



معرفی سیستم آموزش تشخیص و تصحیح ابهام کلمات، برای افزایش کارایی مترجم متن فارسی به زبان اشاره فارسی با استفاده از آنتولوژی

محمد لطافت ابریکوه¹، عبدالرضا رسولی کناری²، محبوبه شمس¹

¹ دانشکده برق و کامپیوتر، دانشگاه صنعتی قم، قم، ایران.

² دانشکده برق و کامپیوتر، دانشگاه صنعتی قم، قم، ایران. (نویسنده مسئول) rasouli@qut.ac.ir

چکیده

مهم‌ترین عامل ارتباط بین انسان‌ها، توانمندی در برقراری ارتباط از روش‌های مختلف مانند گفتار، نوشتار و زبان اشاره می‌باشد. ناتوانی افراد کم‌شنوا و ناشنوا در برقراری ارتباط صمیمانه، هم‌چنین نبود آموزش مرسوم با توجه به نیازهای جامعه، باعث کناره‌گیری این افراد شده است. سیستم‌های مترجم متن به زبان اشاره به منظور برطرف کردن مشکلات برقراری ارتباط بین افراد ایجاد شده‌اند. به دلیل غیرساخت‌یافته بودن زبان فارسی و هم‌چنین مشکلات زبان اشاره فارسی استفاده از سیستم‌های مترجم سایر کشورها عملاً امکان‌پذیر نمی‌باشد. این پژوهش به دنبال معرفی سیستمی می‌باشد که متن فارسی را دریافت کرده و بعد از انجام پردازش زبان طبیعی لازم، آن را به زبان اشاره فارسی تبدیل کند. برای انجام این مهم با چالش‌های متعددی چون کمبود حرکات اشاره به نسبت تعداد کلمات موجود در زبان فارسی و هم‌چنین مشکل ابهام و جناس در برخی کلمات موجود در متون فارسی مواجه هستیم. به منظور رفع مشکلات، معماری سیستم‌های مشابه قابل استفاده نمی‌باشد. بنابراین معماری جدیدی پیشنهاد شده است که از ارتباط بین ماژول‌های برطرف کردن جناس و هم‌چنین ماژول تبدیل متن ورودی به دنیای ناشنویان، با استفاده از آنتولوژی طراحی شده استفاده می‌کند. استفاده از این معماری، نرخ کارایی سیستم مترجم متن فارسی به زبان اشاره فارسی را تا 95٪ بهبود بخشیده و باعث افزایش دقت سیستم مترجم تا سه برابر شده است.

اطلاعات مقاله

مقاله علمی - پژوهشی

دریافت: 05 اسفند 1396

پذیرش: 14 بهمن 1397

واژگان کلیدی:

زبان اشاره فارسی

آنتولوژی

ناشنویان

پردازش زبان طبیعی

تصحیح ابهام

Introduce detection and correction of ambiguity of words system education to increase the efficiency of translator from persian text to persian sign language using ontology

Mohammad Letafat Abrebekouh¹, Abdolreza Rrasouli Konari², Mahboobeh Shamsi¹

¹ Faculty of Electrical and Computer Engineering, Qom University of Technology, Qom, Iran.

² Faculty of Electrical and Computer Engineering, Qom University of Technology, Qom, Iran.

(Corresponding author) rasouli@qut.ac.ir

ARTICLE INFORMATION

Original research paper

Received: 24 February 2018

Accepted: 03 February 2019

Keywords:

Persian Sign Language

Ontology

Deaf

Natural Language Processing

Correction Ambiguity

ABSTRACT

The most important factor through which human beings can form their relationships and connections is using different methods such as speaking, writing and sign language. The point that deaf or semi-deaf people are ignored in communities is because of their disability in forming intimate relationships, and also the lack of common training according to what is needed in the society. The text to sign language translation systems have been created to resolve communication problems among hearing and deaf persons. Due to unstructured Persian language as well as the problems of Persian sign language, using translators' systems of other countries is not practicable. The present study seeks to design a system that receives Persian text as an input and changes it to Persian sign language after all the necessary natural language processing is done. In doing this matter, several challenges are seen on our way: a) The number of Persian sign language words which are few compared to the words in Persian language and b) The ambiguity and pun normally used in Persian language. So, similar systems' architecture could not be applied. Therefore, a new architecture is proposed that uses the relation between the removal pun modules and input text invert module to the deaf world by using designed ontology. Using this architecture, the efficiency of the Persian text to Persian sign language translator system has improved to 95% and increased the accuracy of the translator system by tripling.

1. مقدمه

جریان اصلی جامعه به دلیل عدم توانایی در برقراری ارتباط با جامعه ناشنوایی باعث گوشه‌گیری این افراد شده است. در حال حاضر سالانه 1400 ناشنوای عمیق در کشور متولد می‌شود که بیشتر این کودکان دارای خانواده شنوا هستند. به همین دلیل والدین در سال‌های ابتدایی زندگی کودک، به دلیل عدم آشنایی با زبان اشاره نمی‌توانند الگوی کاملی از زبان اشاره را به آن‌ها آموزش دهند. همانطور که از نام «زبان اشاره» پیداست، در زبان اشاره از اشارات و یا به عبارت دقیق‌تر از حرکات، حالات دست، چشم، و جهت ایستادن، برای انتقال منظور بهره برده می‌شود. چنانچه به طور مثال حرکت دهان یا حالت کلی صورت تغییر کند، معنای کلمه تولید شده نیز می‌تواند تغییر کند، یا به عبارت دیگر کلمه دیگری تولید می‌شود. همین اصل در مورد حرکات انگشتان و دست‌ها نیز صادق است [22]. زبان اشاره برخلاف تصور رایج، زبانی جهانی نیست و به شیوه‌های مختلفی طبقه‌بندی می‌شوند. علاوه بر این در زبان اشاره هم ما با گویش‌ها و الفاظ متفاوتی مواجهیم، یعنی زبان اشاره در واقع یک زبان واحد نیست و در هر کشوری و در هر ناحیه‌ای مختصات خود را دارد. حال با توجه به تکنولوژی‌های اطلاعاتی و ارتباطی جدید توجه به نیازهای ناشنوایان برای برقراری ارتباط با سایر افراد جامعه امر بسیار مهمی است که لازم است بدان توجه شود. سیستم مترجم یک زبان به زبان دیگر که کار ترجمه را به صورت خودکار انجام می‌دهد می‌تواند بسیار مفید باشد. یک سیستم مترجم ماشینی، نیازمند در نظر گرفتن و تطبیق شدن با هر زبانی است. در این سیستم، نیاز است تا جزئیات مربوط به هر دو زبان مبدا و مقصد در نظر گرفته شود تا بتوانیم بهترین رویکرد ممکن را در خصوص ترجمه‌ی ماشینی و همچنین تکنیک‌های پردازش زبان طبیعی انتخاب کنیم. با رشد چشمگیر اسناد منتشر شده در وب در سال‌های اخیر و نیاز اساسی به نگهداری، دسته‌بندی و بازیابی و پردازش آن توجه بیش از پیش به پردازش زبان طبیعی توسط رایانه افزایش یافته است [23].

1.1 زبان اشاره

زبان اشاره نوعی از زبان است که در آن ناشنوایان به جای استفاده از سیستم صوتی و یا به عبارتی گفتار، از دست‌ها برای سخن گفتن بهره می‌بریم. در واقع این دست‌ها هستند که به جای زبان، سخن می‌گویند [1]. زبان اشاره به ندرت توسط شنوا قابل درک می‌باشد، و این یک مانع بزرگ ارتباطی می‌باشد [12].

امروزه مکتب‌های آموزشی مختلفی برای ناشنوایان و کم‌شنوایان وجود دارد. گروهی بر این باورند که تضعیف زبان اشاره و تاکید بر روش‌های دیگر، مخالفت با نظام طبیعی ذهنی کودکان و تاکید بر روش‌های مصنوعی است و گروه دیگر معتقدند تاکید بر زبان اشاره انگیزه کودکان را برای یادگیری زبان گفتاری کاهش می‌دهد. برخی دیگر نیز بر این باورند که در آموزش به ناشنوایان نقش دست‌ها را نمی‌توان نادیده گرفت. از این رو زبان اشاره مهم‌ترین ابزار آموزشی ناشنوایان محسوب می‌شود [24]. زبان اشاره، یک زبان جهانی نیست، و از هر کشوری به کشور دیگر متفاوت است [13]. زبان اشاره فارسی از گویشی به گویش دیگر متفاوت است. زبان اشاره فارسی علی‌رغم افزایش تعداد ناشنوایان و استفاده از زبان اشاره، این زبان پیشرفت آن‌چنانی نداشته و دایره کلمات موجود در آن بسیار محدود است. با توجه به تکنولوژی‌های اطلاعاتی و ارتباطی جدید توجه به نیازهای ناشنوایان برای برقراری ارتباط با سایر افراد جامعه امر بسیار مهمی است که لازم است بدان توجه شود. سیستم مترجم یک فرآیند خودکار است که در آن یک زبان طبیعی به زبان طبیعی دیگری تبدیل می‌شود [30]. یک سیستم مترجم ماشینی، نیازمند در نظر گرفتن و تطبیق شدن با هر زبانی است. در این سیستم، نیاز است تا جزئیات مربوط به هر دو زبان مبدا و مقصد در نظر گرفته شود تا بتوانیم بهترین رویکرد ممکن را در خصوص ترجمه‌ی ماشینی و همچنین تکنیک‌های پردازش زبان طبیعی انتخاب کنیم [2].

