



وزارت فرهنگ و آموزش عالی
پژوهشگاه علوم انسانی و مطالعات فرهنگی

پایان نامه

برای دریافت درجه کارشناسی ارشد

در زبانشناسی همگانی

موضوع

فرهنگ لغت در سیستم انتقالی ترجمه ماشینی

استاد راهنما

دکتر مصطفی عاصی

استاد مشاور

دکتر شهین نعمت زاده

نگارش

الهه دفتری نژاد

نیمسال دوم - سال تحصیلی ۷۸-۱۳۷۷

فهرست مطالب

صفحه	عنوان
الف.....	قدردانی
ج.....	چکیده
د.....	پیشگفتار
ط.....	اختصارات

فصل اول: مقدمه و معرفی سیستم انتقالی

۱.....	۱-۱ اصول نظری
۳.....	۲-۱ ترجمه ماشینی
۴.....	۳-۱ تاریخچه ترجمه ماشینی در جهان
۶.....	۴-۱ تاریخچه مطالعات فرهنگ لغت در ترجمه ماشینی
۹.....	۵-۱ سابقه تحقیق در زبان فارسی
۱۲.....	۶-۱ انواع سیستم‌های ترجمه ماشینی
۱۳.....	۷-۱ بخشهای مختلف سیستم انتقالی و روش کارکرد آن
۱۴.....	۱-۷-۱ بخش تجزیه
۱۵.....	۲-۷-۱ بخش انتقال
۱۵.....	۳-۷-۱ بخش تولید
۱۶.....	۸-۱ کارایی سیستم انتقالی و پروژه‌های انجام شده در جهان
۱۹.....	۹-۱ هدف این طرح

فصل دوم: بخش واژگان در سیستم انتقالی

۲۰	۱-۲ مقدمه
۲۰	۲-۲ انواع واژگان در سیستم انتقالی
۲۱	۱-۲-۲ واژگان تحلیلی
۲۲	۲-۲-۲ واژگان تولیدی
۲۲	۳-۲-۳ واژگان انتقالی
۲۳	۳-۲ محتوای اطلاعاتی مدخلهای واژگانی تک زبانه
۲۵	۱-۳-۲ مقوله واژگانی
۲۶	۲-۳-۲ مکمل ها
۲۶	۳-۳-۲ زیر مقوله ها
۲۸	۱-۳-۳-۲ دانش نحوی
۲۹	۲-۳-۳-۲ دانش معنایی
۳۱	۴-۳-۲ اطلاعات متفرقه و ارائه اسناد و مدارک
۳۱	۴-۳-۲ اطلاعات واژگان انتقالی (دو زبانه)
۳۵	۱-۴-۲ ابهام
۳۶	۱-۴-۲ هم نویسی
۳۷	۲-۴-۲ ناهمگنی های واژگانی و ساختاری
۳۸	۳-۴-۲ اصطلاحات
۴۰	۴-۴-۲ با هم آبی ها

فصل سوم: بخش های ساختواژی (تحلیل گرواژگانی) و قاعده مندی های زبان

۴۳	۱-۳ مقدمه
۴۴	۱-۱-۳ بخش ساختواژی تحلیلی
۴۵	۲-۱-۳ بخش ساختواژی تولیدی

۴۵بخش ساختاروی انتقالی. ۳-۱-۳
۴۶قاعده مندی های زبان. ۲-۳
۴۷قواعد واژه های تصریفی ۱-۲-۳
۵۰قواعد واژه های اشتقاقی ۲-۲-۳
۵۲قواعد واژه های مرکب ۳-۲-۳
۵۴واژه های تخصصی. ۳-۳
۵۶تأثیر متقابل فرهنگ و دستور زبان. ۴-۳

