه به زبان فرانسه از کدام حرف تعریف (ils یا elles) می‌بایست استفاده کنیم. در ترجمه ماشینی با رویکرد زبانشناسی ساختگرا<sup>۲</sup> راه حلی برای مثال بالا وجود ندارد. همچنین در رویکرد مبتنی بر پیکره زبانی نیز نمی‌توان راه حلی یافت که از ترجمه اشتباه و انتخاب ساختار

<sup>۱</sup>. Reality

<sup>۲</sup>. Structural Linguistics

دستوری غلط جلوگیری شود. تنها راه حل<sup>۱</sup> محتمل، استفاده از رویکرد هوش مصنوعی<sup>۱</sup> در ترجمه ماشینی است.<sup>۲</sup>

عوامل خطا در ترجمه همچنین می‌توانند ناشی از تفاوت‌های سبکی<sup>۳</sup> میان دو زبان باشند. همان‌طور که هاجینز (۲۰۰۷، ص ۱۵) با ارائه جملات زیر بر این مطلب تأکید می‌کند، جملات می‌توانند با سبک‌های مختلف در زبان‌های گوناگون ایجاد شوند و این تفاوت‌های سبکی زمانی ملموس‌تر هستند که قصد نویسنده، تأکید بر سبک خاصی در زبان مبدأ باشد.

- [انگلیسی] Advances in technology created new opportunities.
- [ژاپنی] Because technology has advanced, opportunities have been created.
- [ژاپنی] Technology has advanced. There are new opportunities.

بیشتر روش‌های ترجمه ماشینی ویژگی‌های ساختاری زبان مبدأ را به کار می‌گیرند، از این رو در مثال بالا می‌بینیم که چطور یک جمله انگلیسی می‌تواند به سبک‌های مختلف در یک زبان مقصد ترجمه شود. این تنها در محدوده درک مترجم انسانی است که چه سبکی برای محتوای مورد نظر مناسب‌تر می‌باشد. عوامل بروز خطا در ترجمه بسیارند، اما نوعی که ما در این مقاله به آن می‌پردازیم، عوامل فرهنگی موثر در بروز خطا در ترجمه اینترنتی از فارسی به انگلیسی است.

### عوامل خطای فرهنگی در ترجمه اینترنتی فارسی - انگلیسی

ترجمه فعالیتی است که به ناچار حداقل دو زبان و دو فرهنگ مختلف را شامل می‌شود. (توری، ۱۹۷۸: ۲۰۰). اگر این تعریف را بپذیریم اهمیت عوامل فرهنگی در ترجمه را درمی‌یابیم. نایدا (۱۹۶۴: ۱۲۰) به مشکلات مترجم در برگردان عوامل فرهنگی اشاره می‌کند که از جمله آن می‌توان به این موارد اشاره کرد: نبود روشی صحیح جهت یافتن مناسب‌ترین تکنیک انتقال

<sup>۱</sup>. Artificial Intelligence

<sup>۲</sup>. برای اطلاع بیشتر از رویکرد هوش مصنوعی در ترجمه می‌توان به کتاب ویلکس (۱۹۷۲، ص ۳۱) رجوع کرد. او هوش مصنوعی در ترجمه را به نظامی برای تحلیل و تولید معنا جهت یافتن پاسخ دقیق‌تر به مسائل و گره‌های ترجمه‌ای تعریف می‌کند.

<sup>۳</sup>. Stylistic difference

معادل‌های فرهنگی<sup>۱</sup> نهفته در زبان مبدأ و پر کردن شکاف های زبانی<sup>۲</sup> میان دو (یا چند) زبان دخیل در ترجمه.

همچنین نیومارک (۱۹۹۸: ۹۵) در تعریفی از فرهنگ می‌آورد: «فرهنگ همان شیوه و روش زندگی مرتبط به اعضای یک جامعه است که از زبانی خاص به عنوان ابزار برای بیان حالات و رفتار استفاده می‌کنند.» از این رو می‌توان عوامل فرهنگی در ترجمه را به چند عنوان کلی دسته‌بندی کرد که در هر جامعه این عوامل فرهنگی در اکثر اوقات شاخص آن جامعه هستند. به عنوان نمونه، ضرب‌المثل، کنایه، استعاره، تلمیح و اصطلاحات عامیانه از این جمله هستند.