2.1 پردازش زبان طبیعی

پردازش زبان‌های طبیعی راه بسیار جذابی برای ارتباط بین انسان و کامپیوتر است و در صورت عملی شدن آن به طور کامل می‌تواند تحولات شگفت‌انگیزی را در پی داشته باشد. کاربردهای پردازش زبان طبیعی را می‌توانیم به طرق مختلف دسته‌بندی کنیم، این کاربردها را در یک دیدگاه می‌توان به دو دسته کلی تقسیم کرد: کاربردهای نوشتاری و کاربردهای گفتاری.

- استخراج اطلاعات خاص از یک متن
- ترجمه یک متن به زبانی دیگر
- خلاصه‌سازی متن
- یافتن مستندات خاص در یک پایگاه داده نوشتاری
- از کاربردهای گفتاری می‌توان به موارد ذیل اشاره کرد:
- سیستم‌های پرسش و پاسخ انسان و رایانه
- سرویس‌های خودکار ارتباط با مشتری از طریق تلفن

از مضاف واقع می‌شود. البته این وضعیت در مورد صفات پیشین و حروف اضافه برقرار نیست. در این حالات هسته در انتها قرار می‌گیرد. به عبارت دیگر فارسی بهره‌گیر از حروف اضافه پیشین و صفات و اضافات پسین است [26].

4.1 مسائل و مشکلات پردازش زبان فارسی

برخلاف زبان انگلیسی که در آن هم حروف و هم لغات کاملاً متمایز از یکدیگرند، در زبان فارسی پیوستگی میان برخی علائم با لغات وجود دارد و علاوه بر آن تنوع نگارش در کلمات نیز موجود می‌باشد. ریشه‌یابی فعل که یکی از مراحل مهم پیش‌پردازش متن می‌باشد، در زبان فارسی چالش‌های خاص خود را دارد. به عنوان مثال در یک لغت به هم پیوسته هم بن فعل، شناسه، علامت زمان فعل و حتی شناسه‌های مفعولی می‌توان داشت که کار پردازش لغات را پیچیده‌تر می‌نماید به طوری که نمی‌توان از دانش، تجربه و نرم‌افزارهای موجود در این زمینه استفاده نمود و تولید نرم‌افزاری که قادر به حل تمامی این پیچیدگی‌ها باشد، فرایندی زمان بر و مستلزم تلاش فراوان می‌باشد. تفاوت‌های ذاتی زبان‌های گسسته‌ای مانند انگلیسی با زبان‌هایی مانند فارسی (عربی و غیره) که با یکدیگر تفاوت‌های بنیادین در قواعد دستوری دارند، منجر به آن شده است که ادعای اعمال تغییرات در ساختار یک نرم‌افزار انگلیسی و به دست آوردن نتایج خوب برای زبان فارسی لزوماً امکان‌پذیر نبوده و مستلزم آزمایش‌های فراوان برای اثبات صحت آن خواهد بود. [25]

5.1 چالش‌های ذاتی زبان فارسی

زبان فارسی در ساختار و قاعده با زبان انگلیسی متفاوت می‌باشد. برخی از مشکلات ذاتی مربوط به متون فارسی در ذیل طبقه‌بندی شده است [26].

- نبود نکات گرامری تعریف شده همانند آن چه که در زبان انگلیسی وجود دارد. این مساله در ریشه‌یابی و پیش‌پردازش متن تأثیرگذار خواهد بود.
- وجود لغات ترکیبی چند جزئی همانند "آب سرد کن"
- ابهام ساختاری به نحوی که یک کلمه می‌تواند معانی مختلف داشته باشد. به عنوان مثال، شیر سه معنی متفاوت دارد. شیر حیوان، شیر آب، شیر خوراکی.

- سیستم‌های آموزشی به دانش‌آموزان
- سیستم‌های کنترلی توسط صدا
- از دیدگاه دیگر می‌توان پردازش زبان طبیعی را به سطوح زیر تقسیم‌بندی کرد [25]:
- آواشناسی و صداشناسی که به تشخیص آواها و صداها و بازشناسی گفتار می‌پردازد.
- ریخت‌شناسی که به ساختارهای کلمات و ریشه‌یابی واژگان می‌پردازد.
- نحو که به ارتباط کلمات به همدیگر و مباحث دستوری آن‌ها در گروه‌ها و جملات می‌پردازد.
- معناشناسی که به ارتباطات معنایی کلمات می‌پردازد.
- کاربردگرایی که کاربردهای زبان برای رساندن یک مطلب به مخاطب یا مخاطبان، در حالت عملی و یا در نوشتار و گفتار طبیعی می‌پردازد.
- مباحثه که به ارتباطات کلی یک زبان فرای یک یا چند جمله خاص می‌پردازد.

3.1 جایگاه زبان فارسی

جایگاه زبان فارسی در میان زبان‌های دیگر را از سه جنبه‌ی وراثتی (تاریخی)، ناحیه‌ای و رده‌شناختی می‌توان بررسی کرد [3]: از دیدگاه زبان‌شناسی تاریخی، فارسی همراه با زبان‌های هند آریایی، زیر گروه هند - ایرانی را در گروه شرقی زبان‌های هند و اروپایی تشکیل می‌دهند. این زیر گروه شامل زبان‌هایی مانند فارسی، پشتو و کردی می‌باشد. از نظر ناحیه‌ای، به دلیل همسایگی با کشورهای عربی زبان، دارای بسیاری کلمات قرضی و حتی برخی قواعد مشابه با آن‌هاست. فارسی از دیدگاه ویژگی‌های زبانی (رده‌شناختی)، یک زبان پیوندی و ضمیرانداز است. فارسی از راست به چپ نوشته می‌شود و اگرچه در اصل دارای ترتیب فاعل - مفعول - فعل است ولی مملو از استثنائات مجاز در این ترتیب می‌باشد که حاصل فرایندهایی چون نامکانی، بهم ریختگی، حرکت جهت برجسته‌سازی، تاخیر، شکافت و شبه‌شکافت و غیره هستند، و به دلیل استفاده‌ی فراوان، عملاً فارسی را به یک زبان بدون ترتیب تبدیل می‌کنند [27]. در فارسی معمولاً فعل در انتها و هسته در ابتدا قرار می‌گیرد. مثلاً در عبارات وصفی موصوف قبل از صفت و در عبارات اضافی مضاف‌الیه قبل

روابط میان آن‌ها می‌پردازد. مفهوم آنتولوژی در وب معنایی کمی متفاوت از فلسفه است. آنتولوژی از دو واژه Onto به معنی هستی و Logia به معنی مطالعه به وجود آمده است و در کل معنی هستی‌شناسی دارد [4]. آنتولوژی در وب معنایی واژه‌ها و ارتباط بین آن‌ها در دامنه‌ای که استفاده می‌شود را نشان می‌دهد. عناصر اصلی تشکیل دهنده‌ی آنتولوژی عبارتند از: مفاهیم، ارتباط بین آن‌ها و خصوصیات آن‌ها.

آنتولوژی را خیلی کوتاه می‌توان به این صورت تعریف کرد: آنتولوژی، مشخص کردن و تعریف یک مفهوم سازی است. در حالی که کلمات مشخص کردن و تعریف مفهوم‌سازی باعث بحث‌های زیادی شده‌اند، نکته‌ی اساسی در مورد این تعریف از آنتولوژی موارد زیر هستند [5].

- یک آنتولوژی مفاهیم، ارتباطات و سایر مختصاتی که برای مدلسازی یک دامنه مورد نیاز هستند را تعریف می‌کنند.
 - تعریف یک شکل از تعاریف یک فرهنگ لغات نمایش (کلاس‌ها، ارتباطات و ...) را در بردارد که معانی را برای فرهنگ لغات و قیود رسمی برای استفاده‌ی همیشه از آن فراهم می‌کند. یک آنتولوژی لغات و مفاهیمی (معانی) که در تعریف و نمایش یک محدوده‌ای از دانش به کار می‌روند را تعیین می‌کنند و بنابراین معانی را استاندارد می‌کند.
- آنتولوژی‌ها توسط مردم، پایگاه‌های داده و برنامه‌های کاربردی که نیاز به اشتراک‌گذاری اطلاعات یک دامنه‌ی خاص دارند استفاده می‌شود. در زمینه‌ی وب آنتولوژی‌ها یک فهم مشترک از یک دامنه را تأمین می‌کنند. چنین فهم مشترکی برای حل مشکل چند معنایی لازم است. زیرا دو برنامه‌ی کاربردی ممکن است از دو لفظ متفاوت برای یک معنی واحد استفاده کنند و یا بالعکس از یک ترم واحد برای دو مفهوم متفاوت استفاده کنند. در واقع آنتولوژی‌ها همکاری معنایی را فراهم می‌کنند. به عبارت دیگر آنتولوژی، ارتباط بین مفاهیم در اسناد و دنیای واقعی را مشخص می‌کند که با این کار اسناد مربوطه توسط ماشین قابل پردازش می‌شود و اشتراک‌گذاری اطلاعات بین عامل‌ها را تسهیل می‌کند. در واقع می‌توان گفت [6]:

Vocabulary + Structure = Taxonomy

Taxonomy + Relationship, Constraints and Rules = Ontology

Ontology + Instance = Knowledgebase

عناصر مختلف آنتولوژی شامل موارد زیر است [28]:

- عدم وجود قاعده خاص برای تشخیص اسامی و مکان‌های خاص همانند آنچه در زبان انگلیسی موجود می‌باشد.
 - ابهام در معنی کلمه به علت نبود اطلاعات آوایی مانند «مرد» و «مرد»
 - عدم وجود دستورالعمل قطعی برای استفاده از نیم‌فاصله.
 - عدم وجود قواعدی ثابت برای فاصله‌گذاری ترکیبات.
 - ناآگاهی رایانه از نقش لغت در جمله همانند «آیین‌نامه نوشتن» و «آیین‌نامه رانندگی». از آن‌جا که در این ترکیبات، تمام کلمات به تنهایی معنی دارند، رایانه قادر به تعیین مرز لغات نمی‌باشد.
 - وجود کلمه‌های ترکیبی و امکان در نظر گرفته شدن دو کلمه مجزا. به عنوان مثال سیب‌زمینی کلمه مرکبی است که از دو کلمه سیب و زمینی تشکیل شده است.
 - فقدان یک ارزیاب اتوماتیک برای هر قسمت از سیستم‌های پردازش زبان طبیعی.
 - فقدان تجزیه‌کننده اتوماتیک ساختار کلامی متن که قسمت‌های پایه و پیرو متن را تشخیص دهد.
- از جمله مسائل دیگر در زبان فارسی نبود حروف کوچک و بزرگ همانند زبان انگلیسی برای تشخیص اسامی خاص (همانند نام روز، هفته، ماه و مناطق جغرافیایی) می‌باشد [25].

