

طراحی مدل مفهومی ترجمه ماشینی بافت-آگاه: رهیافتی به ترجمه بافت-بنیان^۱

ابراهیم عزتی^۲

چکیده

این پژوهش به صورت میان-رشته‌ای با الهام از مباحث بافت-بنیان در ترجمه از یک سو و با بهره‌گیری از دستاوردهای پردازش طبیعی زبان از طرفی دیگر، سعی در معرفی مدلی مفهومی برای ترجمه ماشینی دارد که بر اساس بافت گفتمان طراحی شده است. با توجه به اهمیت بافت در ترجمه جملات از زبان مبدأ به زبان مقصد، مدل ترجمه ماشینی بافت-آگاه به عنوان مدل پیشنهادی برای رفع مشکلات ترجمه‌ای و جاری ترجمه‌های ماشینی ارائه می‌گردد. مدل طراحی شده نه تنها از مشخصه‌های قابل مشاهده در بافت بهره می‌برد، بلکه از ویژگی‌های فرامتنی و همچنین قابل پیش‌بینی در بافت نیز برای دستیابی به ترجمه صحیح و مناسب‌تر استفاده می‌کند. براین اساس، بافت به سه حوزه کاربردی یعنی (۱) موضوع، (۲) فضای تصویری و (۳) عناصر سازنده دسته‌بندی شد. در این مدل که یک تابع از X به Y است؛ هر واژه، عبارت یا جمله ورودی پس از تبدیل به دنباله X یعنی الگوی قابل پردازش برای ماشین ترجمه، وارد پردازش چهار مرحله‌ای سیستم بافت-آگاه یعنی (۱) کسب بافت، (۲) مدل‌سازی بافت، (۳) استنتاج بافت و (۴) توزیع بافت می‌شود. در این چهار مرحله، ویژگی‌های بافتی صریح و همچنین ویژگی‌های بافتی ضمنی در خصوص جمله ورودی، تحلیل و مشخص می‌گردند. پس از پردازش سیستم بافت-آگاه، خروجی آن به صورت دنباله Z_1 مرتب می‌شود. این دنباله مطابق ویژگی‌های بافتی زبان مبدأ مرتب شده است. سپس دنباله Z_1 به دنباله Z_2 مطابق زبان مقصد برابری می‌شود. در نهایت ساختار Z_2 به ساختار Y که جمله نهایی ترجمه شده به زبان مقصد است، تبدیل می‌شود. نوآوری در مدل پیشنهادی محقق، ادغام ویژگی‌های ترجمه بافت-بنیان با ویژگی‌های سیستم بافت-آگاه در مراحل پردازش است. یکی دیگر از پیشنهادهای خلاقانه این مدل، بررسی میزان رضایت کاربر از ترجمه ارائه شده است که به منظور کیفیت‌سنجی ترجمه و در صورت نیاز اصلاح و بهبود آن در نظر گرفته می‌شود.

واژه‌های راهنما: ترجمه بافت-بنیان، ترجمه ماشینی، سیستم بافت-آگاه، گفتمان

۱. این مقاله در تاریخ ۱۴۰۱/۰۳/۱۴ دریافت شد و در تاریخ ۱۴۰۱/۰۵/۳۱ به تصویب رسید.

۲. عضو هیئت علمی، بخش زبان‌شناسی و زبان‌های خارجی، دانشگاه پیام‌نور، تهران، ایران؛ پست الکترونیک:

e_ezzati@pnu.ac.ir

۱- مقدمه

لارسن معتقد است قالب‌ها سازه‌های روساختی زبان مانند واژه، عبارت یا جمله هستند. اما منظور یا همان نیت متکلم، معیاری است که شکل قالب‌ها یا جمله‌های زبان را مشخص می‌کند. از این رو برای درک منظور یک جمله، به بافتی^۱ که جمله در آن به کار رفته نیاز است (Larson, 1984). طبق دیدگاه فرث^۲ و هالیدی^۳، منظور و بافت فارق از یکدیگر نیستند و بدون بافت، درک دقیق و فرا جمله‌ای معنا و مفهوم (منظور) یک جمله، به درستی میسر نخواهد بود (Stanlaw, 2020).

در علم ترجمه، بافت می‌تواند تکواژ، واژه، عبارت و جمله‌ای قابل مشاهده در یک متن یا مکالمه باشد و یا حتی می‌تواند بخشی غیر قابل مشاهده و استنباطی و قابل پیش‌بینی از یک گفتمان^۴ باشد که به مترجم کمک کند تا به ترجمه دقیقی از یک واژه، عبارت و یا جمله نائل آید (Idlibi, 2019). در حقیقت، فضا و زمان گفتمان، حالات شرکت‌کنندگان، حجم و ابعاد صحنه اتفاقات همه به صورت تصاویری ثابت و متحرک در نظر گرفته می‌شود تا درکی شفاف و متعهد به بافت متن زبان مبدأ به خواننده یا شنونده زبان مقصد ارائه دهد (Song, 2010). به عبارت ساده‌تر، مترجم متناسب با بافت زبان مقصد، معنای جمله صحیح و مناسب زبان مبدأ را به خواننده یا شنونده زبان مقصد انتقال دهد. بنابراین رسیدن به دقیق‌ترین ترجمه میسر نخواهد بود مگر با در نظر گرفتن جامع تمام جنبه‌هایی که در تاویل معنا و مفهوم جمله نقش دارند (Nugroho, 2016; Pranowo, 2020).

۲- بافت و ترجمه ماشینی

۱. context

۲. Firth

۳. Halliday

۴. discourse

اگر بافت گفتمان به دو نوع کلی صریح^۱ و ضمنی^۲ دسته‌بندی شود (قنبریان وهمکاران، ۱۴۰۱) در آن صورت، هر اطلاعاتی که در کنار واژه یا عبارت یا جمله‌ای دیگر آمده باشد و مترجم را در درک بهتر و در نهایت، رسیدن به ترجمه‌ی درست کمک می‌کند، بافت صریح است. چنین بافتی می‌تواند کلمات یا عبارات و یا جملات بکاررفته قبل و یا بعد از واژه، عبارت یا جمله مورد ترجمه باشد که فضا و زمان و مکان و سایر اطلاعات مرتبط با کلمه، عبارت یا جمله مورد نظر برای ترجمه را مشخص می‌کند. بافت ضمنی ویژگی‌هایی هستند که به‌طور آشکار در گفتمان ذکر نشده‌اند ولی برای درک منظور بسیار حائز اهمیت و بر اساس بافت صریح قابل‌استنباط هستند. بنابراین به مواردی همچون موضوع گفتمان، عواطف و احساسات شرکت‌کنندگان و نوع رابطه آنها در گفتمان، شرایط حاکم بر گفتمان، سن و جنسیت، مکان و زمان گفتمان و ویژگی‌های فرهنگی و حتی تاریخی و اجتماعی بافت ضمنی گفته می‌شود (ترکاشوند و قائمی، ۱۳۹۶).