در بررسی این عوامل فرهنگی پل واترز (۱۹۹۹: ۳۵) در کتاب «تحقیق اینترنتی»<sup>۳</sup> خود اشاره می‌کند که می‌شود از یک الگوی تکرارپذیر<sup>۴</sup> مبتنی بر پیکره زبانی اینترنتی برای یافتن خطاهای مرتبط با تفاوت‌های فرهنگی میان دو زبان استفاده کرد و ترجمه ماشینی را به سمتی هدایت کرد که بتواند تناقض‌های فرهنگی موجود را با استفاده از این الگوی منعطف اینترنتی پوشش دهد.

اما پرسش کنونی در ترجمه عوامل فرهنگی آن است که آیا چنین قابلیت‌هایی در یک نرم‌افزار وجود دارد و آیا نرم افزار مترجمی همچون گوگل با وجود این الگوی پیشرفته‌ی جمع‌آوری معادل‌های فرهنگی قادر به ترجمه داده‌های فرهنگی هست؟ جواب این سؤال را می‌توان با آزمون داده‌های فرهنگی در مترجم اینترنتی و مطالعه بازخورد آن بررسی کرد.

در این بررسی، ورودی‌هایی از جنس عناصر فرهنگی در زبان فارسی همچون ضرب‌المثل، اصطلاحات عامیانه، استعاره، کنایه و تلمیح به نرم‌افزار مترجم معیار گوگل داده می‌شود و بازخورد آن با تحلیل هر یک از مثال‌ها بررسی می‌گردد.

### ضرب‌المثل فارسی:

(۱) - کوه به کوه نمی‌رسد، اما آدم به آدم می‌رسد.

(Mountain to Mountain is not, but Adam looks to Adam.) (ت.۱. انگلیسی)

<sup>۱</sup>. Cultural Equivalence

<sup>۲</sup>. Linguistic gaps

<sup>۳</sup>. Internet Research

<sup>۴</sup>. Repetitive Pattern

(۲) - شتر دیدی ندیدی.

(ت.۲.انگلیسی) You have seen camel.

(۳) - آش نخورده و دهن سوخته.

(ت.۳.انگلیسی) Burnt Ash intact and mouth.

ترجمه‌های فوق از ضرب‌المثل‌های فارسی، ترجمه‌ای کاملاً کلمه به کلمه<sup>۱</sup> است و نشان می‌دهد که مترجم گوگل در مواجهه با نمونه‌های فرهنگی مانند ضرب‌المثل‌ها هیچ زمینه تحلیلی (حداقل در زبان فارسی) برای پردازش یک معادل صحیح ندارد. اگر مثال‌های بالا به یک مترجم انسانی ارائه شود، با جابه‌جایی ساختاری و درک مفهوم کلی جمله اصلی، سعی در ایجاد یک معادل معنایی<sup>۲</sup> صحیح می‌کند. همانطور که نایدا و تابر (۱۹۸۲: ۲۰۰) در تعریف معادل پویا<sup>۳</sup> عنوان می‌کنند، «معادل پویا روش ترجمه‌ای است که در آن، مترجم سعی می‌کند معنای متن مبدأ را طوری برگردان کند که همان تاثیر را در زبان مقصد نیز داشته باشد». این تعریف، تأکیدی بر درک انسانی به عنوان مترجم پویا است و نبود این درک در ترجمه اینترنتی مانعی بزرگ محسوب می‌شود.

#### اصطلاحات عامیانه :

(۴) - انگشت به دهان.

(ت.۴.انگلیسی) Finger to mouth.

(۵) - آب خنک خوردن.

(ت.۵.انگلیسی) Eat cold water.

(۶) - از خر شیطان پایین آمدن.

(ت.۶.انگلیسی) Devil of a donkey down.

<sup>۱</sup>. Word for word

<sup>۲</sup>. Semantic Equivalence

<sup>۳</sup>. Dynamic Equivalence



در ترجمه‌های بالا نبود معادل‌های معنایی صحیح از اصلاحات عامیانه‌ی زبان فارسی در زبان انگلیسی مشخص است و نرم‌افزار گوگل طبق ترجمه آماری خود سعی در جستجوی واژگان نزدیک در زبان مبدأ کرده است. در نمونه‌ای دشوارتر و فرهنگی‌تر که مربوط به اصطلاحات رایج زبان فارسی است می‌توان نمونه‌هایی از کنایه‌های رایج فارسی به نرم‌افزار مترجم گوگل ارائه داد.