6.1 مواجهه با چند معنایی و چند نقشی بودن کلمات

برخی کلمات مانند شیر دارای چندین معنی هستند، که با توجه به بافتی که در آن واقع می‌شوند معنی آن‌ها مشخص می‌گردد. بعضی کلمات نیز مانند «در» و «چرا» علاوه بر چند معنی دارای چند مقوله‌ی نحوی یا نقش دستوری هستند. این ویژگی منجر به بالا رفتن سطح ابهام در متن می‌شود [25]. برای مقابله با این‌گونه مشکلات در زبان فارسی باید ارتباط بین کلمات را در آن جمله با استفاده از علم آنتولوژی به دست آورده و با توجه به آن جمله معنی کلمه با هم‌خانواده خود که دارای ابهام نیست جایگزین گردد؛ یا اینکه با ارائه توضیحی بیشتر برای آن کلمه ابهام آن را از بین برد.

7.1 آنتولوژی

آنتولوژی ریشه در فلسفه دارد و مبدا آن را ارسطو می‌دانند. در فلسفه، آنتولوژی شاخه‌ای از علم است که به بررسی موجودات و

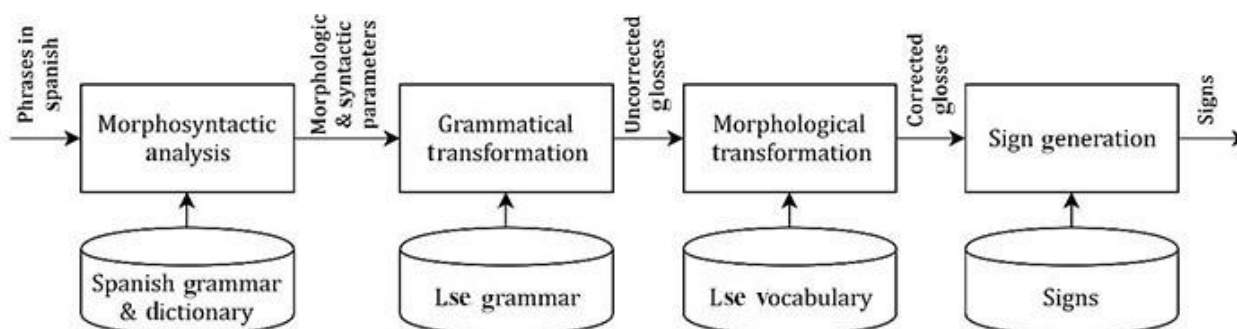
این سیستم از ابزار Xtag برای تجزیه جمله انگلیسی استفاده می‌کند، که در این ابزار یک دیکشنری برای کلمات انگلیسی گنجانده شده است که ملاک تجزیه جملات ورودی خواهد بود [7]. در سال 2009، در دانشگاه زاراگوزای اسپانیا پس از بررسی‌های انجام شده، سیستم ترجمی برای تبدیل خودکار ورودی در زبان اسپانیایی به زبان اشاره اسپانیایی با استفاده از مترجم مجازی ارائه کردند. این سیستم مترجم با استفاده از قوانین گرامری زبان اسپانیایی و با توجه به ویژگی‌های مورفولوژیکی و نحوی، تبدیل کلمات را انجام می‌دهد. منظور از ویژگی‌های مورفولوژیکی در این سیستم همان بررسی ویژگی‌های ظاهری و در اصطلاح ریخت‌شناسی کلمات تشکیل دهنده ورودی بود. این مترجم با یک موتور انیمیشن که در آن یک شخصیت مجازی به عنوان مترجم عمل می‌کرد ترکیب شده بود که در نهایت ورودی را به زبان اشاره اسپانیایی تبدیل می‌کرد. این شخصیت انیمیشنی برای انتقال حالت‌های جمله از ژست‌های خوشحالی، ناراحتی و غیره استفاده می‌کند. این سیستم مترجم زبان اسپانیایی به زبان اشاره اسپانیایی توسط ماژول‌هایی در زبان ++C نوشته شده بود. این سیستم ویژگی‌های نحوی، مورفولوژیکی و معنای کلمات را در نظر می‌گرفت. معماری این سیستم در شکل 1 آورده شده است.

- ماژول آنالیز مورفولوژیکی: این ماژول وظیفه استخراج روابط و وابستگی‌های نحوی را دارد.
- ماژول تبدیل گرامری: این ماژول وظیفه تغییر برخی بلوک‌ها و ترتیب کلمات با توجه به اطلاعات نحوی جمع‌آوری شده را دارد.
- ماژول مورفولوژیکال: وظیفه تبدیل کلمات به کلمه معادل در دنیای ناشنویان را دارا می‌باشد.

- نمونه‌ها: اشیای ابتدای یا نمونه‌های اولیه که مبنای ساخت آنتولوژی می‌باشد.
- کلاس‌ها: مجموعه‌ها، مفاهیم، انواع اشیای یا انواع چیزهای ابتدای تشکیل دهنده نمونه.
- خاصیت‌ها: جنبه‌ها، ویژگی‌ها، خصوصیات یا پارامترهایی که آن اشیای (کلاس‌ها) مربوط به اشیای می‌توانند داشته باشند.
- ارتباطات: روش‌هایی که در آن کلاس‌ها و نمونه‌هایی می‌توانند با یکدیگر ارتباط داشته باشند.
- جملات تابعی: ساختارهای پیچیده‌ای که از یک ارتباط مشخص شکل می‌گیرند، که می‌توانند به جای یک کلمه یا کلمه‌ی خاص یک مورد استفاده قرار گیرند.
- قیدها: توضیحاتی که به صورت رسمی بیان می‌شوند تا مشخص کنند که چه چیزی باید صحیح باشد تا این که یک حکم به عنوان ورودی مورد پذیرش قرار گیرد.

8.1 پیشینه پژوهش

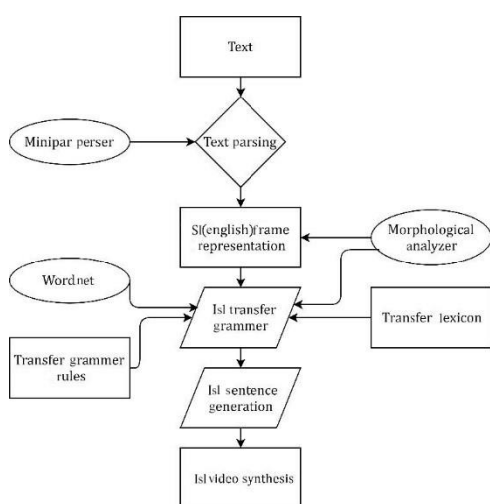
در سال 2003، چند تن از دانشجویان دانشگاه استلنبوش در آفریقای جنوبی یک سیستم مترجم ماشینی معرفی نمود که متن به زبان انگلیسی را دریافت کرده و آن را به زبان اشاره آفریقای جنوبی تبدیل می‌کرد را معرفی نمود. این سیستم متن به زبان انگلیسی را دریافت کرده و با استفاده از دیکشنری زبان مبدا تمامی انواع کلمه را شناسایی می‌کند. در مرحله بعد با استفاده از دیکشنری زبان مبدا-مقصد و هم‌چنین قوانین گرامری آن را به زبان مقصد تبدیل می‌کند. در این مرحله معتبر بودن جمله در زبان اشاره بررسی می‌شود و در نهایت در صورت تائید اعتبار جمله، ماژول گرافیکی آن را به زبان اشاره آفریقای جنوبی تبدیل می‌کند.



شکل 1. معماری تبدیل زبان اسپانیایی به زبان اشاره اسپانیایی

Fig. 1. Architecture of translation system of Spanish to Spanish sign language

این سیستم برای نمایش خروجی برنامه که زبان اشاره بنگال می‌باشد از ویدئوهای از قبل تهیه شده استفاده می‌کند. این مترجم به دنبال استانداردسازی Basl در تمام مناطق بنگلادش می‌باشد در سال 2010، آقایان داسگوپتا و باسو در هندوستان به منظور تبدیل اتوماتیک متن به زبان اشاره هندی یک چارچوب برای انتقال متن انگلیسی به زبان اشاره هندی ارائه دادند که این روش از قوانین گرامری انتقال برای زبان اشاره هندی به منظور تولید جمله بهره می‌برد. این سیستم که Ingit نامیده می‌شود یک سیستم ترجمه برای دامنه جملات استفاده شده در راه آهن می‌باشد. در این سیستم جمله ورودی ابتدا پیش‌پردازش و آنالیز می‌شود تا وابستگی‌های آن شناسایی شود. سپس با استفاده از ماژول مربوط به قواعد دستوری، قواعد دستوری جمله ورودی ارزیابی و شناسایی می‌شود. ماژول بعدی وظیفه انتقال ساختار جمله ورودی به زبان اشاره هندی را دارا می‌باشد. در نهایت خروجی مرحله قبل در قالب ویدئویی ضبط شده نمایش داده می‌شود. معماری این سیستم در شکل 3 نمایش داده شده است [9].