فصل چهارم: ارائه مدل فرهنگ انتقالی برای زبان فارسی

۵۸ ۱-۴ مقدمه
۵۸ ۲-۴ منابع و ساختار واژگانی سیستم یورترا (Eurotra)
۶۵ ۳-۴ روش کار
۶۶ ۱-۳-۴ انتخاب مدخل های نمونه و طراحی جدول ها
۶۷ ۱-۱-۳-۴ مقوله واژگانی اسم
۷۴ ۲-۱-۳-۴ مقوله واژگانی فعل
۸۰ ۳-۱-۳-۴ مقوله واژگانی صفت
۸۴ ۴-۱-۳-۴ مقوله واژگانی قید
۸۵ ۵-۱-۳-۴ مقوله واژگانی ضمیر
۹۱ ۶-۱-۳-۴ مقوله واژگانی حرف اضافه
۹۲ ۷-۱-۳-۴ مقوله واژگانی حرف ربط
۹۳ ۸-۱-۳-۴ مقوله واژگانی حرف تعریف
۹۴ ۹-۱-۳-۴ جدول های اصطلاحات و باهم آبی ها
۹۵ ۱۰-۱-۳-۴ وندهای تصریفی و اشتقاقی
۹۷ ۱۱-۱-۳-۴ بخش انتقال

۴-۴ مسيرهای معادل يابی ورودی‌های نمونه ۱۰۲

فصل پنجم: نتیجه‌گیری و پیشنهادات

۱-۵ انتخاب سیستم و محدود کردن حیطه تحقیق ۱۱۰

۲-۵ اهمیت تهیه فرهنگ و محدود کردن موضوعی ۱۱۰

۳-۵ راههای گسترش و تکمیل سیستم ۱۱۱

۴-۵ نتیجه حاصله از مدل سازی فرهنگ ۱۱۲

۵-۵ اهمیت مباحث زبانی در بحث‌های زبانشناسی رایانه‌ای و ترجمه ماشینی ۱۱۲

واژه نامه انگلیسی - فارسی ۱۱۴

واژه نامه فارسی - انگلیسی ۱۲۲

کتابنامه ۱۳۱

قدردانی

سپاس از پرودگار متعال که هیچ حرکتی جز به حول قوه لم یزلی اش باقی نخواهد ماند و سپاس از هر آنکه به من راه و رسم بهتر اندیشیدن و بهتر زیستن را آموخت. شایسته است در آغاز این رساله و در پایان دوره سه ساله تحصیلات کارشناسی ارشد از تمامی استادان بسیار ارجمندی که در این مدت در خدمتشان بودم و در کلاسهای پر بارشان بسیار آموختم، به خاطر زحمات و محبتهای بی دریغشان صمیمانه سپاسگزاری کنم. حقیقتاً سپاس از نکته آموزشی‌های بسیار و تأثیرگذاری فراوان این بزرگواران هرگز در قالب عبارات نمی‌گنجد و نیز سپاس بی پایان من در مقابل زحمات و تلاشهای این استادان گرانقدر بسیار ناچیز است.

از جناب آقای دکتر مصطفی عاصی که راهنمایی این پایان نامه را بر عهده داشتند و در این مدت من بیش از پیش از دانش و تجربه ایشان بهره گرفتم بسیار سپاسگزارم. ایشان همواره با سعت صدر زاید الوصفی به سئوالات من پاسخ گفتند و راهنمایی‌های بسیار ارزنده‌ای در تمامی زمینه‌ها خصوصاً در انتخاب موضوع، تهیه منابع و پیاده سازی طرح ارائه نمودند که بدون آن انجام این رساله هرگز میسر نبود، همچنین رساله حاضر را با دقت بسیار مورد بررسی قرار دادند و مجموعه‌ای از نارسائیهای کار را متذکر شدند. شایسته است بار دیگر از تمام زحمات ایشان تشکر کنم.

از سرکار خانم دکتر شهین نعمت زاده که مشاوره پایان نامه حاضر را بر عهده گرفتند و علی رغم ضیق وقت با دقت و نکته سنجی بسیاری که خاص ایشان است پایان نامه حاضر را مطالعه نمودند و نظرات سودمند و پر ارزشی در جهت ارائه بهتر و روانی مطلب بیان داشتند بسیار سپاسگزارم.

از کلیه مسئولین محترم انجمن حکمت و فلسفه، پژوهشگاه علوم انسانی و مطالعات فرهنگی، خصوصاً مسئولین محترم کتابخانه پژوهشگاه علوم انسانی که کمکهای فراوانی به من نمودند بی نهایت سپاسگزارم.

در خاتمه لازم می‌دانم از پدر، مادر و همسر مهربانم که از آغاز راه مشوق من بودند و

نیز به لطف کمک و همیاری ایشان طی این مسیر بر من ممکن گشت قدردانی و تشکر
نمایم.