بافت در ترجمه ماشینی^۳ بر اساس الگوی پردازش زبان طبیعی^۴ تعریفی منحصر دارد. بافت در پردازش زبان طبیعی، هر گونه اطلاعاتی است که در جهت توصیف و تشریح موقعیت یک ورودی استفاده می‌شود. در ترجمه ماشینی این ورودی‌ها همان عناصر زبانی هستند که به زبان ماشین کدگذاری کدگشایی می‌شوند (Joshi et al., 2020). به‌طور کلی، اگر فرایند ترجمه، به سه مرحله ترجمه، ویرایش و تهیه نسخه نهائی دسته‌بندی شود، ماشین‌های ترجمه با تمام پیشرفت و توسعه در فناوری اطلاعات هنوز در مرحله اول فرآیند یعنی ترجمه هستند. بدین منظور باز هم نیاز به وجود متخصصین زبان و مترجمان حرفه‌ای در مراحل ویراستاری و تهیه نسخه نهائی ترجمه، مبرم و ضروری است (Kenny, 2019). البته هدف از ترجمه، چه ماشینی و یا غیر، ارائه محصولی بدون نقص از لحاظ انتخاب واژگان و عبارات، جمله‌بندی و در کل ترجمه‌ای کاربردی

۱. explicit

۲. implicit

۳. machine translation

۴. natural language processing

در انتقال درست پیام و متناسب با فرهنگ و بافت زبانی^۱ زبان مقصد است. (Bender & Good, 2020). در این خصوص، ماشین‌های ترجمه در ترجمه اصطلاحات زبانی^۲، استعارات^۳، ابهامات واژگانی^۴، پیچیدگی‌های نحوی^۵، ترجمه عبارات فعلی^۶، ترجمه همخط‌ها^۷ و تفاوت‌های زبانی مابین زبان‌ها، ساختارهای غیر دستوری^۸ و مستتر^۹ و مواردی مشابه با چالشی جدی و اساسی روبرو هستند (Alzeebaree, 2020).

برای مثال در ترجمه جمله‌ای که مادری به پسرش می‌گوید:

Why don't you take the garbage out?

بر اساس عبارت "why don't you" متوجه می‌شویم که این جمله هرچند ظاهراً سؤالی دارد ولی هدفش درخواست از کسی برای انجام کاری است. این نوع اطلاعات در بافت به صراحت بیان شده است. فرض کنید جملات قبلی این پاره‌گفتار به صورت ذیل است:

John was playing a videogame; meanwhile her mother said: Why don't you take the garbage out?

باتوجه به اطلاعات بیان شده در بافت یا همان جملات قبلی می‌توان حدس زد که گفتمان در زمان بعد از صرف غذا و احتمالاً شام رخ داده است. چون معمولاً گذاشتن سطل آشغال در جلوی درب منزل در هنگام شب اتفاق می‌افتد. «جان» پسر خانواده است و به مادرش در کار خانه کمک چندانی نمی‌کند و بازیگوش است. چون درست بعد از شام سراغ بازی رایانه‌ای رفته است.

-
۱. linguistic context
 ۲. idioms
 ۳. metaphors
 ۴. lexical ambiguities
 ۵. syntactic complications
 ۶. phrasal verbs
 ۷. homographs
 ۸. ungrammatical structures
 ۹. implicit

می‌توان حدس زد که در خانه کس دیگری نیست و این مادر و پسر در خانه تنها زندگی می‌کنند. حضور پدر خانواده کم‌رنگ است شاید پدر خانواده تا دیر وقت کار می‌کند و یا اینکه فوت شده است و این مادر است که بار تمام کارها را به‌تنهایی به دوش می‌کشد. همچنین هر چند پسر در اواخر سن نوجوانی است ولی هنوز در دوران کودکی خود غرق است. حالا با توجه به این توضیحات، کدام ترجمه برای جمله بالا مناسب‌تر است؟

الف) چرا سطل اشغال رو بیرون نمی‌داری؟

ب) میشه لطفاً سطل اشغال رو بیرون بذاری!

در ترجمه با توجه به این اصل که تأثیری که خواننده یا شنونده از یک جمله در زبان مبدأ می‌گیرد باید در ترجمه‌اش نیز حفظ شود و همان تأثیر را درک کند، ترجمه «ب» مناسب‌تر و کاربردی‌تر است. هم‌اکنون فرض کنیم که پسر درخواست مادر را پشت گوش انداخته و همچنان مشغول بازی است. در این زمان مادر پس از مدتی برای بار دوم همان جمله را به پسر می‌گوید:

Why don't you take the garbage out?!

در جمله فوق، بافت، به موقعیت تکرار درخواستِ مادر اشاره می‌کند و این بار درخواست با حالتی طعنه‌آمیز است. حال کدام ترجمه مناسب‌تر است؟

الف) میشه لطفاً سطل اشغال رو بیرون بذاری!

ب) بد نیست سطل اشغال رو بیرون؟!!

در این موقعیت، مترجم باید جملهٔ زبان مبدأ را بر اساس بافت ضمنی و استنباطی به گونه‌ای ترجمه کند که لحن طعنه‌آمیز مادر و خستگی‌اش از تنها کار کردن تا دیروقت و بی‌خیالی پسرش به این موضوع، در ترجمه نمایان شود. بنابراین می‌توان اذعان داشت که ترجمه «ب» چنین ویژگی‌های را دارد هرچند آن ویژگی‌ها به‌طور آشکار در متن بیان نشده‌اند.

مثالی دیگر برای همخط‌ها یعنی کلماتی که دارای نگارشی یکسان ولی با معنی متفاوت:

Go stand at the bow if you're feeling seasick.

ترجمه GOOGLE TRANSLATE:

اگر احساس دریازدگی می‌کنید، در کمان بایستید.

در ترجمه فوق کلمه همخط *bow* کمان ترجمه شده است؛ در صورتی که منظور عرشه جلوی کشتی است.

مثال برای کلمه مبهم:

Call me a taxi, please.

ترجمه GOOGLE TRANSLATE:

لطفاً یک تاکسی به من زنگ بزنید.

آیا منظور این است که با تلفن یک تاکسی برای من بگیرید یا مفهومی استعاری به معنای مرا یک تاکسی صدا بزنید؟! آیا واژه *call* باید زنگ زدن ترجمه شود یا صدا زدن؟ آیا تاکسی نام کسی است یا یک وسیله نقلیه؟!

حال اگر منظور گوینده گرفتن یک تاکسی برایش باشد، و گوینده در گفتمانی برای بار چهارم بگوید:

Call me a taxi, please.

در این صورت، آیا ترجمه عبارت فوق با ترجمه همان عبارت که برای بار اول بیان شده است، فرقی ندارد؟ اگر فرق دارد، چگونه باید به این تفاوت دست یافت تا ترجمه‌ای صحیح از عبارات داشته باشیم؟

مثال برای عبارت مبهم:

I saw someone on the hill with a telescope.

ترجمه GOOGLE TRANSLATE:

شخصی را روی تپه با تلسکوپ دیدم.

مشخص نیست که آیا گوینده یا نویسنده روی تپه بود یا کسی که تلسکوپ داشت؟!

نظر به توضیحات ارائه شده، چگونه می‌توان به ترجمه‌ای صحیح در ترجمه ماشینی دست یافت که هم ویژگی‌های خاص زبان مبدأ را با تحلیلی جامع دریافت نماید و هم در زبان مقصد ترجمه‌ای دقیق و متعهد به زبان مبدأ را ارائه دهد؟ در پاسخ به این سؤال، باید خاطر نشان کرد که نائل آمدن به ترجمه‌ای چنین دقیق و بی‌نقص تاکنون نه برای انسان ممکن بوده است و نه برای ماشین، هر چند محقق، ترجمه بافت-بنیان را به‌عنوان رویکردی پیشنهادی جهت رفع ابهامات در ترجمه و معیار دست یافتن به ترجمه‌ای صحیح معرفی می‌نماید. بر این اساس، هدف این پژوهش ارائه مدلی مفهومی^۱ در ترجمه ماشینی است که مشکلات موجود آنها را بر اساس سیستم بافت-آگاه^۲ که خود مستخرج از ترجمه بافت-بنیان است، مرتفع سازد و راه را برای محققین حوزه ترجمه و فناوران علاقه‌مند به ترجمه ماشینی هموارتر نماید.