شکل 3. معماری سیستم مترجم Ingit

Fig. 3. Architecture of Ingit translation system

به منظور ارزیابی این سیستم از کارشناسان زبان‌شناسی برای درک مفهوم خروجی استفاده شده است. نتایج اعلام شده از سوی این کارشناسان در مجموع قابل قبول بوده است.

در سال 2010، محققان دانشگاه Iit هندوستان، نمونه اولیه سیستم تبدیل متن انگلیسی به زبان اشاره هندی را معرفی کردند. این سیستم می‌توانست با بهبود و شکل‌گیری برای مردم ناشنای هند

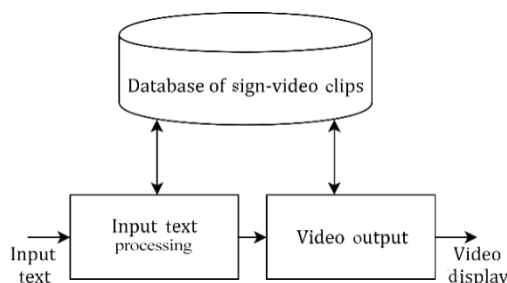
• ماژول تولید اشاره: وظیفه نمایش جمله ساخته شده در مراحل قبل را دارد.

ارزیابی این مترجم زبان اسپانیایی به زبان اشاره اسپانیایی توسط دو معلم مدرسه که از مترجم‌های زبان اشاره بودند، انجام شد. دقت و سرعت نمایش حرکات در جمله‌هایی که به صورت آزمایشی به سیستم داده شده بود بررسی شد. در مجموع 92 جمله که شامل 561 کلمه بود توسط سیستم مترجم مورد آزمایش قرار گرفت [8].

در سال 2009، عده‌ای از دانشجویان بنگلادشی یک سیستم مترجم برای تبدیل متن بنگلادشی به زبان اشاره بنگال ایجاد کردند. آن‌ها برای کاهش موانع ارتباطی بین افرادی که دارای اختلال شنوایی هستند با افراد عادی جامعه، نرم‌افزار مترجمی را توسعه دادند که متن ورودی را به زبان بنگال دریافت کرده و به طور خودکار به زبان اشاره بنگالی (Basl) که توسط انجمن بنگالی در سراسر بنگال غربی ایجاد شده است تبدیل می‌کند. فرهنگ لغت، که مبنای کار مترجم است، در حدود هزار کلمه را

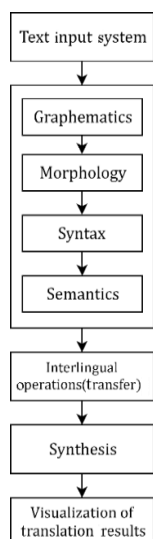
شامل می‌شود که این کلمات، عمدتاً ساده هستند و از کتاب‌های درسی کودکان جمع‌آوری شده‌اند. روش کار این مترجم اینگونه است که کلمه یا جمله ساده در بخش مربوطه وارد شده و پردازش‌ها و جایجایی‌های لازم در ساختار کلمه یا جمله صورت می‌گیرد و در نهایت اشاره معادل آن کلمه یا جمله نمایش داده می‌شود. معماری این مترجم در شکل 2 آورده شده است. معماری این مترجم شامل سه ماژول اصلی است که عبارتند از:

1. ماژول پردازش متن ورودی
2. ماژول پایگاه‌داده‌هایی از کلیپ‌های تصویری
3. ماژول خروجی ویدئویی



شکل 2. معماری تبدیل متن به زبان اشاره بنگلادشی

Fig. 2. Architecture of text to basl translation

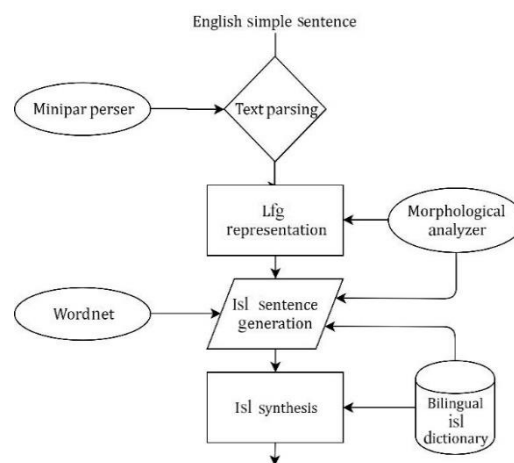


شکل 5. معماری سیستم تبدیل زبان روسی به زبان اشاره روسی
Fig. 5. Architecture of Russian text to rsa translation system

در مرحله تحلیل گرافیکی عناصر متن ورودی از هم جداسازی می‌شود. در مرحله آنالیز مورفولوژیکی از روش‌های مبتنی بر دیکشنری گرامری زبان روسی که حاوی کلمات کلیدی این زبان است بهره می‌برد. در تحلیل نحوی، ارتباط بین کلمات تعریف و بررسی می‌شود که در این تحلیل از تحلیل معنایی نیز استفاده می‌شود. روش اصلی تجزیه و تحلیل نحوی و معنایی آن‌ها، درخت وابستگی متن ورودی می‌باشد. در ماژول بعدی با استفاده از درخت وابستگی مرحله قبل، متن ورودی به زبان اشاره روسی (Rsa) تبدیل می‌شود. در ماژول بعدی ساختار جمله در زبان اشاره روسی مورد بررسی قرار می‌گیرد. در نهایت توسط ماژول تصویرسازی به خروجی زبان اشاره روسی تبدیل می‌شود.

در سال 2014، در آزمایشگاه دانشگاه Cochon در ایالت کوچی هندوستان نمونه اولیه از یک مترجم زبان مالایی به زبان اشاره مالایی معرفی شد. این سیستم پیشنهادی متن به زبان مالایی را به عنوان ورودی دریافت کرده و زبان اشاره متناظر با آن را تولید می‌نماید. زبان اشاره خروجی در این سیستم توسط انیمیشن و با استفاده از مدل‌های تولیدی کامپیوتر عرضه می‌شود. سیستم پیشنهادی آن‌ها با گرفتن متن به عنوان ورودی، کار را شروع می‌کرد. در مرحله اول، برچسب زنی اجزای کلام را انجام می‌دهد. در طول این مرحله، اجزای مختلف سخن شناسایی می‌گردد. سپس نتیجه این مرحله به قسمت بهینه‌ساز سپرده می‌شود. ماژول بهینه‌ساز کلمات نامطلوب را حذف می‌کند. خروجی این ماژول، ورودی بخش ریشه‌یاب می‌باشد که وظیفه یافتن ریشه کلمه را

کمک شایانی باشد. این سیستم جمله انگلیسی را در قالب متن دریافت کرده و پس از تجزیه و تحلیل نحوی و ایجاد ساختار مناسب به زبان اشاره هندی مربوط به متن ورودی تبدیل می‌کرد. از آن جایی که زبان اشاره هندی هیچ شکل نوشتاری ندارد خروجی این برنامه در قالب ویدئوهای از قبل آماده شده (ضبط شده) نمایش داده می‌شود. سیستم جاری برای انتقال متن به زبان اشاره هندی از سیستم انتقال مبتنی بر معماری بهره می‌برد. همان‌گونه که بیان شده این سیستم نمونه اولیه از یک سیستم بزرگ می‌باشد و تنها جملات ساده انگلیسی را به عنوان ورودی دریافت کرده و زبان اشاره هندی آن را نمایش می‌دهد. این سیستم از معماری خاصی برای انجام این کار استفاده می‌کند که در شکل 4 آن را مشاهده می‌کنید.



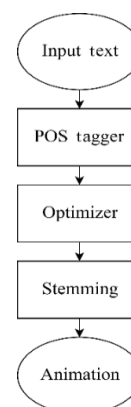
شکل 4. معماری سیستم تبدیل متن هندی به زبان اشاره هندی
Fig. 4. Architecture of text to isl translation system

در سال 2013، جمعی از دانشجویان دانشگاه Novosibirsk در روسیه سیستم تبدیل و ترجمه زبان روسی به زبان اشاره روسی را برای کسانی که دارای اختلالات شدید شنوایی هستند را معرفی نمود. در آن زمان گام‌های اولیه این سیستم بررسی شد و در مرحله بررسی قرار گرفت. ساختار کلی سیستم مورد نظر آن‌ها که ترجمه کامپیوتری زبان روسی را به زبان اشاره روسی (Rsa) انجام می‌داد به صورت شکل 5 تعریف شده بود:

ماژول‌های اصلی این ساختار عبارتند از:

- تفسیر متن منبع
- انتقال به زبان مقصد
- ترکیب زبان اشاره
- مشاهده نتایج با استفاده از زبان اشاره

دارد. در نهایت ماژول انیمیشن خروجی را نمایش می‌دهد. معماری این سیستم در شکل 6 آورده شده است.



شکل 6. معماری تبدیل زبان مالایی به زبان اشاره مالایی

Fig. 6. Architecture of Malayalam language to sign language

برای به تصویر درآوردن زبان اشاره مالایی، سیستم از مدل سه بعدی که تولید آن به صورت کامپیوتری بوده است، استفاده می‌کند. این سیستم به‌طور مستقیم ورودی را به زبان اشاره تبدیل می‌کند که این مسئله پیچیدگی تبدیل را کاهش می‌دهد [10]. به طور کلی در جدول 1 خلاصه‌ای از تمام سیستم‌های ارائه شده، همراه با مزایا و معایب هریک بیان شده است.