چکیده

پایان نامه حاضر با عنوان "فرهنگ لغت در سیستم انتقالی ترجمه ماشینی" در ۱۳۳ صفحه ارائه شده است که شامل ۵ فصل و نیز بخش‌های پیشگفتار، اختصارات، واژه نامه‌های انگلیسی - فارسی و فارسی - انگلیسی و کتابنامه است.

پس از پیشگفتار فصل اول با عنوان "مقدمه و معرفی سیستم انتقالی" قرار دارد. در این فصل پس از بیان مباحث نظری پیرامون سیستم انتقالی، اشاره مختصری به جایگاه ترجمه ماشینی در زبان‌شناسی شده است سپس تاریخچه کوتاهی از ترجمه ماشینی و نیز مطالعات فرهنگ لغت در سیستم‌های ترجمه ماشینی در جهان آمده است و نیز به همین نحو، بخشی از سابقه نسبتاً کوتاه تحقیقات ترجمه ماشینی در زبان فارسی ارائه شده است. در این فصل انواع سیستم‌های ترجمه ماشینی معرفی شده و سپس سیستم انتقالی که سیستم مورد بحث در این رساله می‌باشد توصیف شده است. در پایان این فصل تعداد بسیار محدودی از پروژه‌هایی که با استفاده از سیستم انتقالی طراحی شده و نیز در بخش فرهنگ لغت دارای مدلی مشابه طرح حاضر - مدل انتقالی - می‌باشند معرفی شده‌اند.

در فصل‌های دوم و سوم با عنوانهای "بخش واژگان در سیستم انتقالی" و "بخش‌های ساختارهای (تحلیل گر واژگانی) و قاعده مندی‌های زبان"، به توصیف جزء به جزء فرهنگ لغت در سیستم انتقالی پرداخته شده است. در فصل دوم پس از توصیف کلی بخش‌های واژگانی، اطلاعات مورد نیاز برای هر مدخل توضیح داده شده است. پر واضح است که این موارد که از منابع مختلفی گردآوری شده‌اند بر حسب دقت مورد نظر و نوع زبان می‌تواند به فهرست کامل‌تری تبدیل شود. سپس بحثی پیرامون مسائلی چون ابهام، هم نویسی، ناهمگنی‌های واژگانی، اصطلاحات و با هم آبی‌ها ارائه شده است. از آنجائی که تک تک این موارد در زمره زمینه‌های تحقیقی این رشته محسوب می‌شوند لذا این بحث نسبتاً محدود و مختصر است. در فصل سوم پس از توصیف کلی بخش تحلیل گر واژگانی، فرایندهای واژگانی و قاعده مندی‌های زبان معرفی شده است. در

توصیف این فرآیندها که عبارتند از؛ تصریف، اشتقاق منظم و ترکیب منظم، نحوه برخورد سیستم مذکور با هر یک از این موارد و نیز نحوه پردازش هر مورد توضیح داده شده است. در پایان این فصل بحثی پیرامون واژه‌های تخصصی و نیز تأثیر متقابل فرهنگ لغت و نحو سیستم بر یکدیگر قرار دارد.

در فصل چهارم با عنوان "ارائه مدل فرهنگ انتقالی برای زبان فارسی"، نخست یک بررسی اجمالی از ویژگی‌های واژگانی سیستم یورترا^(۱) به عمل آمده است. سیستم یورترا، سیستم ترجمه ماشینی اتحادیه اروپاست که در این طرح به عنوان مدل قرار گرفته است. نوع این سیستم انتقالی بوده و نیز طرح بخش واژگانی آن مشابه طرح حاضر است. البته با توجه به آنکه در اجرای پروژه یورترا کشورهای مختلفی شرکت دارند، اختلافاتی در طرح بخش واژگانی آن مشاهده می‌شود. مثلاً در بخش‌هایی از آن کلیه واژه‌ها و حتی برخی از عبارات به صورت قالبی وارد شده است و پردازشگر واژگانی وجود ندارد لیکن همین واژه‌ها و اصطلاحات در بخش‌های دیگری از این طرح ممکن است به نسبت‌های مختلفی - از جدا کردن وندهای تصریفی تا تجزیه وندهای اشتقاقی منظم و ترکیب منظم - دارای ذخیره اطلاعاتی لازم برای انواع پردازش‌های واژگانی باشند. در ادامه فصل چهارم روش کار (بخش زبانی) توصیف شده است و نیز طرح جدول‌ها که در واقع پیاده سازی تقسیم بندی‌های زبانی از طریق نرم افزار اکسس^(۲) می‌باشد، ارائه شده است. در پایان این فصل نمونه‌هایی از ورودی‌های فرضی و مسیرهای معادل‌یابی آنها آورده شده است. این نمونه‌ها مهم‌ترین قابلیت‌های این مدل را نشان می‌دهند.