۳- ترجمه بافت-بنیان

ترجمه ماشینی بافت-بنیان ضمن در نظر گرفتن بافت صریح و ضمنی در گفتمان برای رسیدن به ترجمه‌ای درست و متناسب، ماهیت الگوی بافت-بنیان را بر سه اصل کاربرد موضوعی^۳، فضای تصویری^۴ و عناصر سازنده^۵ جمله استوار می‌سازد (نیلی‌پور، ۱۳۷۸ و خان‌جان، ۱۳۹۵). کاربرد موضوعی در واقع به موضوعاتی اشاره می‌کند که منظور کلی جمله است. جملات بر اساس اشاره به موضوعاتی همچون عشق و نفرت، شادی و غم، موسیقی، رویا، غذا، مسافرت، خانواده، شغل، ورزش و سایر موضوعات فردی یا اجتماعی و فرهنگی دیگر در گفتمان تولید می‌شوند. فضای تصویری متشکل از تصویری است که جمله چه به صورت ثابت یا متحرک از خود نمایش می‌دهد. در واقع ماهیت متحرک یا ثابت بودن تصویر جمله بر اساس ماهیت فعل آن تعیین

۱. conceptual model

۲. context-aware system

۳. usage

۴. image

۵. constituents

می‌شود. اگر فعل جمله ماهیتی کنشی^۱ (حرکت داشتن) مانند «دویدن و پریدن» و یا فرایندی^۲ (تغییر حالتی به حالتی دیگر) مانند «مردن» و یا ترکیب آن دو یعنی کنشی-فرایندی^۳ (هم حرکتی و هم تغییردهنده حالت) مثل «کشتن» داشته باشد، فضای تصویری جمله متحرک است ولی اگر فعل صرفاً به حالتی^۴ بدون حرکت و تغییر مانند «گرم بودن یا سرد بودن» اشاره کند فضای تصویری جمله ثابت است. همچنین در فضای تصویری جمله علاوه بر موارد مذکور، مواردی دیگری هم وجود دارد از جمله ماهیت فضا و زمان^۵ جمله که مشخص می‌کند کی و کجا گفتمانی اتفاق افتاده است و اینکه شرکت کنندگان در گفتمان چه کسانی هستند و چه رابطه‌ای با هم دارند و سیاق و سبک گفتمان چگونه است؛ رسمی است یا غیر رسمی. در نهایت عناصر سازنده بافت به اجزا تشکیل‌دهنده جمله اشاره دارند که در رأس آن فعل است. فعل مرکز و هسته اصلی جمله است که با تشخیص آن می‌توان به فاعل و مفعول جمله نیز دست پیدا کرد (Ashengo et al., 2021).

۴- ترجمه ماشینی بافت-آگاه

۴-۱- ترجمه ماشینی

به طور اجمالی در ترجمه ماشینی می‌توان به سه رویکرد آماری^۶، تحت وب معنایی^۷ و عصبی^۸ اشاره کرد. ترجمه ماشینی آماری در سال ۱۹۴۹ معرفی شد که البته اکثر رویکردهای دیگر ترجمه ماشینی یا بر اساس این روش ساخته شدند یا از تکنیک‌های آن استفاده کرده‌اند. این

۱. action

۲. process

۳. action-process

۴. state

۵. space and time

۶. statistical

۷. semantic web-based

۸. neural

رویکرد از قواعد آماری و الگوریتم‌های یادگیری ماشین^۱ استفاده می‌کند. مکانیزم ترجمه ماشینی آماری بدین نحو است که گروهی از کلمات یا نشانه‌ها را از زبان مقصد به گروهی از کلمات و نشانه‌ها در زبان مبدأ ترجمه می‌کند. به عبارت دیگر، کلمات زبان گروه A به کلمات زبان گروه B در زبان مبدأ ترجمه می‌شوند. در این ترجمه، اگر جملات ترجمه شده قبلی ذخیره شده باشند، صحت و دقت ترجمه تا حدود ۱۰۰ درصد هم بالا می‌رود. در غیر این صورت، ترجمه از صحت بالایی برخوردار نخواهد بود. یکی از خاص‌ترین ویژگی این رویکرد، ذخیره‌سازی معادل کلمات در متون ترجمه شده از زبان مبدأ به مقصد است. (Koehn, 2009). رویکرد ترجمه ماشینی تحت وب معنایی تقریباً در همان دوره شکل‌گیری ترجمه ماشینی آماری پدید آمد. البته ماهیت هستی-شناسی^۲ این رویکرد در دهه ۱۹۹۰ شکل گرفت ولی ویژگی تحت وب معنایی این نوع رویکرد به سال‌های اخیر برمی‌گردد. این نوع رویکرد به وجود آمده است که از تطبیق صرف واژه‌ها فراتر رفته و جست‌وجو را براساس موضوع، ارتباط میان داده‌ها، نوع داده‌ها، و کیفیت‌های دیگر انجام می‌دهد. این رویکرد ابتدا به تجزیه و تحلیل معنایی متون می‌پردازد و سپس تجزیه و تحلیل نحوی متون را انجام می‌دهد. در مواردی که تجزیه و تحلیل نحوی صورت نگیرد، تجزیه و تحلیل تکواژشناسی جایگزین می‌شود. در واقع هم برای ویژگی‌های نحوی و هم تکواژشناسی از فایل لغتنامه‌ای خود استفاده می‌کند. همچنین در این نوع ترجمه ماشینی نیازی به ذخیره همه اطلاعات نیست، بلکه تنها الگوها و ویژگی‌های معنایی ذخیره می‌شوند (França et al., 2021; Moussallem et al., 2018). جدیدترین نوع ترجمه ماشینی که از سال ۲۰۱۳ به بعد وارد عرصه رقابت با دیگر ترجمه‌های ماشینی شد ترجمه ماشینی عصبی است. روش کار این رویکرد بر اساس علم آمار و شبکه عصبی^۳ یا الگوریتم‌های یادگیری عمیق^۴ می‌باشد. تفاوتش با ماشین ترجمه

۱. machine learning

۲. ontological

۳. neural network

۴. deep learning

آماري، استفاده از قسمت‌های کدگذار^۱ و کدگشا^۲ می‌باشد. هر چند برای تجربه و تحلیل نحوی جملات از روش ترجمه آماری استفاده می‌کند. در واقع، ابتدا با استفاده از روش شبکه عصبی بازگشتی^۳ داده‌های ورودی را به واژه‌ها و تکواژها کدگذاری می‌کند، سپس برای معادلیابی آنها از روش ترجمه ماشینی آماری استفاده می‌کند. بعد از معادلیابی واژگان و تکواژها، آنها را به شبکه عصبی بازگشتی ارسال می‌کند تا به زبان مبدأ کدگشایی شوند. یکی از ویژگی‌های بارز این نوع ترجمه ماشینی، ابهام زدایی با استفاده از روش ان-دنباله^۴ است (Koehn, 2017).

۴-۲- سیستم بافت-آگاه

به‌طور کلی، سیستم بافت-آگاه از دهه ۹۰ به بعد معرفی شد. در سال ۱۹۹۴، تایمر^۵ و شیلیتار^۶ بافت-آگاه را سیستمی می‌دانستند که شامل مکان، افراد و اشیاء نزدیک باشد. در سال ۱۹۹۷، براون^۷ و بووی^۸ زمان و دما را آن افزودند. در سال ۱۹۹۸، پاسکو^۹ بافتی مبتنی برحالت‌های فیزیکی و انتزاعی مورد علاقه کاربر را در یک موجودیت خاص، معرفی کرد. در سال ۲۰۰۸، دی^{۱۰} و هاکیلا^{۱۱}، بافت-آگاه را نظامی می‌داند که با استفاده از بافت، قادر است اطلاعات و/یا خدمات مرتبط به فعالیت‌های کاربر را ارائه دهد (Dey & Häkkinen, 2008).