2. روش پژوهش

برای ایجاد یک سیستم مترجم از زبانی به زبان دیگر دو روش وجود دارد [14]:

جدول 1. خلاصه از سیستم‌های پیشین

Table 1. A summary of past work

Title	Years	Method	Advantages	Disadvantages
South African sign language machine translation system	2003	<ul style="list-style-type: none"> Pos tagging Syntactic tree 	<ul style="list-style-type: none"> Use animation marking 	<ul style="list-style-type: none"> Failure to identify the relationship between words Lack of identification semantic ambiguity
Automatic translation system to Spanish	2009	<ul style="list-style-type: none"> Feeling analysis 	<ul style="list-style-type: none"> Use animation High accuracy 	<ul style="list-style-type: none"> Lack of identification semantic ambiguity
An English to Indian machine translation system	2010	<ul style="list-style-type: none"> Convert text to sign language 	<ul style="list-style-type: none"> Proper architecture Detect the relationship between words 	<ul style="list-style-type: none"> Memory overhead Lack of identification semantic ambiguity
A prototype Malayalam to sign language automatic	2014	<ul style="list-style-type: none"> Pos tagging Animation 	<ul style="list-style-type: none"> Pos tagging Use animation Word stemming 	<ul style="list-style-type: none"> Lack of identification semantic ambiguity

نرمالسازی است قطعا انجام شده است و نتیجه این ماژول در حال تجزیه شدن می باشد.

- ماژول پیش پردازش
- ماژول بررسی ابهام و جناس
- ماژول نمایش اشارات

1.1.2 نرمال سازی

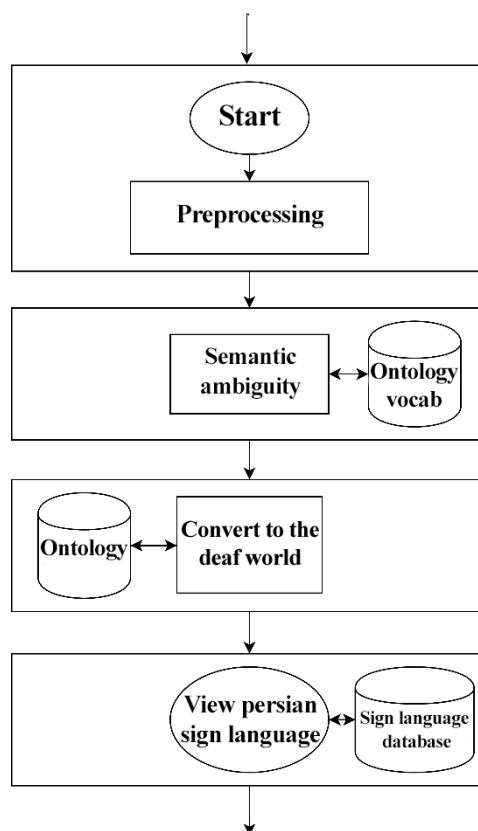
در اولین گام باید متون برای استفاده در گام های بعدی به شکلی استاندارد درآیند [16]. از آن جا که متون مختلف ممکن است بسیار به هم شبیه باشند اما به دلیل تفاوت های ساده ظاهری از نظر ماشین متفاوت ببینند؛ به همین دلیل سعی شده است این تفاوت های ساده ظاهری برطرف گردد. برای رسیدن به این هدف، قبل از مقایسه متون، پیش پردازش هایی روی آن ها انجام می شود. طبیعتا هر چه این پیش پردازش ها قوی تر باشد، نتایج حاصل از مقایسه متون قابل اطمینان تر خواهد بود [11]. لازم به ذکر است که از آن جایی که زبان فارسی جزو زبان های غیر ساخت یافته است با مشکلات بسیار بیشتری نسبت به سایر زبان ها مواجه خواهیم شد. متون غیر ساخت یافته، متونی هستند که پیش فرض خاصی در مورد قالب آن ها نداریم و آن ها را به صورت مجموعه ای مرتب از جملات در نظر می گیریم.

در پردازش رسم الخط زبان فارسی، با توجه به قرابتی که با رسم الخط عربی دارد، همواره در تعدادی از حرف ها مشکل وجود دارد که از جمله آن ها می توان به حروف «ک»، «ی»، و... اشاره نمود. در اولین گام باید مشکلات مربوط به این حروف را برطرف ساخت. برخی از این اصلاحات در ذیل آورده شده است [23]:

- بررسی همزه و انواع مختلف املاهای موجود و اصلاح هر کدام (به عنوان مثال تبدیل و به و، ی به ی، ا به ا، ا به ا و ...)
- حذف تشدید از واژه ها
- تبدیل ارقام عربی و انگلیسی به معادل فارسی
- اصلاح اعراب و حذف فتحه، کسره و ضمه و همچنین تنوین ها
- حذف نیم فاصله های تکراری
- حذف نویسه ی «ـ» که برای کشش نویسه های چسبان مورد استفاده قرار می گیرد. مانند تبدیل «بر» به «بر»
- چسباندن پسوندهای «تر»، «ترین» و ... به آخر واژه ها
- تبدیل «ۀ» به «هی»
- حذف فاصله ها و نیم فاصله های اضافه به کار رفته در متن

2.1.2 تجزیه متن به اجزا

Token کردن ابزاری برای شکستن متن ورودی بر اساس واحدهای با معنی مانند کلمه است [11]. نمادهای معناداری مانند Space، Tab و... لازمه ایجاد این ابزار در زبان فارسی می باشد؛ این کار به



شکل 7. معماری پیشنهادی

Fig. 7. Proposed architecture

1.2 ماژول پیش پردازش

ماژول پیش پردازش که اولین ماژول در معماری پیشنهادی سیستم ترجمه متن فارسی به زبان اشاره فارسی است خود دارای ماژول های زیر می باشد، که عبارتند از:

- نرمال سازی
- تجزیه متن به اجزا
- تجزیه متن به جملات
- استاندارد سازی
- ریشه یابی
- برچسب گذاری
- تجزیه متن

هریک از این ماژول ها که در ادامه در مورد آن ها صحبت خواهیم کرد به صورت گام به گام انجام می شود، یعنی به طور مثال زمانی که ماژول تجزیه متن در حال انجام است ماژول قبلی آن، که

7.1.2 تجزیه متن

این ماژول از دستورات گرامری زبان بهره می‌برد. ماژول پارسر باعث تشخیص گروه‌های تشکیل‌دهنده جملات ورودی می‌باشد. تجزیه و تحلیل جمله ورودی و شکستن آن به اجزای تشکیل‌دهنده آن مانند گروه‌های اسمی، فعلی، قیدی و غیره که می‌توان از این گروه‌بندی درخت تجزیه جمله ورودی را رسم کرد. حال کار ماژول پیش‌پردازش به اتمام رسیده است. خروجی کاربردی این ابزار در سیستم مترجم، کلمات تشکیل‌دهنده متن ورودی می‌باشد. ابزار پیش‌پردازش متن در سیستم مترجم، چندین جمله در قالب یک متن را دریافت می‌کند و خروجی آن نرمال شده کلمات تشکیل‌دهنده آن جملات می‌باشد [20].

2.2 ماژول تشخیص و تصحیح جناس کلمات

برای برطرف کردن ابهام و جناس بین کلمات از آنتولوژی استفاده می‌شود. به طور کلی دو روش خودکار و دستی برای ایجاد آنتولوژی استفاده می‌شود. روش‌های خودکار از روش‌های متفاوتی برای استخراج آنتولوژی از متون موجود استفاده می‌کنند و بیش‌تر این روش‌ها از تکنیک‌های پردازش زبان طبیعی استفاده می‌کنند. اکنون روش خودکاری که آنتولوژی قابل قبولی تولید کند، معرفی نشده است. بنابراین برای ایجاد آنتولوژی از روش‌های دستی و یا نیمه‌خودکار استفاده می‌شود مستقل از روش ایجاد آنتولوژی می‌توان از متدولوژی خاصی برای ایجاد آنتولوژی استفاده کرد. متدولوژی استاندارد برای ایجاد آنتولوژی وجود ندارد و بسته به کاربرد آنتولوژی از متدولوژی‌های متفاوتی استفاده می‌شود. در این بخش یکی از این متدولوژی‌ها که روند آن در شکل 8 نشان داده شده است، مطرح می‌شود. در روند ایجاد آنتولوژی به صورت بازگشتی عمل می‌شود، یعنی در هر مرحله ممکن است نیاز به ایجاد اصلاحات و بازگشت به مرحله قبلی باشد.

1.2.2 تشخیص محدوده

مراحل مختلف توسعه‌ی یک آنتولوژی را با تعریف دامنه و حوزه‌ی آن آغاز می‌کنیم. این کار شامل پاسخ‌گویی به سوالات پایه زیر است [29]:

- آنتولوژی در حال توسعه، چه دامنه‌ای را پوشش می‌دهد؟
- برای چه منظوری از این آنتولوژی استفاده می‌کنیم؟

منظور شناسایی واحدهای مستقل معنایی می‌باشد. سپس براساس انتخاب هرکدام از این واحدها متن بر اساس آن‌ها شکسته خواهد شد [18].

3.1.2 تجزیه متن به جملات

با استفاده از علامت‌های نظیر « . » ، « ، » ، « ! » ، « ؟ » ، « : » و به کارگیری برخی دستورات گرامری زبان فارسی و در نظر گرفتن برخی لغات آغازکننده، مرز جملات ورودی تشخیص داده می‌شود. این پردازش باتوجه به پایه بودن جمله در بسیاری پردازش‌های پیش‌رو بسیار دارای اهمیت می‌باشد.