فصل پنجم با عنوان "نتیجه‌گیری و پیشنهادات" شامل نتایج و پیشنهادات شخصی نگارنده است و پس از آن نیز واژه‌نامه‌های انگلیسی - فارسی و فارسی - انگلیسی و کتابنامه قرار دارد.

1- Eurotra

2- Access

پیشگفتار

پایان نامه حاضر با موضوع فرهنگ لغت در ترجمه ماشینی به شاخه میان رشته‌ای زبانشناسی رایانه‌ای مربوط می‌گردد. به طور کلی ترجمه ماشینی به عنوان یکی از انواع پردازش‌های زبانی که از طریق رایانه صورت می‌پذیرد، علاوه بر آنکه بخشی از رشته فوق‌الذکر را به وجود می‌آورد، با رشته هوش مصنوعی نیز مربوط می‌گردد. این بحث حیطه‌ای کاربردی را برای رشته‌های زبانشناسی، رایانه، ترجمه، اصطلاح‌شناسی و هوش مصنوعی ایجاد می‌کند. اگر با ساده‌ترین تعریف زبانشناسی را مطالعه علمی زبان به عنوان سیستم ارتباطی انسان در نظر بگیریم، زبانشناسی رایانه‌ای شاخه جدیدی از زبانشناسی خواهد بود که حاصل تلفیق رشته‌های زبانشناسی و رایانه است. این شاخه میان رشته‌ای رویکردی در زبانشناسی است که با بهره‌گیری از رایانه، روش‌های ریاضی گونه را در امر مطالعه زبان وارد می‌کند. زبانشناسی رایانه‌ای شامل تحلیل داده‌های زبانی، ترجمه ماشینی، تولید الکترونیکی گفتار مصنوعی (تحلیل گفتار) و تشخیص خودکار گفتار انسان می‌باشد. بنابراین ترجمه ماشینی از طریق شاخه میان رشته‌ای زبانشناسی رایانه‌ای با رشته زبانشناسی در ارتباط است (DLTAL, 1985).

بحث در زمینه فرهنگ لغت در سیستم‌های ترجمه ماشینی حیطه‌ای تخصصی‌تر است که علاوه بر زمینه‌های فوق‌الذکر به دانش فرهنگ نگاری؛ یعنی، مطالعه اقلام واژگانی یک زبان و بررسی معانی و روابط آنها و نیز تغییرات فرم و معانی آنها در طول زمان نیازمند است. فرهنگ‌های لغت در سیستم‌های ترجمه ماشینی، فرهنگ‌هایی درونی می‌باشند که در واقع پایگاه‌های اطلاعاتی عظیمی را به وجود می‌آورند. این پایگاه‌ها بزرگترین بخش هر سیستم هستند. حجم و کیفیت این فرهنگ‌ها بیش از هر بخش دیگری بر حیطه، پوشش و کیفیت ترجمه تولید شده تأثیر می‌گذارد. اگر بخواهیم این فرهنگ‌ها را با فرهنگ‌های لغت معمولی مقایسه کنیم، می‌توانیم بگوئیم که حداقل اطلاعات موجود در فرهنگ‌های رایانه‌ای می‌بایست بیش از اطلاعات موجود در فرهنگ‌های معمولی برای هر مدخل باشد. بنابراین این بخش‌هایی از اطلاعات ذخیره شده

در این دو نوع فرهنگ کاملاً مشابه است. بحث فرهنگ لغت در ترجمه ماشینی مبحث گسترده‌ای است که می‌تواند به مطالعه و بررسی انواع گوناگونی از این نوع فرهنگ‌ها بپردازد. این فرهنگ‌ها به لحاظ تفاوت در دیدگاه‌های نظری سیستم، نوع سیستم، نرم‌افزار به کار رفته در تهیه بانک داده‌ها و زبان برنامه نویسی که در تهیه آنها به کار رفته است از یکدیگر متفاوت می‌شوند. مباحث نظری در مورد فرهنگ‌های رایانه‌ای از جمله زمینه‌هایی است که قدمت چندانی نداشته و نیاز به تحقیقات بیشتری دارد. می‌توان گفت که به جز برخی از گرایش‌های جدید در این زمینه، هیچ یک از مدل‌های موجود از پایگاه نظری مستقل و واحدی برخوردار نمی‌باشد.