۱. encoding

۲. decoding

۳. recurrent neural network(RNN)

۴. n-gram

۵. Theimer

۶. Schilitar

۷. Brown

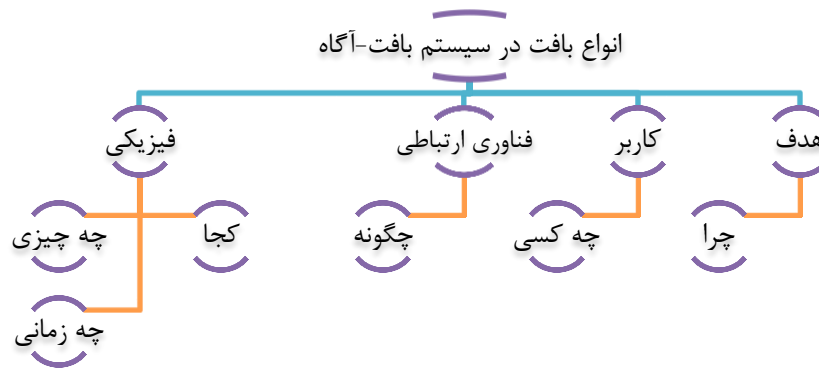
۸. Bovey

۹. Pascoe

۱۰. Dey

۱۱. Häkkinen

همان‌طوری که بافتِ گفتمان، انواعی دارد، در سیستم بافت-آگاه نیز تنوع بافتی وجود دارد. پوسلاد^۱ برای چنین بافتی، تقسیم‌بندی ذیل را ارائه می‌دهد (شکل ۱) (Poslad, 2009):



۱- شکل انواع بافت در سیستم بافت-آگاه

طبق دیدگاه پوسلاد، بافت به‌طور کلی به چهار نوع تقسیم می‌شود: (۱) فیزیکی، (۲) فناوری اطلاعاتی و ارتباطی، (۳) کاربر، (۴) هدف. در بافت فیزیکی هرگونه اطلاعاتی که در پاسخ به سؤالات «چه چیزی؟، کجا؟، و چه زمانی؟» باشد، مد نظر قرار می‌گیرد. در بافت فناوری اطلاعاتی و ارتباطی، اطلاعات در پاسخ به سوال «چگونه؟» جمع‌آوری می‌شوند. در بافت کاربر، اطلاعات مطابق پاسخ به سوال «چه کسی؟» دریافت می‌شوند. و در نهایت، بافت هدف، بر اساس پاسخ به سوال «چرا؟» دریافت می‌شوند؛ به‌عبارت دیگر، چرا چنین بافتی در تحلیل داده‌ها ضروری است؟ و چرا باید از این بافت استفاده کرد؟ و هدف کاربر از به‌کارگیری چنین بافتی چیست؟

۴-۳- مراحل پردازش بافت-آگاه

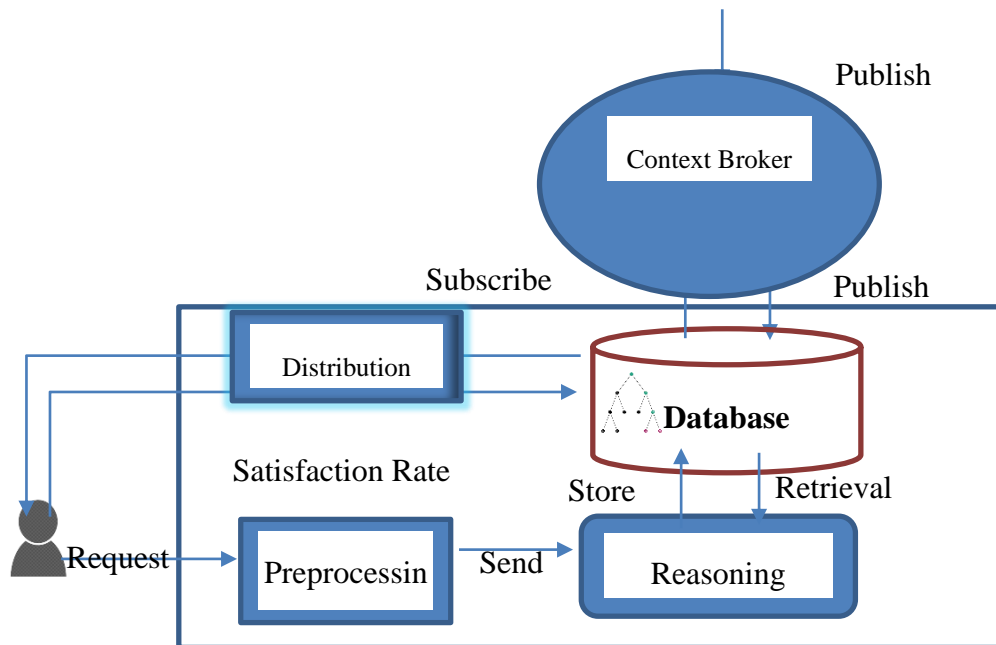
فرآیند چرخه حیات سیستم بافت-آگاه شامل ۴ مرحله می‌باشد (Pradeep & Krishnamoorthy, 2019). مدل کلی مراحل انجام کار مطابق شکل زیر (شکل ۲) است. ابتدا، در

۱. Poslad

مرحله کسب بافت^۱، داده‌ها از سنسورهای مختلف جمع‌آوری می‌شود و در بروکر بافت^۲ که بر اساس مدل انتشار/اشتراک^۳ است، داده‌های سنسورها در کانال‌هایی بر اساس موضوع قرار می‌گیرند (Lazidis, et al., 2022). در مرحله دوم، داده‌ها برای ذخیره در پایگاه‌داده^۴ مربوط به بافت، مدل‌سازی^۵ می‌گردند (Gollagi, et al., 2020). این مدل بر اساس یک ساختار درختی مطابق شکل (۱) در پایگاه‌داده و در محلی مربوط به بافت ذخیره می‌گردد تا در مراحل ترجمه از آن استفاده شود. مرحله سوم از چرخه حیات بافت، مربوط به استنتاج^۶ بافت است. در این مرحله، بنا بر درخواست کاربر برای ترجمه، پس از پیش‌پردازش^۷ متن و تعیین جملات، فعل، فاعل و سایر ارکان جمله با توجه به ارتباط بین بافت و سایر ارکان، استنتاج انجام می‌گیرد. برای استنتاج از روش قاعده-بنیان^۸ استفاده می‌گردد. در مرحله چهارم یعنی توزیع بافت^۹، نتایج نهایی به کاربر نمایش داده می‌شود و جهت بهره‌وری از نظرات کاربر و پردازش بهینه‌تر ترجمه‌های بعدی، میزان رضایت کاربر نیز مشخص می‌گردد (De Mato, et al., 2020). چنین ساختاری در این مدل پیشنهادی، میزان تأثیر ترجمه ارائه‌شده را بر اساس احساس مخاطب اندازه‌گیری می‌کند؛ بازخوردی که امکان سنجش آن در ترجمه‌های انسانی وقت‌گیر و پرهزینه است (Sezer, et al., 2018).