### کنایه‌ها :

(۷) - شما که روی ما رو سفید کردی !

(ت.۷.انگلیسی) You're white on us!

(۸) آدم هم اینقدر ناخن خشک!

(ت.۸.انگلیسی) Adam also nails so dry!

(۹) - علف زیر پایم سبز شد!

(ت.۹.انگلیسی) Green grass under my feet were!

در کنایه شماره ۷ منظور از سفید کردن صورت شخصی، رنگ زدن آن نیست و مشخص است که در ترجمه اینترنتی فوق انتظار نداریم که فردی در زبان مقصد متوجه معنی جمله ترجمه شده بشود. با توجه به مثال‌های فوق، کنایه‌ها مشکل‌ترین عنصر فرهنگی برای مترجم گوگل محسوب می‌شوند چراکه نبود درک انسانی در یافتن معنای جمله در زبان مبدأ، خروجی کاملاً اشتباه و متفاوت با جمله اصلی بدست می‌دهد. مخاطب زبان مقصد نمی‌تواند به هیچ عنوان متوجه معنای صحیح جمله اصلی بشود و ممکن است از ترجمه اینترنتی انجام شده سوء برداشت کند.

در استعاره‌ها و تشبیه‌ها نیز عملکرد گوگل گرچه بهتر از ضرب‌المثل‌ها است، اما در عین حال خیلی ابتدایی می‌نماید.

**استعاره و تشبیه :**

(۱۰) مثل کبک سرش را زیر برف می‌کند.

(ت.۱۰.انگلیسی) Quebec, like his head is under the snow.

(۱۱) به سفیدی برف می‌ماند.

(ت.۱۱.انگلیسی) The white snow remains.

(۱۲) صورتش به سفیدی برف است.

(ت.۱۲.انگلیسی) Snow White has to face.

در مثال شماره ۱۰ کلمه «کبک» به اشتباه به شهر کبک کانادا ترجمه شده است که خود ناشی از وجود مشکل دیگری در نظام نوشتاری زبان فارسی است که هم برای زبان آموزان فارسی و هم برای مترجم رایانه‌ای تولید اشکال می‌کند (ر. ک. مستند دبیرخانه شورای عالی اطلاع رسانی، ۱۳۸۸: ۱۲). در مجموع نرم‌افزار گوگل، در ترجمه‌ی این نمونه‌های ساده فرهنگی نیز دچار خطا شده است.

آیا گوگل در زبان‌های دیگر هم چنین خروجی نامناسبی از ضرب‌المثل‌ها، استعاره، کنایه و اصطلاحات عامیانه تحویل می‌دهد؟ پاسخ به این سؤال کمی دشوار است، چراکه پیکره زبانی گوگل برای زبان‌ها متفاوت است. اما استفاده از رویکرد آماری در یافتن معادل‌های رایج، گوگل را ناگزیر به ترجمه انعطاف‌ناپذیر و کلمه به کلمه می‌کند و این موردی است که در ترجمه به زبان‌های دیگر نیز مشهود است.

در مقوله ترجمه ماشینی در مطالب فرهنگی دیگر نظیر تلمیحات نیز خروجی‌های مترجم گوگل نامفهوم است و برداشت کلی از ترجمه‌های بالا ممکن است برداشتی منفی و ناامیدکننده از مترجم گوگل باشد. یکی از مهم‌ترین دلایل این ناتوانی در زمینه فرهنگی می‌تواند کوچک بودن پیکره زبانی فارسی و نبود یک بانک اطلاعاتی رسمی و جامع در داخل ایران باشد. از این رو دایره زبانی مترجم گوگل محدود به جمع‌آوری اطلاعات از محیط مجازی اینترنتی و مستندات رسمی سازمان‌های جهانی است، که اطلاعات آن غنای زبانی کافی را ندارد و سطح زبانی آن محدود به

برخی اصطلاحات و جملات رایج در متون فارسی است. این معضل همواره ترجمه اینترنتی به زبان فارسی را در برگردان نمونه‌های فرهنگی با مشکل مواجه کرده است.