4.1.2 استانداردسازی

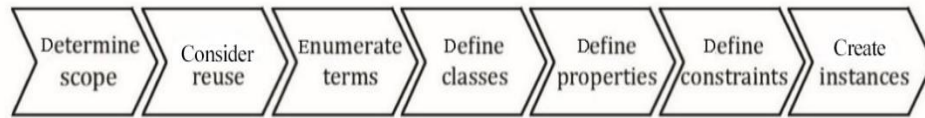
ساختار زبان فارسی به‌گونه‌ای است که شکل نوشتاری کلمات و شکل گفتاری آن‌ها با یکدیگر متفاوت است. تمام پردازش‌ها بر روی متون بر روی شکل نوشتاری کلمات صورت می‌گیرد. بنابراین در سیستم مترجم متن فارسی به زبان اشاره فارسی نیاز داریم که اگر متن ورودی به شکل گفتاری نوشته شده باشد آن را به معادل خود در شکل نوشتاری آن کلمه تبدیل نماییم. در ماژول Formal تمام پردازش‌های پیشین بر روی متن ورودی انجام شده است. این ماژول تمام کلمات و اصطلاحات عامیانه که در گفتار فارسی از آنان استفاده می‌شود را با شکل نوشتاری خود در زبان فارسی جایگزین می‌کند.

5.1.2 ریشه‌یابی

این ماژول ابزاری برای ریشه‌یابی لغات و تشخیص نوع کلمه ساخته شده از آن (اسم مکان، اسم زمان، حالت فاعلی، حالت مفعولی و غیره) می‌باشد. در این ماژول ما به تک‌واژه‌ها نگاه می‌کنیم و به کلمات اطراف آن لغت کاری نداریم در فارسی به این کار ریشه‌یابی لغوی می‌گویند [19].

6.1.2 برچسب‌گذاری

این ماژول ابزاری برای مشخص کردن نوع کلمات از قبیل: اسم، صفت، قید، فعل و... می‌باشد. به عمل برچسب‌گذاری واژگانی به کلمات و نشانه‌های تشکیل‌دهنده یک متن *Pos tagging* می‌گویند، به صورتی که این برچسب‌ها نشان‌دهنده نقش کلمات و نشانه‌ها در جمله می‌باشد.



شکل 8. متدولوژی ایجاد آنتولوژی

Fig. 8. Methodology for ontology creation

- تمایل داریم در مورد چه واژگانی صحبت کنیم؟
 - این واژه‌ها چه ویژگی‌هایی دارند؟
 - درباره این واژه‌ها چه می‌خواهیم بگوییم؟
- دو گام بعدی شامل توسعه سلسله مراتب کلاس‌ها و تعریف ویژگی‌های مفاهیم بسیار نزدیک به هم و در هم تنیده هستند. به گونه‌ای که تمایز میان آن‌ها و قائل به تقدم و تاخر شدن میان آن‌ها بسیار سخت است. به‌طور معمول با ایجاد تعریف برای تعدادی محدود از مفاهیم در سلسله مراتب آغاز می‌کنیم و سپس با توصیف ویژگی‌های آن‌ها (مفاهیم ایجاد شده) کار را ادامه می‌دهیم. این دوگام، مهم‌ترین گام‌ها در فرآیند طراحی آنتولوژی هستند.

4.2.2 تشخیص کلاس‌ها و ساختار آن‌ها

با استفاده از واژه‌های به دست آمده کلاس‌ها و زیرکلاس‌های آن‌ها را مشخص می‌کنیم. برای این کار می‌توان از روش پایین به بالا، بالا به پایین یا ترکیبی از این دو روش استفاده کرد.

- رهیافت بالا به پایین: فرآیند توسعه بالا به پایین با تعریف مفاهیم عام در دامنه آغاز می‌کند. سپس با ایجاد زیرکلاس‌های خاص‌تر از این مفاهیم، فرآیند توسعه را ادامه می‌دهد.

- رهیافت پایین به بالا: فرآیند توسعه پایین به بالا با تعریف کلاس‌های بسیار خاص آغاز می‌کند. برگ‌های سلسله مراتب کلاس‌ها سپس با گروه‌بندی این کلاس‌ها در قالب مفاهیم عام‌تر، فرآیند توسعه را ادامه می‌دهد.

- رهیافت ترکیبی: فرآیند توسعه ترکیبی، ترکیبی از روش‌های بالا به پایین و پایین به بالا محسوب می‌گردد. در این روش، ابتدا مفاهیم برجسته و حائز اهمیت تعریف می‌شوند. سپس با عمومی‌سازی (ترکیب) و یا خصوصی‌سازی (تجزیه) این مفاهیم، فرآیند توسعه ادامه می‌یابد.

هیچ یک از این سه روش، به صورت ذاتی از مابقی بهتر نیست. رهیافتی که جهت توسعه آنتولوژی اتخاذ می‌گردد، کاملاً به دیدگاه شخصی طراح، از دامنه وابسته است. اگر یک توسعه

- برای چه نوع مشکلاتی، اطلاعات آنتولوژی باید قادر به پاسخ‌گویی باشد؟
- باید به این نکته توجه داشت که پاسخ به این سوالات در فرآیند طراحی آنتولوژی ممکن است تغییر کنند. اما در هر حال، به محدودسازی حوزه‌ی مدل کمک خواهند کرد. به‌طور کلی الگوریتم مراحل تعیین و تثبیت دامنه مورد نظر برای آنتولوژی در شکل 9 آورده شده است.
- همان‌گونه که در الگوریتم مشاهده می‌کنید، تا زمانی که دامنه شناسایی شده برای آنتولوژی، پاسخ‌گوی تمام سوالات آن حوزه نباشد مراحل شناسایی دامنه ادامه پیدا می‌کند.

1. Algorithm determine the domain

2. start
3. Design questions related to determine the domain
4. Receive answers to questions for determine the domain
5. Set domain
6. For all query in domain
 - a. If donot answers the query
 - b. Goto 3
7. End

شکل 9. الگوریتم تعیین دامنه آنتولوژی

Fig. 9. Algorithm determine the domain

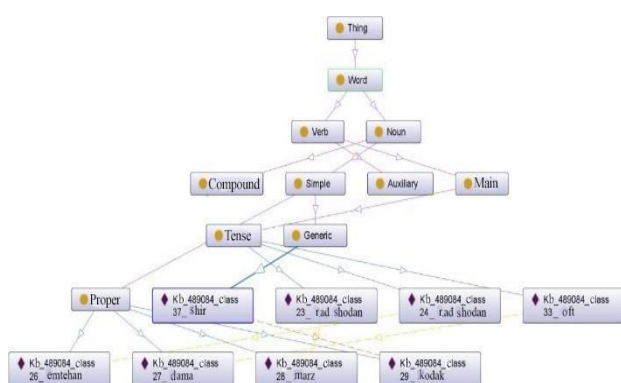
2.2.2 استفاده مجدد

در این مرحله مطالعه‌ی جامعی بر روی آنتولوژی موجود انجام و مشخص می‌شود که از کدام یک از آنتولوژی‌های موجود می‌توان برای ایجاد آنتولوژی مورد نظر کمک گرفت. استفاده‌ی مجدد این مزیت را دارد که صحت آنتولوژی‌های قبلی در عمل بررسی شده و از ابزارهایی که بر اساس آنها کار می‌کنند نیز می‌توان استفاده کرد [21].

3.2.2 تشخیص واژه‌ها

تهیه لیستی از تمامی واژه‌هایی که تمایل داریم برای آن‌ها جمله بسازیم یا در یک کاربرد درباره‌ی آنها توضیح دهیم نیز کاری مفید محسوب می‌گردد:

در ویژگی‌های کلمه شیر آورده شده است که این کلمه دارای ابهام و جناس می‌باشد، به منظور رفع جناس آن از کلمه هم‌معنی یا به اصطلاح معنی دوم این کلمه استفاده می‌شود. حال آنتولوژی در دامنه مورد نظر ایجاد شده است و آماده استفاده می‌باشد. برای تشخیص و تصحیح جناس با استفاده از آنتولوژی و جایگزینی کلمات جناس‌دار با معادل ساده و بدون جناس خود مراحل الگوریتم شکل 12 باید طی شود. به طور کلی خروجی این گام از معماری پیشنهادی، همان جمله ورودی است، با این تفاوت که تمام کلمات جناس‌دار و ابهام‌دار با معادل بدون جناس خود جایگذاری می‌شود.



شکل 11. بخشی از آنتولوژی طراحی شده

Fig. 11. Part of the ontology is designed

1. Algorithm recognition semantic ambiguity correction it
2. Input (text)
3. Output (correction semantic ambiguity text)
4. begin
5. Select domain
6. Create ontology for domain
7. For all word in text do
8. Tokeniz text & index word
9. While index word == ontology groups
10. if related words == related ontology groups
11. Replacement other sensus in groups for index word
12. Integrate words in text format
13. End

شکل 12. الگوریتم شناسایی و تصحیح ابهام و جناس

Fig. 12. Semantic ambiguity recognition & correction algorithm

3.2 تبدیل به دنیای ناشنوایان

زبان اشاره فارسی دارای الفبای مخصوص به خود ناشنوایان می‌باشد و اشارات تمام حروف الفبا برای آنان موجود است، به علاوه در زبان اشاره حرکات در نظر گرفته شده برای کلمات محدود است و تمام کلمات اشاره مستقیم ندارد. زبان اشاره کامل و بدون نقصی

دهنده، دیدگاه بالا به پایین و سیستماتیک به دامنه داشته باشد، در این صورت بهتر است از روش بالا به پایین بهره ببرد. اما به نظر می‌رسد که رهیافت ترکیبی برای بسیاری از توسعه‌دهندگان آنتولوژی‌ها راه‌حل ساده‌تری محسوب می‌گردد و تمایل بیشتری به استفاده از این روش وجود دارد. جدا از این که کدام یک از این رهیافت‌ها را مورد استفاده قرار می‌دهیم، معمولاً با تعریف کلاس‌ها آغاز می‌کنیم. از لیستی که در گام سوم ایجاد کرده‌ایم واژه‌هایی را انتخاب می‌کنیم که اشیاء مستقل را توصیف می‌کنند. این واژه‌ها «کلاس» را مشخص کرده و نقاط لنگرگاه در سلسله مراتب کلاس‌ها را تشکیل می‌دهند. به عبارت دیگر کلاس‌ها مجموعه‌ای از نمونه‌ها است که به توصیف دقیق موجودیت‌ها اشاره دارد. به طور کلی الگوریتم مراحل ساخت و ایجاد آنتولوژی در شکل 10 آورده شده است؛