سیر تحولی فرهنگ‌های رایانه‌ای، به طور کلی متأثر از سیر تحولی موجود در سیستم‌های ترجمه ماشینی بوده است. اگر سیستم‌های ترجمه ماشینی را به ترتیب زمانی و بر حسب یک تقسیم بندی سنتی به دو دسته اصلی؛ یعنی، سیستم‌های مستقیم و غیر مستقیم تقسیم کنیم آنگاه فرهنگ‌های رایانه‌ای همگام با این خط سیر دارای یک مسیر تحولی و تکاملی مشخص و قابل بررسی خواهند بود. این سیر تکاملی در فصل اول این پایان نامه به طور مختصر بررسی شده است. تهیه فرهنگ‌های رایانه‌ای و نیز سیستم‌های ترجمه ماشینی و یا هرگونه سیستم پردازشگر زبان، تنها از طریق پروژه‌های متعدد و گروه‌های محققین در زمینه‌های تخصصی و گرایشهای داخلی رشته‌های مختلف قابل انجام است. نقش زبان‌شناسان در چنین طرح‌هایی مشخص کردن ویژگی‌ها و روندهای زبانی طرح است. نیاز به همکاری توأم گروه‌های زبان‌شناس با گرایش‌های معنی‌شناسی، فرهنگ نگاری، آواشناسی، دستور زبان و سایر زمینه‌ها، تأکیدی بر ضرورت نقش این گروه‌ها در چنین پروژه‌هایی است. این گروه‌ها چهار چوب اصلی چنین طرح‌هایی را مشخص می‌کنند. همکاری توأم گروه‌های محققین در زمینه‌ها و گرایش‌های گوناگون زبانی نکته‌ای است که در پروژه‌های مختلف ترجمه ماشینی در جهان مشاهده می‌شود.

به عنوان مثال بخش مطالعات و ازگانی پروژه اگلس^(۱)، حاصل همکاری یک تیم ۲۵ نفره از متخصصین این رشته می‌باشد. نکته دیگری که ذکر آن حائز اهمیت می‌باشد آن است که پرداختن به مبحث ترجمه ماشینی به عنوان موضوعی که به یک بحث میان رشته‌ای مربوط می‌گردد، آگاهی از هر دو زمینه تخصصی را برای محققین آن الزامی می‌سازد. هاجنیز در کتاب "ترجمه ماشینی: گذشته، حال، آینده" ضمن اشاره بدین نکته، بیان می‌کند که آگاهی از حداقل یک زبان برنامه‌نویسی برای زبان‌شناسان این رشته ضروری می‌باشد (Hutchins, 1986).

انگیزه نگارنده در انتخاب موضوع این تحقیق، برداشتن گامی هر چند ناچیز در جهت انجام طرح ترجمه ماشینی برای زبان فارسی است. به نظر می‌رسد که ضرورت تحقیق در زمینه ترجمه ماشینی موضوعی است که نیاز به توضیح فراوان ندارد و این در حالی است که نقش زبان‌شناسی در جهت ایفای رسالت خویش و نیز به کارگیری دست‌آوردهای خود در جهت برآورد ضرورت‌های تحقیقی از طریق چنین پروژه‌هایی - و نیز سایر تحقیق‌های کاربردی در دیگر زمینه‌های بین رشته‌ای - بهتر قابل اجراست. این مهم خود نقش به‌سزایی در شناساندن و معرفی این رشته و کاربردهای آن دارد. با توجه به سابقه تحقیق در زمینه ترجمه ماشینی برای زبان فارسی - که البته محدود به پایان نامه‌هایی است که در مقاطع تحصیلی مختلف در چند دانشگاه انجام شده است - مشخص می‌شود که دانشجویان سایر رشته‌ها خصوصاً رشته کامپیوتر توجه و علاقه خاصی به این مبحث مبذول داشته‌اند لیکن برطرف کردن موانع متعددی که در مسیر چنین پروژه‌هایی قرار دارد ضرورت انجام تحقیق‌های مشابهی را از سوی متخصصین، صاحب نظران و نیز دانشجویان زبان‌شناسی، جهت تهیه و گردآوری منابع و اطلاعات و نیز ایجاد سابقه‌ای در این زمینه مطرح می‌سازد.