همانطور که گفته شد پیکره زبانی مورد استفاده در گوگل یک پیکره پویاست که به طور لحظه‌ای روزآمد می‌شود. این بدان معناست که میلیون‌ها نفر روزانه از نرم افزار مترجم گوگل استفاده می‌کنند و همانطور که در تصویر مشاهده می‌کنید، هر یک از استفاده کنندگان قادرند حتی به تصحیح دایره واژگان و اصطلاحات گوگل اقدام نمایند و با کلیک بر روی «ارائه ترجمه بهتر»، جمله‌ی ترجمه شده را اصلاح کنند. گوگل معتقد است که با این روش می‌تواند دایره پیکره زبانی خود را از هر زبانی گسترش دهد و هر یک از استفاده کنندگان از نرم‌افزار ترجمه گوگل را به یک مترجم بومی و فعال تبدیل کند.

این رویکرد گوگل گرچه موجه به نظر می‌رسد، اما گاهی ترجمه‌های اصلاح شده سهواً و یا حتی عمداً ناصحیح انجام می‌شوند و ذخیره شدن این ترجمه‌ها به عنوان ترجمه صحیح، برخی اوقات اعتبار زبانی کافی را از جملات ترجمه شده می‌گیرد. گاه حتی این موضوع، دستخوش تصمیمات سیاسی یک جامعه و یا قوم می‌شود. ترجمه واژه Gulf به واژه کاملاً مجعول و ناصحیح «خلیج ع.ر.ب.ی» یکی از اقدامات ناشایستی بود که این امکان گوگل در پی داشت.

## نتیجه گیری

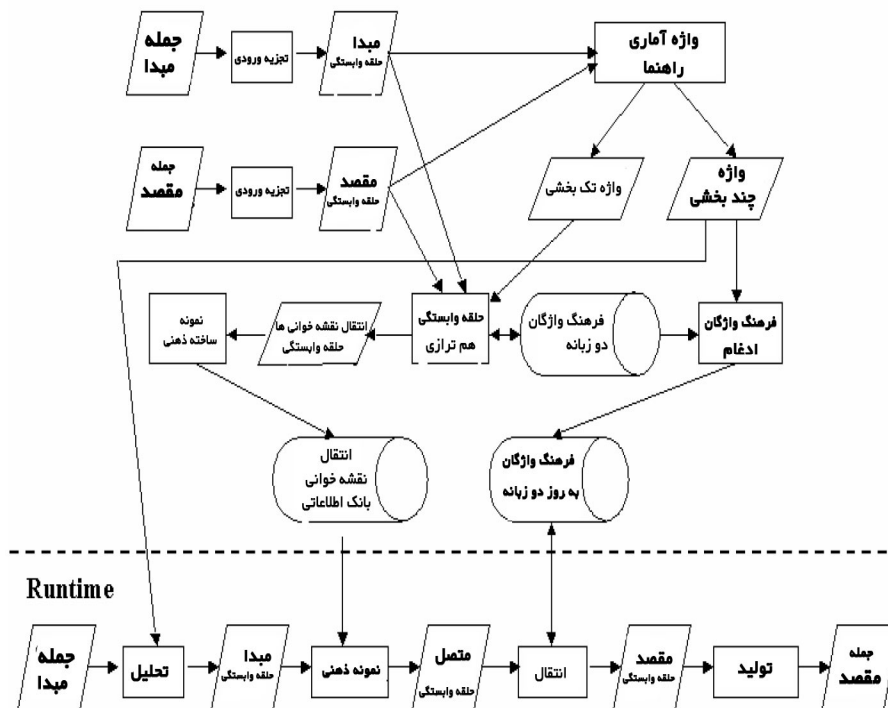
مترجم اینترنتی گوگل مبتنی بر چارچوب آمار-بنیاد است. این چارچوب آماری<sup>۱</sup>، دسترسی به اطلاعات تحلیلی از زبان‌های شریک در ترجمه را به نرم‌افزار مترجم می‌دهد تا در فرآیندی بسیار

<sup>1</sup>. Contribute a better translation

<sup>2</sup>. Statistical framework

پیچیده بتواند به خروجی مناسب برسد. این روش در حالت کلی، متن مبدأ را دریافت می‌کند و با روش «پایین به بالا»<sup>۱</sup> و الگوریتم نمودار تقطیع چندسویه<sup>۲</sup>، خروجی خود را پس از تحلیل و بهینه‌سازی نشان می‌دهد.

عملیات بالا به صورت کلی و نظری بیان شده است اما در حالت اجرا، رشته ورودی دریافت می‌شود و پس از تبدیل به نمودار وابستگی در میان بانک اطلاعاتی حجیم زبانی، معادل مناسب ترجمه جستجو می‌شود.