1. Algorithm create ontology
2. Input (domain)
3. Output (ontology for domain)
4. begin
5. For all word in text do
6. Tokeniz text & index word
7. Property detection for words
8. Group by according to similar properties
9. Selected header group according to referrals maximum
10. Create sample for groups
11. End

شکل 10. الگوریتم ایجاد آنتولوژی

Fig. 10. Ontology construction algorithm

یکی از مهم‌ترین مراحل ایجاد یک آنتولوژی شناخت دامنه و حوزه‌کاری آن می‌باشد. با شناخت دامنه ایجاد آنتولوژی می‌توان واژه‌های محدود به آن حوزه را شناسایی کرد. پس از شناسایی واژه‌ها، با استفاده از یکی از رهیافت‌های ذکر شده، وخواص شناسایی شده برای آن‌ها، کلاس‌بندی انجام می‌شود. در گام بعدی اعضای هر یک از کلاس براساس میزان ارجاع به آن واژه شماره-گذاری می‌شود.

در شکل 11 آنتولوژی طراحی شده در مقیاس کوچک را مشاهده می‌کنید.

این آنتولوژی با استفاده از ابزار طراحی پرونچ ایجاد شده است. در این طراحی از رهیافت ترکیبی استفاده شده است. کلاس‌ها و زیرکلاس‌ها را مشاهده می‌کنید که در ساختار سلسله مراتبی قرار گرفته‌اند. در سطح آخر و برگ این ساختار نمونه‌ها قرار گرفته‌اند که در این جا برای مثال در سطح آخر کلمه "شیر" و کلمه کودک را مشاهده می‌کنید که با یکدیگر ارتباط داشته و به دلیل این که

موردنظر در آنتولوژی موجود باشد، پس از یافتن کلمه در آنتولوژی به سراغ یافتن کلاس مربوط به آن کلمه می‌رویم، زیرا در هر کلاس تنها سرگروه آن کلاس در دنیای ناشنوایان شناخته شده است و اشاره نظیر دارد، از این رو برای تبدیل کلمه به دنیای ناشنوایان، سرگروه کلمه موردنظر در کلاس را یافته و جایگزین کلمه موردنظر می‌شود. با این کار کلمه ورودی هم در دنیای ناشنوایان دارای اشاره معادل می‌باشد. حال اگر کلمه مورد نظر در آنتولوژی مورد نظر تعریف نشده باشد به آن معنی است که این کلمه در دنیای ناشنوایان فارغ از اشاره معادل می‌باشد، به همین دلیل برای نمایش آن در دنیای ناشنوایان از الفبای مربوط به زبان اشاره که در آنتولوژی تعریف شده است، استفاده می‌شود. لازم به ذکر است که برای انجام این عمل کلمه ورودی این بار به حروف تشکیل دهنده خود تجزیه می‌شوند. و در نهایت به صورت هجای حروف نمایش داده می‌شوند.

مراحل بیان شده برای انجام تبدیل متن ورودی به همتای خود در دنیای ناشنوایان را در الگوریتم شکل 13 نیز مشاهده می‌کنید.

1. Algorithm translator
2. Input(text)
3. Output(text in deaf word)
4. Preprocessing input text
5. Tokenize text
6. Compare word in ontology
7. If (word in ontology)
 - a. Replace word for word in deaf word
 - b. View sign language
8. Else
 - a. Parsing word for alphabet
 - b. View alphabet sign word
9. End

شکل 13. الگوریتم تبدیل گفتار فارسی به دنیای ناشنوایان

Fig. 13. Translator module algorithm

4.2 نمایش اشارات

برای نمایش زبان اشاره دو روش وجود دارد [15]، یکی استفاده از فیلم‌های از قبل تهیه شده و دیگری استفاده از انیمیشن می‌باشد. هر دو روش مشکلات و محاسن مربوط به خود را دارد. استفاده از فیلم‌های از قبل تهیه شده دارای سادگی در پیاده‌سازی و همچنین به دلیل استفاده از اشخاص حقیقی که دارای تسلط کافی به زبان اشاره می‌باشد دارای دقت و سرعت بالایی می‌باشد. اما به دلیل

نیست، همچنین از گویشی به گویشی دیگر دارای تفاوت می‌باشد. در ایران دو نوع زبان اشاره وجود دارد:

- زبان اشاره استاندارد
- زبان اشاره طبیعی

زبان اشاره استاندارد که در مدارس ناشنوایان از این زبان استفاده می‌شود، همچنین اخبار ناشنوایان بر اساس همین زبان بیان می‌شود دارای همان گرامر زبان فارسی است، و هرکلمه در همان مکان گرامری خود به زبان اشاره تبدیل می‌شود. با این تفاوت که تمام کلمات موجود در زبان فارسی اشاره معادل نداشته و باید از معادل اشاره‌دار کلمات استفاده شود. همچنین برخی کلمات هیچ‌گونه اشاره مستقیمی ندارد و برای نمایش آن دسته از کلمات در زبان اشاره از هجی کردن الفبای آن کلمه استفاده می‌شود. اما زبان اشاره طبیعی که قدمت طولانی‌تری نسبت به زبان اشاره استاندارد دارد دارای ساختار گرامری مخصوص به خود می‌باشد که با ساختار گرامری زبان فارسی متفاوت است. در سیستم مترجم متن فارسی به زبان اشاره فارسی از زبان اشاره استاندارد استفاده شده است. برای این که ماژول تبدیل به خوبی وظیفه خود را انجام دهد نیازمند داشتن اطلاعات از زبان و دنیای مقصد یعنی زبان اشاره و دنیای ناشنوایان دارد. زبان اشاره فارسی به دلیل عدم گسترش آن و همچنین عدم به‌روز کردن حرکات و اضافه کردن حرکات جدید به دایره حرکات اشاره به نسبت کلمات در گفتار فارسی بسیار محدود هستند، به همین دلیل ممکن است چندین کلمه در گفتار فارسی وجود داشته باشد که دارای یک اشاره معادل می‌باشد. همچنین به دلیل محدودیت اشارات در ایران برخی کلمات که معادل مستقیم یا غیرمستقیم برای آن وجود ندارد به ناچار برای نمایش آن از هجی کردن استفاده می‌شود. برای برطرف کردن این مشکل نیز به سراغ استفاده از آنتولوژی می‌رویم. دامنه این آنتولوژی کلمات فارسی می‌باشند که اشاره نظیر داشته و مستقیماً برای ناشنوایان قابل درک هستند. برای تبدیل کلمات به همتای خود در دنیای ناشنوایان از تکنیک استفاده از آنتولوژی استفاده می‌شود. آنتولوژی مربوط به ماژول تبدیل به مانند آنتولوژی مربوط به ماژول تشخیص و تصحیح جناس می‌باشد. برای استفاده از ماژول تبدیل متن از گفتار فارسی به همتای خود در دنیای ناشنوایان، در ابتدا متن ورودی پیش‌پردازش می‌شود. خروجی این کار تجزیه متن ورودی به کلمات تشکیل دهنده خود می‌باشد که ملاک تشکیل آنتولوژی می‌باشد. پس از تجزیه متن ورودی، بررسی می‌شود که هریک از کلمات در آنتولوژی مورد نظر تعریف شده است یا خیر؟ اگر کلمه

همان‌گونه که مشاهده می‌کنید، 42٪ جملاتی که برای تست سیستم مترجم در نظر گرفته شده‌اند دارای جناس می‌باشد. این جملات برای درک توسط سیستم مترجم و بالا بردن کارایی سیستم و کاهش خطا در روند ترجمه باید رفع ابهام شده و جناس آن‌ها برطرف شود. به عنوان نمونه در جمله «خبر رد شدن علی در امتحان مثل بمب ترکید.» کلمه «رد شدن» در این‌جا دارای جناس می‌باشد، این کلمه دارای دو مفهوم «عبور کردن» و «مردود شدن» می‌باشد. در این‌جا با توجه به کلمات تشکیل‌دهنده و ارتباطی که بین کلمه «رد شدن» و «امتحان» وجود دارد، مشخص می‌شود که در این جمله مفهوم «مردود شدن» مدنظر می‌باشد. برای همین منظور باید جمله نمونه به «خبر مردود شدن علی در امتحان مثل بمب ترکید.» تغییر کند؛ جمله نتیجه دیگر دارای ابهام و جناس نمی‌باشد.

اگر که سیستم تنها از ماژول تشخیص و رفع جناس در جملات برای ترجمه متن ورودی به زبان اشاره فارسی استفاده می‌کند، با این فرض تعداد جملات مجموع تست به 16 جمله کاهش می‌یابد. در این مجموع 13 جمله دارای جناس است که برای ترجمه به زبان اشاره فارسی و درک آن توسط ناشنوایان باید ابهام و جناس آن برطرف گردد. البته 3 جمله برای انتقال مفهوم، علاوه بر ماژول تشخیص و رفع ابهام و جناس نیازمند ماژول دیگری نیز می‌باشد. بنابراین تنها 10 جمله تنها با رفع ابهام و جناس آن توسط ناشنوایان قابل درک می‌باشد. همچنین 3 جمله بدون جناس می‌باشد، که این 3 جمله بدون استفاده از هیچ ماژول و پردازشی توسط ناشنوایان قابل درک می‌باشد. به طور کلی برای ارزیابی کارایی ماژول تشخیص و تصحیح ابهام و جناس در جملات، همچنین میزان تاثیر آن در دقت پردازش متن از معیار دقت و صحت که برای ارزیابی سیستم‌های بازیابی اطلاعات است، همچنین برای ارزیابی سیستم‌های که بصورت کیفی می‌باشند و برای ارزیابی نیازمند نتایج کمی هستیم از این نوع ارزیابی استفاده می‌کنیم. در نهایت به منظور ارتباط این دو معیار از میانگین هارمونیک آن دو که معیار F نامیده می‌شود بهره می‌بریم که کارایی سیستم مترجم خودکار متن فارسی به زبان اشاره فارسی را بررسی نمائیم.