Sensors (Acquisition)

۱. context acquisition
۲. context broker
۳. publish/subscribe
۴. database
۵. modelling
۶. reasoning
۷. preprocessing
۸. rule-based
۹. context distribution



شکل ۲- پردازش در سیستم بافت-آگاه

۴-۴- مدل پیشنهادی ترجمه ماشینی بافت-بنیان

به طور مختصر، در طراحی مدل ترجمه ماشینی بافت-بنیان (شکل ۳) ضمن بهره‌گیری از پردازش سیستم بافت-آگاه، از شاخص‌های ترجمه بافت-بنیان نیز استفاده شده است. به عبارت دیگر، سیستم ابتدا متن ورودی را بررسی و تجزیه^۱ می‌کند و سپس با پردازش در سیستم بافت-آگاه بافت-بنیان، پاسخی قابل پیش‌بینی به کاربر ارائه می‌دهد (Fernandes et al., 2021). در حقیقت، در ترجمه ماشینی بافت-بنیان، جمله ورودی را پردازش می‌کند که توسط کاربر، به سیستم معرفی می‌شود و پس از تجزیه به ساختار قابل پردازش ماشین، به سیستم بافت-آگاه ارسال می‌گردد (شکل ۲):



شکل ۳- مدل پیشنهادی ترجمه ماشینی بافت-بنیان بر اساس سیستم بافت-آگاه

در مرحله نخست، جمله ورودی که توسط کاربر به سیستم داده می‌شود، با کمک مدل‌های نحوی (Chomsky, 1957 & Chomsky, 1965)، به دنباله X تجزیه می‌گردد. فرایند تجزیه جمله ورودی بر اساس پردازش‌های کلمه-محور، عبارت محور و جمله-محور صورت می‌گیرد. در این پردازش‌ها جمله ورودی با توجه به ویژگی‌های صرفی و نحوی زبان مبدأ از جمله نوع کلمات مانند کلمات متشکل از تکواژهای واژگانی متشکل از اسم، فعل، صفت، قید و یا تکواژهای دستوری مثل حروف تعریف و اضافه و غیره و همچنین نوع عبارات برای مثال اسمی و فعلی و یا حرف اضافه‌ای و میزان وابستگی کلمات وابسته به هسته عبارت و از جمله انواع جمله مانند جمله ساده، مرکب و یا پیچیده و مواردی این چنین، به صورت دنباله X مرتب می‌شوند. سپس دنباله X یا ورودی تجزیه‌شده، وارد مراحل پردازش بافت آگاه می‌گردد.

نکته قابل توجه در مدل پیشنهادی محقق (شکل ۲) این است که تبدیل جمله ورودی به شکل دنباله X ، جهت ارتقا، تسریع و تسهیل فرایند پردازش صورت می‌گیرد. چرا که جمله ورودی به صورت یک جمله زبانی وارد سیستم می‌شود، بنابراین به منظور قابل خوانش بودنش برای ماشین، به ساختاری سازگار با زبان ماشین مانند دنباله X تبدیل می‌شود. چنین مرحله‌ای مسیر را برای پردازش‌های بعدی تسهیل می‌کند. نوآوری دیگر در مدل پیشنهادی محقق، ادغام ویژگی‌های ترجمه بافت-بنیان با ویژگی‌های سیستم بافت-آگاه در مراحل پردازش است. چنین ساختاری در راستای کیفیت بخشی به ترجمه نهائی و در واقع تناسب ترجمه با ویژگی‌های بافتی همچون موضوع جمله ورودی، فضای تصویری و عناصر سازنده آن می‌باشد که در مدل مذکور در تبدیل دنباله Z_1 به Z_2 و در نهایت تبدیل به ترجمه نهائی پیشنهادی یعنی Y نمایان می‌گردد. یکی دیگر از

ویژگی‌های مؤثر این مدل، بررسی میزان رضایت کاربر از ترجمه ارائه شده است که به منظور کیفیت‌سنجی ترجمه و در صورت نیاز اصلاح و بهبود آن در نظر گرفته می‌شود. هم‌اکنون جهت توضیح موارد فوق مثالی ارائه می‌شود تا مراحل مدل پیشنهادی را گام‌به‌گام شرح دهد.

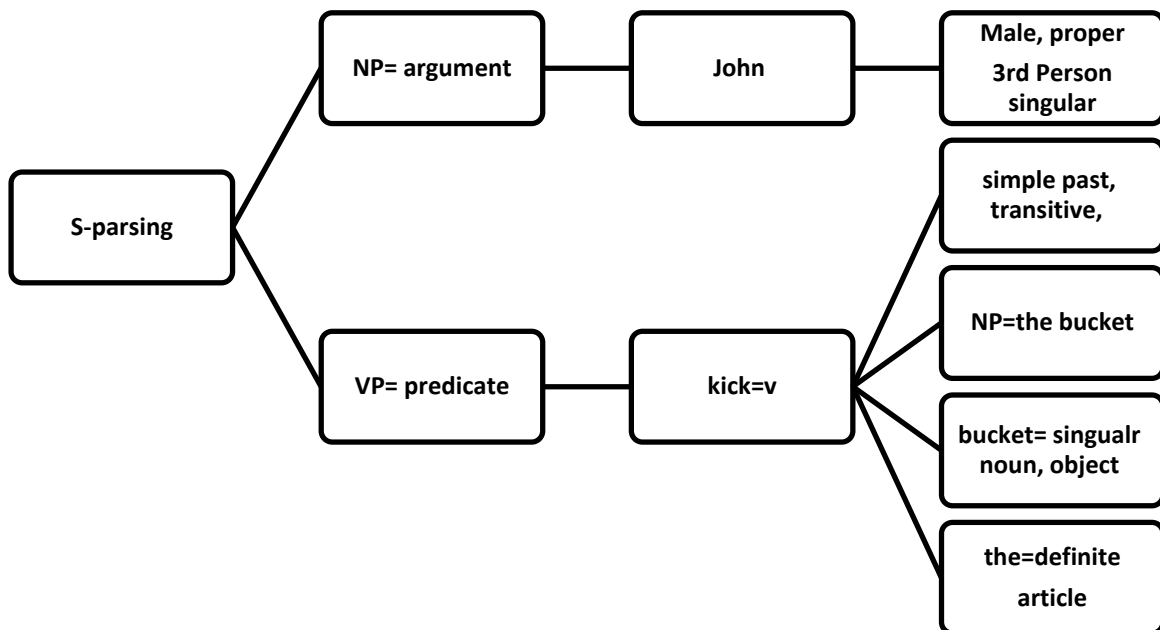
جمله:

John kicked the bucket.

را در گفتمان:

- A) You know, unfortunately we've lost John.
- B) Are you sure?!
- C) Yeah, John kicked the bucket.

به عنوان مثال در نظر می‌گیریم. بنابراین برای دستیابی به دنباله X، جمله ورودی به صورت ذیل تجزیه می‌شود (شکل ۴):



شکل ۴- تجزیه جمله ورودی به صورت دنباله X

در راستای انضمام الگوی ترجمه بافت بنیان در سیستم بافت-آگاه باید عواملی همچون کاربرد موضوعی جمله فضای تصویری و عناصر سازنده جمله ورودی را با مراحل پردازش سیستم بافت-آگاه مطابقت داد.