طرح فرهنگ لغت در پروژه حاضر اگرچه پس از آشنایی نگارنده با نرم‌افزار اکسس

1- Eagles

و نیز آشنایی محدود با زبان برنامه نویسی V.B^(۱) ارائه شده است لیکن به هر حال طراحی بانکهای اطلاعاتی، برنامه نویسی و نیز تهیه الگوریتمهای نحوی و واژگانی چنین پروژه‌هایی برعهده متخصصین رایانه است (Hutchins, 1986). بنابراین طراحی بانک اطلاعاتی این پروژه که جهت تکمیل مدل سازی آن ارائه شده است به هر حال خالی از هرگونه اشکال نخواهد بود.

در پایان ذکر این نکته ضروری است که با توجه به تلاش‌ها، تحقیق‌های گسترده و پروژه‌های مختلفی که درمورد سایر زبان‌ها برای تهیه سیستم‌های ترجمه ماشینی و همچنین سایر پردازش‌های زبانی انجام پذیرفته است، به نظر می‌رسد که زبان فارسی با توجه به میراث فرهنگی، هنری، علمی و قدمت و سابقه تاریخی و نیز سایر زمینه‌های ارزشمند آن، نیاز به مطالعه و بررسی بیشتری در این زمینه دارد. امید است که طرح حاضر گامی هرچند کوچک در این راستا باشد.

1- Visual Baise

اختصارات

واژه	علامت اختصاری
adjective	adj
adverb	adv
affix	aff
agreement	agr
argument	arg
auxiliary	aux
category	cat
causative	caus
classifying	class
collective	coll
collocation	col
comparative	compar
complimentizer	comp
concessive	concess
conditional	cond
conjunction	conj
context	C
countable	count
derivation	der
derivational	deriv
determiner	det
diacritical	diacr
direct object	DO
discontinuous	discon
ditransitive	ditr
documentation	doc
English Monolingual Dictionary	EMD
English-Persian Dictionary	EPD
equivalent	eq
female	F
genitive	gen

واژه	علامت اختصاری
identification	ID
imperative	imper
indicative	indic
indirect object	IO
inflection	inf
inflectional	inflec
intransitive	intr
interrogative	introg
irregular	irreg
linking	link
location	loc
male	M
miscellaneous	mis
morpheme	morph
morphological component	mc
morphology	morphol
negative	neg
noun	n
noun phrase	np
number	no
object	obj
participle	part
perfect	perf
Persian Monolingual Dictionary	PMD
plural	plu
position	pos
possesive	poss
prefix	pref
preposition	p
present	pres
pronoun	pro
qualitative	qual

واژه	علامت اختصاری
reading number	rno
reflexive	refl
regular	reg
restricted semantic feature	rsf
semantic	SEM
singular	sing
subject	subj
subjunctive	subjunc
subordinating conjunctions	subconj
suffix	suf
superlative	superl
syntactic	SYN
transitive	tr
verb	v

فصل اول

مقدمه و معرفی سیستم انتقالی

با توجه به اهمیت ارتباطات در عصر حاضر و نیز حجم وسیع اطلاعاتی که هم روزه در میان کشورها مبادله می شود، نیاز و اهمیت زبان به عنوان وسیله ارتباطی و ترجمه به عنوان وسیله عبور از سد موانع ارتباطی مشخص تر می شود. نیاز شرکتهای تجاری مختلف به ترجمه متن های مشابه میان زبانهای بسیار متنوع، نیاز مترجمین به افزایش سرعت کار خود از طریق کاهش یکنواختی های کار ترجمه، احساس نیاز علوم مختلف از جمله علوم انسانی به کاربرد رایانه و علاقه به بهره گیری از سرعت و دقت فراوان آن در زمینه های مختلف و مسائلی از این دست به همراه انگاره دیرپای ترجمه ماشینی که حتی سابقه آن را تا قبل از اختراع رایانه می توان جستجو کرد، همه و همه توجه مناسبی در توجه روز افزون به ترجمه ماشینی و مسائل مربوط به آن، از جمله، فرهنگ لغت می باشد.