نمودار ۱. مترجم ماشینی آماری (تحقیق میکروسافت ۲۰۰۷)

نمودار شماره ۱ شمایی جامع از فرآیند انجام ترجمه ماشینی توسط نرم‌افزارهای مترجم اینترنتی گوگل و یا میکروسافت است. این حلقه بر اساس حلقه وابستگی<sup>۳</sup> چیده شده است که

<sup>۱</sup> Bottom-Up

<sup>۲</sup> Multi-path chart parsing algorithm

<sup>۳</sup> Dependency loop

مانند روش اندازه گیری «بلو» در پردازش و آزمون درستی اطلاعات ترجمه شده موثر است. اگر تعریف هیرو یوکی کاجی (۱۹۸۷) را از حلقه وابستگی بپذیریم، ساختاری است واسطه‌ای میان دو زبان. در این تعریف، ژرف‌ساخت میانی در دو زبان، نماد مناسبی برای حلقه وابستگی به شمار می‌رود و می‌تواند در دو جهت مبدأ و مقصد تعریف شود.

این تعاریف کلی از فرآیندهای آماری ترجمه ماشینی، دیدگاهی سطحی از فرآیند ترجمه ماشینی گوگل به دست می‌دهد که بر خلاف نظر عام، ترجمه‌ای کلمه به کلمه نمی‌باشد و در پس آن، پردازش معنا در دو زبان و تحلیل آماری پیچیده رخ می‌دهد.

اطلاعات فنی و تأیید شده از نرم‌افزارهای مترجم، بسیار کم هستند و این موضوع نتیجه‌گیری را دشوار می‌کند. اما با مطالعه این نرم‌افزارها و فرآیندهای کلی آن می‌توان پیش‌بینی کرد که اینگونه نرم‌افزارها روند رو به رشدی خواهند داشت و به خاطر استفاده از محیط اینترنت، پیکره زبانی پویا و دائماً در حال ارتقا خواهند داشت. از این رو پیکره‌های زبانی پویا در ترجمه اینترنتی به عنوان راه حلی در دسترس مطرح می‌شوند. اینگونه پیکره‌های زبانی برخط، می‌توانند اطلاعات خود را از دو طریق کسب کنند و به عنوان منبع اصلی خود در ترجمه ماشینی برخط استفاده کنند:

الف. استخراج پویا و خودکار اطلاعات

ب. یا دریافت اطلاعات از طریق کاربران.

اطلاعات موجود در محیط اینترنت به صورت لحظه‌ای در حال تغییر و افزایش است. از این رو شاید نتوان این منبع زبانی را نادیده گرفت و همانطور که تصور می‌شود نرم‌افزارهایی مانند مترجم گوگل این نکته ظریف را نادیده نگرفته‌اند و از این منبع بسیار گسترده استفاده کرده‌اند. با آنکه استفاده از پیکره زبانی برخط فواید بسیار زیادی برای نرم‌افزارهای مترجم دارد و از آن جمله می‌توان به به‌روز بودن دایره اطلاعات زبانی آن‌ها اشاره کرد اما موضوع نگران کننده در استفاده از پیکره زبانی برخط، غیر رسمی بودن و اطلاعات بعضاً غیر استاندارد آن می‌باشد که منجر به رواج لغات و صورت‌های زبانی نادرست در ترجمه می‌شود.

پرسشی که هنوز باقی می‌ماند این است که تا چقدر این اطلاعات گردآوری شده از محیط اینترنت و بانک زبانی گوگل می‌تواند غنای زبانی کافی برای ایجاد یک پیکره زبانی معیار را داشته باشد؟