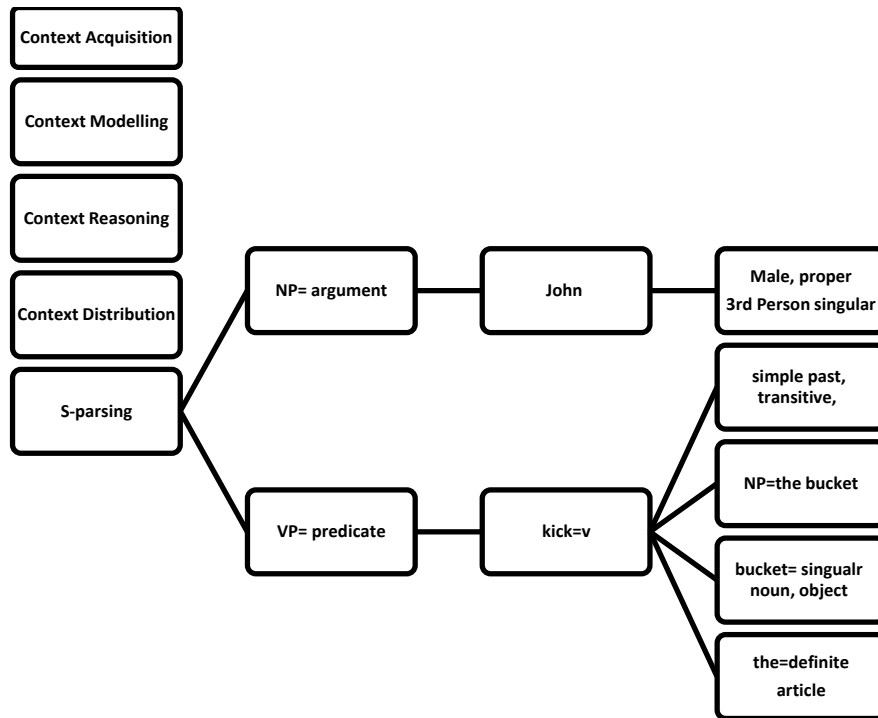
طبق الگوی ترجمه ماشینی براساس سیستم بافت-آگاه (شکل ۵)، در مرحله اول یعنی کسب بافت، کلیه اطلاعات مرتبط با جمله ورودی زبان مبدأ، به صورت نظامی شبکه‌ایی از پیکره^۱ جمع‌آوری می‌شوند. به عبارتی دیگر، زمانی که دنباله X وارد اولین مرحله سیستم بافت-آگاه می‌شود، ابتدا به جای برابرگزینی کلمه به کلمه از زبان مبدأ به زبان مقصد، کاربرد موضوعی جمله مورد بررسی قرار می‌گیرد. بنابراین مطابق با سیری در پیکره مرتبط با کاربرد موضوعی جمله، مروری بر واژه‌نامه تک‌زبانۀ زبان مبدأ در تبیین معنای کلمات و بررسی چگونگی ارتباط جملات قبلی و بعدی با یکدیگر بر اساس ارجاعات و اشارات، بافت صریح و بافت ضمنی جمله ورودی تشخیص داده می‌شود. برای مثال، در جمله "John kicked the bucket." کاربرد موضوعی جمله، برای موقعیتی است که اشاره به "to die" (مردن) دارد. در واقع در این مرحله مشخص می‌شود که جمله مذکور دارای یک اصطلاح^۲ غیررسمی و قدیمی است. بنابراین منظور از "to kick the bucket" یعنی "to die" و معنای تحت‌اللفظی آن مد نظر نیست. در صورتی که این جمله در "GOOGLE TRANSLATE" به فارسی این‌گونه به صورت لغت به لغت ترجمه می‌شود: «جان سطل را لگد زد.»

در مرحله دوم یعنی مرحله مدل‌سازی بافت، اطلاعات برگرفته از مرحله قبل، استانداردسازی می‌شوند. در واقع، سیستم در این مرحله اطلاعاتی را انتخاب می‌کند که متناسب با ویژگی‌های معنایی جمله ورودی باشد. به عبارت دیگر، اطلاعات باید معتبر و با روایی مناسب باشند و از بافت جمله ورودی برگرفته شده باشند. علاوه‌براین، اطلاعات جمله ورودی نباید در تعارض با یکدیگر باشند تا الگوی واقعی بافت جمله زبان مبدأ به‌طور دقیق و صحیح ترسیم شود. سپس باید متناسب با اطلاعات معتبر به‌دست‌آمده، معادل دقیق کلمات و عبارات انتخاب شوند و ساختار نحوی مناسب برای چیدمان کلمات نیز مشخص گردد تا کلمات مانند تکه‌های مختلف یک پازل در کنار یک‌دیگر قرار گیرند و طرحی منسجم و پایا از ترجمه جمله ورودی ارائه دهند. باتوجه‌به جمله مثال، در این مرحله مشخص می‌شود که در فضای تصویری اصطلاح "to kick the bucket" نوعی کنش یعنی "to hit something/someone with foot" دیده می‌شود و این

۱. corpus

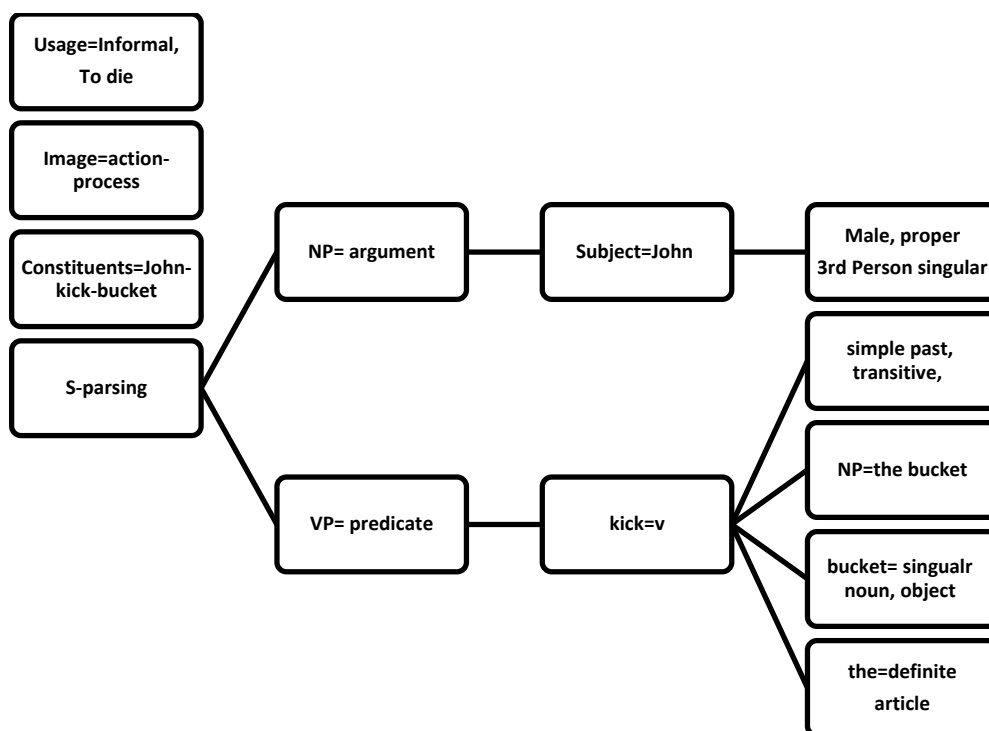
۲. idiom

اصطلاح علاوه بر داشتن فعلی که ماهیت کنشی دارد، در شخص یا شیئی نیز تغییر حالت یا فرایند تبدیل از یک حالت به حالتی دیگر به وجود می‌آورد که در این مثال، "to kick"، حالت و وضعیت "bucket" را تغییر می‌دهد.



شکل ۵- الگوی ترجمه ماشینی بر اساس سیستم بافت-آگاه

در مرحله سوم یا مرحله استنتاج بافت، اطلاعات ضروری و مؤثر در بافت زبان مبدأ در راستای رسیدن به ترجمه‌ای دقیق مشخص شده و بر اساس میزان رابطه منطقی با جمله ورودی منظم می‌شوند. برای مثال، با توجه به عناصر سازنده اصطلاح "to kick the bucket" مشخص می‌گردد که این اصطلاح دارای فعلی است با ماهیت حرکت با پا که در وضعیت شیئی دیگر (سطل) تغییر حالت ایجاد می‌کند. بنابراین، در این مرحله منسجم‌ترین و پیوسته‌ترین اطلاعات در کنار هم مرتب می‌شوند تا به مرحله نهائی یعنی مرحله نمایش نهائی به کاربر انتقال داده شوند. در این مرحله، ابتدا الگوی نهائی ترجمه جمله ورودی زبان مبدأ به صورت دنباله Z_1 ارائه می‌شود (شکل ۶).