فرهنگ لغت تقریباً بخش مرکزی هر سیستم است. سیستم های مختلف بر حسب نوع، به فرهنگهای مختلفی نیاز دارند. این فرهنگها به لحاظ اصول نظری که سیستم بر پایه آن طراحی شده است و نیز به لحاظ شکل عمومی، پوشش اطلاعات، میزان جزئیات و دقت صوری که در توصیف واژگانشان به کار رفته است با یکدیگر بسیار متفاوتند.

۱-۱ اصول نظری

به طور کلی زبانشناسی صورتگرا^(۱) الهام بخش مبانی نظری تحقیقات ترجمه ماشینی است چراکه منطبق ساختن قواعد زبانی با اصول ریاضی و قواعد صوری، امکان خود کار کردن عملیات ترجمه را به وجود می آورد. دستورهای متن آزاد^(۲) که در اصل پایه و اساس دستور صورتگرا^(۳) است، منبع اصلی برای تهیه برنامه‌های کامپیوتری می‌باشد. هم اکنون نیز اکثر سیستم‌های ترجمه ماشینی بر پایه اصول دستور گشتاری و دستور وابستگی^(۴)، توأم با تعدیلاتی که ناشی از وام‌گیری از سایر نظریه هاست، طراحی شده‌اند. اغلب این سیستم‌ها به لحاظ اصول نظری ترکیبی از گرایشها و مدل‌های مختلف هستند. سیستم‌هایی که در کلیه روندهایشان تنها به یک نظریه واحد پایبندند. بیشتر حالت تجربی دارند و نیز سیستم‌هایی وجود دارند که دارای هیچ گونه پایگاه نظری نمی‌باشند- که البته مورد انتقاد است- (Hutchins, 1986).

سیستم‌های انتقالی در بدو پیدایش پیرو نظریه دستور گشتاری چامسکی بوده است لیکن به تدریج به دلایل مختلف، از جمله، تولید صورتهای میانی فراوان که پس از تولید می‌بایست حذف می‌شدند، تغییراتی در گرایشهای نظری این سیستم‌ها پدید آمد. در حال حاضر می‌توان علاوه بر تأثیر دستور گشتاری، تأثیر دستور زایشی، نظریه حالت و تحلیل‌های معنایی را بر حسب نوع سیستم به میزان‌های مختلف در آن مشاهده کرد. البته هم اکنون نیز مشخصه‌های گرایش ساختاری یا بازگشتی^(۵) که عبارت است از؛ تغییر متن زبان مبدأ به متن زبان مقصد با استفاده از روندهای تکرار شونده نمودهای متنی، در این سیستم‌ها یک ویژگی مشخص کننده است. ساختار لایه‌ای سیستم انتقالی و نیز ترتیب مشخص در ارائه سطوح، نمایانگر ویژگی ساختاری این سیستم است. در این سیستم سطح "i+1" حاصل اعمال روندها و قواعد تکرار شونده بر سطح "i" می‌باشد.

1- Formal linguistics

2- Context-Free grammars

3- Formal grammar

4- Dependency grammar

5- Structural or recursive approach

تغییر مرحله به مرحله ساختارها، هم در بخش انتقال و هم در بخش‌های تک زبانه نحو و واژگان مشاهده می‌شود (Eynde, 1993).