ما چه بخواهیم و چه نخواهیم، چه این کار را صحیح بدانیم یا غیر علمی، کاربران زیادی در کشور ما از ترجمه اینترنتی برای تسریع در فرایند دریافت اطلاعات استفاده می‌کنند و متون مختلف را به ویژه از انگلیسی با این روش به فارسی ترجمه می‌کنند. تعداد این افراد هم روز به روز در حال افزایش است. به این گروه باید علاقه‌مندان به زبان فارسی در دیگر کشورها را نیز افزود. پس شاید جامعه زبان‌شناسان و متخصصین زبان فارسی و رایانه باید در این راستا احساس وظیفه بیشتری بکنند و خدماتی را در این جهت برای پاسداشت زبان فارسی انجام دهند. راه حلی که برای بهبود این امر به نظر می‌رسد تنظیم ساز و کاری برای تهیه و تولید پیکره‌های معتبر از زبان فارسی و ارائه آن به شرکت‌های ارائه دهنده خدمات ترجمه اینترنتی است تا به جای استفاده از منابع متفرقه و غیر معتبر از داده‌های کارشناسی شده و تهیه شده به روش علمی برخوردار گردند. این کار در حال حاضر به طور بسیار محدود و پراکنده و گاه موازی در قالب طرح‌های پژوهشی کوچک در حال انجام است؛ اما این مهم مستلزم برنامه‌ریزی دقیق و گسترده و صرف هزینه متناسب و استفاده از نیروهای مجرب است. شایسته است فرهنگستان زبان و ادب فارسی و شورای عالی اطلاع‌رسانی به این امر توجه ویژه مبذول دارند.

### منابع:

- خداپرستی، فرج الله. (۱۳۸۰) ابداع و طراحی سیستم رایانه‌ای جهت ترجمه متون علمی زبان انگلیسی به زبان فارسی. مرکز اطلاعات و مدارک علمی ایران.
- دبیرخانه شورای عالی اطلاع‌رسانی جمهوری اسلامی ایران. (۱۳۸۸). جمع‌آوری اطلاعات چالش‌ها و روش‌های ترجمه ماشینی زبان انگلیسی به فارسی و بالعکس. مستند شماره ۱۹۰/۲۵۳۷/۱/۱. تهران: دانشگاه علم و صنعت ایران.

- Aue, Anthony. et. al. (2004). Statistical Machine Translation Using Labeled Semantic Dependency Graphs, *Microsoft Research: Proceedings of TMI 2004*.
- Baker, M. (1992), *In Other Words*, London: Rutledge.
- Gear, David. (2009), Google Official research blog
- Granger, S. et. al.(2008), *Corpus-based Approach to Contrastive Linguistics and Translation Studies*. Amsterdam: Rodopi.
- Huston, S. (2006) Corpus Linguistics, In Keith Brown, *Encyclopedia of Language and Linguistics*. Elsevier (2006) USA.
- Hutchins, John (2007), Machine translation: problems and issues (panel at conference, 13 December 2007. Chelyabinsk, Russia.).
- Hutchins, W. John; and Harold L. Somers (1992). *An Introduction to Machine Translation*. London: Academic Press.
- Jing-yi, Wang. (2006) A Discussion on the Promotion of Machine Translation by Multi-engine Method Based on Dynamic Language Corpus. *US-China Foreign Language*. ISSN1539-8080, USA
- Kaji, Hiroyuki. (1987), A Japanese-to- English Machine Translation System Based on Semantics. *Proceedings of Machine Translation Summit*, pp.55-60, 1987.
- Meyer, C. (2002), *English Corpus Linguistics: An Introduction*. Cambridge: Cambridge University Press.
- Nani (2002) Annual Meeting of the Association for Computational Linguistics. July 6-12. 2002, Philadelphia, PA, USA
- Newmark, P. (1998) *A Textbook of Translation*. New York: Prentice Hall.
- Nida, Eugene A. and C.R. Taber (1969 / 1982) *The Theory and Practice of Translation*, Leiden: E. J. Brill.
- Och, F (2006), Statistical machine translation live. Retrieved from: Google Research blog, April 28, 2006.

Och, F. and Ney, H. 2002. Discriminative Training and Maximum Entropy Models for Statistical Machine Translation. *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*.

Richardson, S. et. al. (2001) Achieving commercial-quality translation with example-based methods. *Proceedings of MT Summit VIII*: 293-298.

Silberman, S. (2004), Machine Translation: AI Methods for Translating from One Language to Another. Retrieved August 5, 2011, from [www.aai.org/AITopics/pmwiki/pmwiki.php/AITopics/MachineTranslation](http://www.aai.org/AITopics/pmwiki/pmwiki.php/AITopics/MachineTranslation).

Somers, Harold. (2001) Three perspectives on MT in the classroom. *MT Summit VIII Workshop on Teaching Machine Translation* (Santiago de Compostela Spain), 25-29.

Wilks, YA (1972). *Grammar, Meaning, and the Machine Analysis of Language*. London: Rutledge & Kegan Paul.