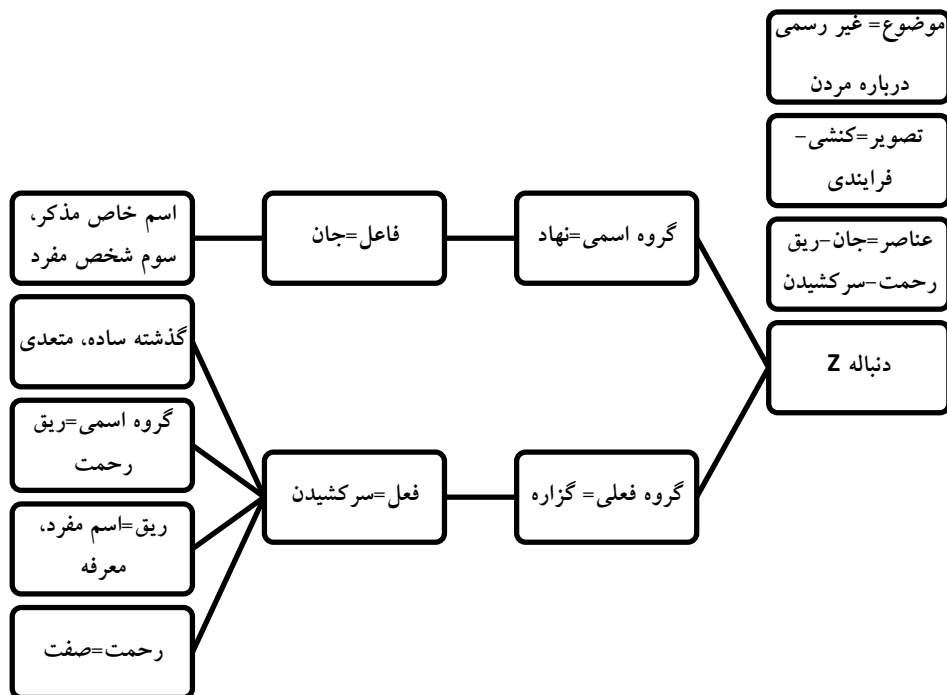
شکل ۶- دنباله Z₁ به زبان انگلیسی برگرفته از سیستم پردازش بافت-آگاه

همان‌طوری که در شکل (۶) نشان داده می‌شود، دنباله Z₁ یک ساختار مفهومی آماده برای ترجمه و برابرگزینی به زبان مقصد است که بر اساس سیستم بافت-آگاه و مطابق بافتِ زبان مبدأ شکل یافته است و با دنباله Z₂ (شکل ۷) یعنی ساختار ترجمه‌شده به زبان مقصد همپوشی یا برازش^۱ دارد. سپس، برای ترجمه ساختار Z₁ به ساختار Z₂، برابرگزینی واژگانی باید مطابق با اطلاعات بافتِ زبان مبدأ، صورت گیرد. به نحوی که به جای اینکه در ترجمه اصطلاح "to kick the bucket" به جای اینکه واژگان تک تک ترجمه شوند، ابتدا کاربرد موضوعی اصطلاح در زبان فارسی نیز واکاوی می‌شود (بافت ضمنی). در زبان فارسی برای موضوع «مردن» اصطلاحات متعددی وجود دارد مانند «دار فانی را وداع گفتن»، «غزل خداحافظی را خواندن»، «رخت بر بستن»، «دست از جهان شستن»، «ریق رحمت را سرکشیدن»، «جان به جان آفرین تسلیم کردن» و غیره. اما ترجمه‌ای را سیستم بافت-آگاه در این مرحله انتخاب می‌کند که مطابق اطلاعات بافت، هم با کاربرد موضوعی و فضای تصویری اصطلاح یعنی بیانِ مردن کسی در سیاق غیررسمی و کهن

۱. fitting

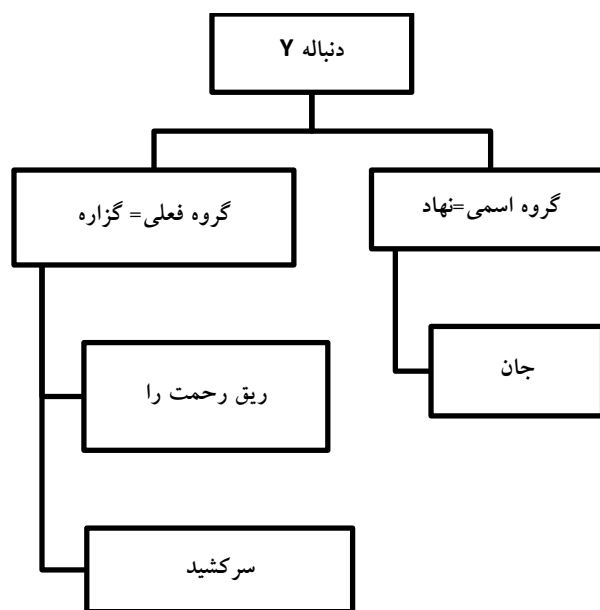
(بافت ضمنی) و هم با عناصر سازنده اصطلاح (بافت صریح) همخوانی داشته باشد. بنابراین، در بین موارد پیشنهادی بالا، نزدیک‌ترین گزینه برابر با توجه به موارد مورد نظر، ترجمه «ریق رحمت را کشیدن» برای "to kick the bucket" است.

بدین نحو، تبدیل دنباله Z_1 یعنی الگوریتم شکل (۵) به Z_2 پس از معادلیابی به زبان فارسی به صورت ذیل نشان داده می‌شود (شکل ۷):



شکل ۷- الگوریتم دنباله Z_2 به زبان فارسی

در نهایت، دنباله Z_2 متناسب با دستور زبان فارسی به دنباله Y (جمله ترجمه شده به زبان مقصد) مرتب می‌شود و به عنوان خروجی نهائی سیستم، توزیع می‌گردد، (شکل ۸):



شکل ۸- دنباله Y خروجی نهائی سیستم بافت-آگاه به زبان مقصد

و ترجمه پیشنهادی و تاییدشده مدل مفهومی بافت-آگاه برای جمله انگلیسی:

John kicked the bucket.

جمله «جان ریق رحمت را سرکشید.» است.

همان طوری که پیشنهاد می شود مدل مفهومی ترجمه ماشینی بافت-بنیان بر اساس سیستم بافت-آگاه یک تابع^۱ از X به Y است. اولین مرحله پردازش، تجزیه جمله ورودی به ساختاری قابل پردازش برای ماشین ترجمه است یعنی تبدیل جمله ورودی به دنباله X. سپس اطلاعات بافتی این دنباله طی چهار مرحله پردازش سیستم بافت-آگاه که عبارتند از: کسب، مدل سازی، استنتاج و توزیع بافت مشخص می گردند. خروجی پردازش شده سیستم بافت-آگاه به صورت دنباله های Z_1 و Z_2 مرتب می شود. بعد از این مرحله، دنباله Z_2 که ترجمه پیشنهادی سیستم بافت-آگاه برای جمله ورودی است وارد مرحله نهائی پردازش می شود که مطابق دستور زبان مقصد به شکل دنباله Y تنظیم می گردد.

۵- نتیجه‌گیری

سیری در تحولات علم فناوری اطلاعات از آغاز با محاسبات مبتنی بر کاراکتر تا محاسبات عبارت-محور و اکنون بافت-آگاه، در ترجمه ماشینی نیز گذار از پردازش‌های صورت-محور یعنی کلمه-بنیان تا بافت-بنیان را به ذهن متبادر می‌نماید. بنابراین، غایت منظور محقق ارائه مدلی مفهومی و پیشنهادی برای ترجمه ماشینی است که هم جنبه‌های زبانشناسانه ترجمه بافت-بنیان را در خود جای داشته باشد و هم ویژگی‌های فناورانه بافت-آگاه را. برای تحقق این منظور، محقق به زبان ساده، مدلی مفهومی را برای تکمیل رویکرد ترجمه ماشینی عصبی طراحی کرده است که ویژگی‌های بافت-بنیان گفتمان با توجه به انواع صریح و ضمنی، در ترجمه نمایان شوند. و از طرفی سیستم بافت-آگاه را که یک مدل پردازشگر زبان طبیعی را نیز برای ترجمه ماشینی طراحی نماید. این مدل ضمن ارائه رهنمونی برای دستیابی به ترجمه‌ای صحیح، می‌تواند نقاط ضعف ترجمه ماشینی به‌ویژه رویکرد ترجمه ماشینی عصبی را نیز در خصوص ترجمه کلمات همخط، واژگان و عبارات مبهم، عبارات فعلی، اصطلاحات و استعارات رفع کند و در ارتقا و بهبود آنها مؤثر باشد. بدین نحو می‌توان اذعان داشت که با این مدل پیشنهادی، برابرگزینی مناسب‌تری برای کلمات و جملات زبان مبدأ انجام می‌گیرد و می‌توان ترجمه دقیق‌تری از متن زبان مبدأ ارائه نمود. در خاتمه، شایان ذکر است که کارایی چنین مدل با همت فناورانی مشخص و میسر خواهد شد که این مدل پیشنهادی را به مدلی فناورانه تبدیل کنند تا به‌طور عملیاتی قابل اجرا باشد.