از آنجایی که سیستم‌های انتقالی معمولاً نحو-بنیاد^(۱) می‌باشند و نه واژه‌بنیاد^(۲) و واحد ترجمه در آنها جمله است و نه واژه یا تکواژ، بنابراین بحث اصول و دیدگاه‌های نظری این سیستم‌ها بیشتر در مباحث نحوی مطرح می‌گردد و مباحث واژگانی آنها نیز بالطبع پیرو همان اصول نظری است، بدین مفهوم که ساختار لایه‌ای نحو در فرهنگ لغت نیز منعکس می‌شود و اطلاعات مربوط به انواع سطوح در آن ذخیره شده و در هر مرحله ارائه می‌گردد. لازم به ذکر است که در برابر گرایش بازگشتی گرایش محدودیت بنیاد^(۳) قرار می‌گیرد و آن عبارت است از؛ قرار دادن محدودیت‌هایی بر معادل‌یابی ترجمه‌ای بدون هیچ گونه نظری پیرامون روندهای تکرار شونده یا ترتیبی (Sadler, 1993). طرح پیشنهادی دستور واژگانی نقشمند^(۴) یکی از معدود نمونه‌های این گرایش است.

۱-۲ ترجمه ماشینی

ترجمه ماشینی حیطه‌ای کاربردی برای رشته‌های تخصصی کامپیوتر، زبان‌شناسی، ترجمه و اصطلاح‌شناسی می‌باشد. این حیطه کاربردی هم اکنون بخشی از رشته جدید زبان‌شناسی رایانه‌ای را به وجود آورده است. زبان‌شناسی رایانه‌ای به مطالعه هرگونه همکاری رایانه در پردازش و برخورد با زبان‌های طبیعی می‌پردازد. ترجمه ماشینی همچنین با رشته دیگر یعنی هوش مصنوعی در ارتباط است. زبان‌شناسی رایانه‌ای و هوش مصنوعی باعث خلق روش‌ها و نظریه‌های ویژه و نیز اصطلاحات جدید و درهم آمیخته بوده‌اند (Lehrberger & Bourbeau, 1988).

ترجمه ماشینی بدون شک یکی از جالب‌ترین زمینه‌های تحقیقی است که به لحاظ

1- Syntax-based

2- Word-based

3- Constraint-based approach

4- Lexical Functional Grammar (LFG)

اجتماعی، سیاسی، اقتصادی، تجاری، ذهنی و فلسفی حائز اهمیت است. هدف این رشته تحقیقی رسیدن به حجم زیادی از ترجمه، در مدت کوتاه، با قیمت اقتصادی است. واقعیت ترجمه ماشینی فشرده مطالبی است که در پشت گزارشهای تحقیقی و برنامه‌های اجرایی و عملی نهفته است لیکن از آنجایی که ترجمه ماشینی از آغاز با ادعاهای بزرگ و انتظارات گزاف به صورتی جالب متصور شده است لذا گاهی تعابیر نادرستی در مورد آن به چشم می‌خورد بنابراین بهتر آن است که یافته‌های کنونی این سیستم‌ها را با متناسب کردن انتظاراتمان ارج نهیم (Arnold et al, 1994).

۱-۳ تاریخچه ترجمه ماشینی در جهان

مشارکت رایانه در ترجمه زبانهای طبیعی پس از جنگ جهانی دوم با یادداشت معروف وارن وی‌ور^(۱) در سال ۱۹۴۹ آغاز شد، یادداشتی که طی آن بیان شده بود که می‌توان ترجمه را از طریق تغییر در وظایف کدهای ماشین انجام داد. حجم وسیعی از منابعی که هم اکنون در اختیار ترجمه ماشینی به ویژه در آمریکا قرار دارد پس از گزارش آلپاک^(۲) در سال ۱۹۶۶ به دست آمده است، گزارشی که در آن توصیه شده بود که بر اساس یافته‌های کنونی، ترجمه ماشینی می‌بایست جایگاه خود را در یکی از زمینه‌های علمی رشته عمومی زبانشناسی رایانه‌ای پیدا کند.

در ایالات متحده آمریکا در دهه‌های ۱۹۵۰ و ۱۹۶۰ ترس از پیشرفتهای فنی شوروی (به ویژه پس از پرتاب نخستین اسپوتنیک^(۳) در سال ۱۹۵۷) حمایت‌های دولتی و نظامی از ترجمه ماشینی (روسی - انگلیسی) رو به فزونی نهاد. در این دو دهه همچنین گروههای مختلفی در انگلستان و اروپای غربی، شوروی و اروپای شرقی و ژاپن و چین به تحقیق پیرامون ترجمه ماشینی و انجام پروژه‌های مختلف پرداختند. در این دهه‌ها

1- Warren Weaver

2- ALPAC (Automatic Language Processing Advisory Committee).

3- Sputnik