منابع:

ترکاشوند، فرشید و قائمی، نجمه (۱۳۹۶). چالش‌های بافت فرهنگی در ترجمه از فارسی به عربی: با تمرکز بر گلستان سعدی و روضة الورد. پژوهش‌های ترجمه در زبان و ادبیات عربی، ۱۷ (۲۷)، ۷۹-

۹۸.

خان‌جان، علیرضا (۱۳۹۵). مسئله بافت در ترجمه. فصلنامه مترجم، ۲۴ (۵۸)، ۷۶-۹۳.

قنبریان شیاده، مریم، ذوالفقاری، محسن و حیدری، حسن (۱۴۰۱). خوانش نشانه‌شناسی لایه‌ای در فرایند

معنایی داستان مدرن «ابوالهول» از فرشته ساری. متن پژوهی/ادبی، ۲۶ (۹۳)، ۹۷-۱۲۲.

نیلی پور، رضا (۱۳۷۸). کاربرد اصطلاحها و تعبیرها در ترجمه. تهران: دانشگاه پیام نور.

Alzeebaree, Y. (2020). Lexical and Structural Ambiguity in Machine Translation: An Analytical Study. *Eastern Journal of Languages, Linguistics and Literatures*, 1(1).

Ashengo, Y. A., Aga, R. T., & Abebe, S. L. (2021). Context based machine translation with recurrent neural network for English–Amharic translation. *Machine Translation*, 1–18.

Bender, E. M., & Good, J. (2020). A grand challenge for linguistics: Scaling up and integrating models. *White paper contributed to NSF's SBE*, 1–1.

Chomsky, N. (1957). *Syntactic Structures*. The Hague: Mouton.

Chomsky, N. (1965). *Aspects of the theory of syntax*. M.I.T. Press.

De Matos, E., Tiburski, R. T., Moratelli, C. R., Johann Filho, S., Amaral, L. A., Ramachandran, G., Krishnamachari, B., & Hessel, F. (2020). Context information sharing for the Internet of Things. *Computer Networks*, 1(16). <https://doi.org/10.1016/j.comnet.2019.106988>

Dey, A. K., & Häkkinen, J. (2008). Context-awareness and mobile devices. User interface design and evaluation for mobile technology, 1, 205–217.

França, R. P., Monteiro, A. C. B., Arthur, R., & Iano, Y. (2021). An Overview and Technological Background of Semantic Technologies. *Advanced Concepts, Methods, and Applications in Semantic Computing*, 1–21.

Gollagi, S.G., Math, M.M. & Daptardar, A. (2020). A survey on pervasive computing over context-aware system. *CCF Trans. Pervasive Comp. Interact*, 2(1), 79–85. <https://doi.org/10.1007/s42486-020-00030-6>

Idlibi, D. (2019). *The importance of context in translation. A study of some selected English and Arabic terms and concepts*. GRIN Verlag.

Joshi, P., Santy, S., Budhiraja, A., Bali, K., & Choudhury, M. (2020). The state and fate of linguistic diversity and inclusion in the NLP world. *arXiv preprint arXiv:2004.09095*

Kenny, D. (2019). Machine translation. In *Routledge Encyclopedia of Translation Studies* (pp. 305–310). London: Routledge.

Koehn, P. (2009). *Statistical machine translation*. Cambridge University Press.

Koehn, P. (2017). Neural machine translation. *arXiv preprint arXiv:1709.07809*

Larson, M. L. (1984). *Meaning-based translation: A guide to cross-language equivalence*. University press of America.

- Lazidis, A., Petrakis, E.G.M., Chouliaras, S., Sotiriadis, S. (2022). Open-Source Publish-Subscribe Systems: A Comparative Study. In: Barolli, L., Hussain, F., Enokido, T. (eds) *Advanced Information Networking and Applications. AINA 2022. Lecture Notes in Networks and Systems*, vol 449. Springer, Cham. https://doi.org/10.1007/978-3-030-99584-3_10
- Moussallem, D., Wauer, M., & Ngomo, A.-C. N. (2018). Machine translation using semantic web technologies: A survey. *Journal of Web Semantics*, 51, 1–19.
- Nugroho, A. B. (2016). Meaning and translation. *JEE, Journal of English and Education*, 1(2).
- Poslad, S. (2009). *Ubiquitous computing: Smart devices, environments and interactions* (1st ed.). Wiley Publishing, p. 491.
- Pradeep, P., & Krishnamoorthy, S. (2019). The MOM of context-aware systems. *Computer Communications*, 1(137):44–69. <https://doi.org/10.1016/j.comcom.2019.02.002>
- Pranowo, P. (2020). The role of context in the interpretation of pragmatic meaning. *Retorika: Journal Bahasa, Sastra, dan Pengajarannya*, 13(2).
- Sezer, O. B., Dogdu, E. & Ozbayoglu, A. M. (2018). "Context-Aware Computing, Learning, and Big Data in Internet of Things: A Survey," in *IEEE Internet of Things Journal*, vol. 5, no. 1, pp. 1–27, Feb. 2018, doi: 10.1109/JIOT.2017.2773600
- Song, L. (2010). The role of context in discourse analysis. *Journal of Language Teaching and Research*, (1) 6: 876–879. doi: 10.4304/jltr.1.6.876-879
- Stanlaw, J. (2020). Halliday, Michael. *The International Encyclopedia of Linguistic Anthropology*, 6(1), 1–4.

Developing a Conceptual Model of Context-Aware Machine Translation: An Approach Derived from Context-Based Translation^۱

Ebrahim Ezzati^۲

Abstract

In this study, a context-based translation consisting of "usage, image and constituents" was proposed to reflect the intended and cultural aspects of a source sentence more precisely. Accordingly, the study is concerned with considering the perspectives of the context-based translation in visualizing and developing a conceptual model of machine translation in which a context-aware system has been of concern. This model resembles a function from X to Y . That is, there is an X -sequence, a machine-friendly sequence, as point of departure. Next, the given sequence is processed based on four modules of a context-aware system as follows: context acquisition, context modelling, context reasoning and context distribution. The processed outputs of the context-aware system are transformed into two sequential sequences: Z_1 , a sequence arranged based on the source language context and Z_2 , the translated sequence of Z_1 into target language. Then, the translated sequence (Z_2) is converted into the Y -sequence, as the final proposed translation of the input. To enclose, the innovative aspects of the proposed model concerned a combined system processing based on both context-based and context-aware characteristics, not to mention the ending feedback sector receiving user's satisfaction rate promotes the quality level of offered translation.

Keywords: Context-aware system, context-based translation, discourse context, machine translation

^۱. This paper was received on 23.05.2022 and approved on 22.08.2022.

^۲ Academic Member, Department of Linguistics and Foreign Languages, Payame Noor University, Iran; email: e_ezzati@pnu.ac.ir