معیار دقت: به حاصل تقسیم «تعداد مستندات بازیابی شده واقعاً» بر «تعداد کل مستندات بازیابی شده» گفته می‌شود.

$$\text{Precision} = \frac{10}{17} = 0.77 \quad (1)$$

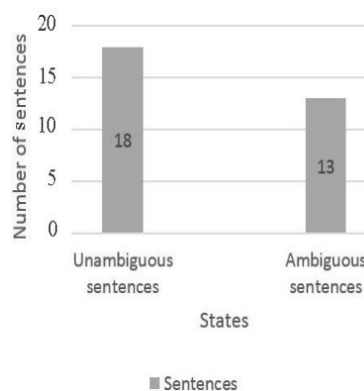
فضای ذخیره‌سازی دارای سرباز حافظه است. البته برای استفاده در یک دامنه محدود و کوچک استفاده از فیلم بسیار مناسب‌تر و بهتر است. استفاده از انیمیشن برای نمایش زبان اشاره دارای پویایی بیشتر می‌باشد و دچار سرباز حافظه نیز نمی‌باشد، اما پیاده‌سازی آن پیچیده و دشوار می‌باشد [17]. همچنین به دلیل نمایش آن توسط کاراکترهای انیمیشنی دقت پایین‌تری نسبت به حالت قبل دارد. این نوع نمایش برای استفاده در دامنه بزرگ مناسب‌تر است.

3. نتایج و بحث

برای ارزیابی سیستم مترجم متن فارسی به زبان اشاره فارسی از 31 جمله خبری که در اخبار ناشنوایان مورد استفاده قرار گرفته بود و ترجمه آن جملات با استفاده از مترجم زبان اشاره صورت گرفته است، استفاده شده است. لازم به ذکر است که انتخاب جملات با تأیید مربیان و مترجمین زبان اشاره نیز رسیده است. همچنین خروجی سیستم مترجم خودکار متن فارسی به زبان اشاره فارسی و خروجی هریک از ماژول‌ها، توسط مترجمین زبان اشاره فارسی بررسی و کارایی و دقت انتقال مفاهیم و کلمات به دنیای ناشنوایان ارزیابی شده است که در ادامه بیان می‌شود.

به منظور ارزیابی ماژول تشخیص و تصحیح ابهام کلمات، 13 جمله در مجموع تست وجود دارد، همچنین 18 جمله برای درک توسط ناشنوایان نیازمند استفاده از ماژول تبدیل می‌باشند.

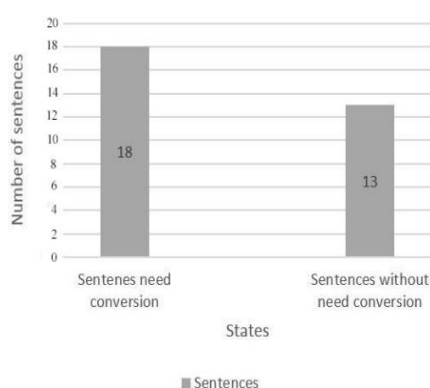
در نمودار شکل 14 نحوه قرارگیری جملات دارای جناس را در جملات تشکیل‌دهنده مجموعه تست سیستم مترجم مشاهده می‌کنید.



شکل 14. جملات مجموعه تست برای ماژول تصحیح ابهام

Fig. 14. Test set sentences for ambiguity correction module

به علاوه در جملات در نظر گرفته شده 18 جمله وجود دارد که قابل درک برای ناشنویان نیستند و نیازمند تبدیل برخی کلمات تشکیل دهنده آن جملات به معادل آن‌ها در دنیای ناشنویان داریم. این تبدیل کلمات به این دلیل انجام می‌شود که تمام کلمات در دنیای ناشنویان، معادل نداشته، قابل نمایش توسط زبان اشاره نمی‌باشد. در شکل 16 نحوه قرارگیری جملات نیازمند تبدیل را در مجموعه تست برای ارزیابی سیستم مترجم را مشاهده می‌کنید؛



شکل 16. جملات مجموعه تست برای ماژول تبدیل

Fig. 16. Test set sentences for convert module

همان‌طور که مشاهده می‌کنید 58٪ جملاتی که برای ارزیابی سیستم در نظر گرفته شده است، نیازمند تبدیل کلمات تشکیل دهنده آن‌ها به معادل خود در دنیای ناشنویان می‌باشد. برای مثال در جمله «پنج موقع را برای دعا و حاجت خواستن غنیمت شمارید» چندین کلمه این جمله در دنیای ناشنویان ناشناخته می‌باشد و نیازمند تبدیل آن کلمات به معادل خود در دنیای ناشنویان می‌باشد. برای مثال کلمات «موقع» و «غنیمت شماردن» معادل ندارد، این معادل‌گیری توسط آنتولوژی که ایجاد شده است انجام می‌گیرد، کلمه «موقع» با کلمه «زمان» و کلمه «غنیمت شماردن» با کلمه «ارزش نهادن» جایگزین می‌شود تا تمام کلمات تشکیل دهنده این جمله در دنیای ناشنویان شناخته شده باشد و توسط زبان اشاره قابل نمایش گردد. اگر این‌که در سیستم مترجم متن فارسی به زبان اشاره فارسی تنها از ماژول تبدیل جمله گفتار فارسی به هم‌تای خود در دنیای ناشنویان استفاده کند، را مورد بررسی قرار می‌دهیم. به طور کلی 21 جمله در مجموع تست قرار دارد، که 18 جمله آن توسط ناشنویان قابل درک نمی‌باشد و نیازمند استفاده از ماژول تبدیل هستند. لازم به

معیار صحت: به حاصل تقسیم «تعداد مستندات بازایی شده درواقع با ربط» بر «تعداد کل مستندات مرتبط موجود» گفته می‌شود.

$$\text{Recall} = \frac{13}{16} = 0.81 \quad (2)$$

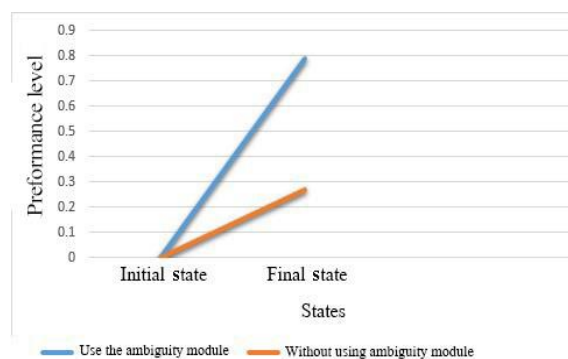
با بهره‌گیری از این دو معیار به ارزیابی کارایی استفاده از ماژول تشخیص و رفع جناس در جملات موجود در مجموعه تست با استفاده از متد F می‌پردازیم:

$$\text{F-measure} = \frac{2 \times 0.77 \times 0.81}{0.77 + 0.81} = 0.79 \quad (3)$$

براساس این محاسبات استفاده از ماژول تشخیص و تصحیح ابهام و جناس در متن کارایی سیستم مترجم متن فارسی به زبان اشاره فارسی به صورت خودکار را به 79٪ افزایش می‌دهد، یعنی در صورت استفاده از این ماژول دقت سیستم مترجم برای انتقال مفاهیم به دنیای ناشنویان و درک مفاهیم توسط مترجمین زبان اشاره به صورت خودکار تا 79٪ افزایش می‌یابد. این درحالی است که در صورت عدم استفاده از این ماژول همان‌گونه که در رابطه زیر مشاهده می‌کنید، کارایی 27٪ می‌باشد.

$$\text{F-measure} = \frac{2 \times 0.1875 \times 0.5}{0.1875 + 0.5} = 0.27 \quad (4)$$

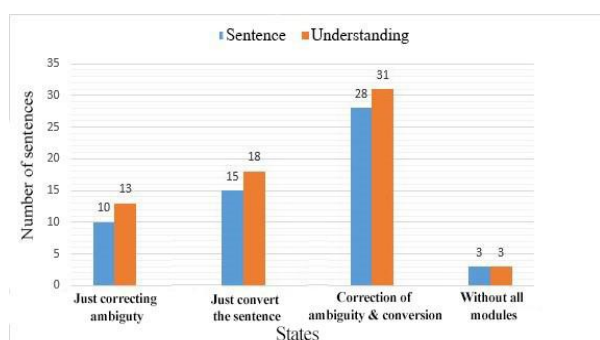
بنابراین استفاده از ماژول جناس به تنهایی کارایی سیستم مترجم متن فارسی به زبان اشاره فارسی به صورت خودکار را تقریباً 3 برابر افزایش می‌دهد. در نمودار شکل 15 میزان کارایی سیستم مترجم به منظور انتقال خودکار مفاهیم به ناشنویان و مترجمین زبان اشاره با استفاده از ماژول تشخیص و رفع ابهام و جناس هم‌چنین بدون استفاده از این ماژول را مشاهده می‌کنید.



شکل 15. میزان کارایی سیستم با استفاده از ماژول تصحیح ابهام

Fig. 15. System performance using ambiguity correction module

کارایی سیستم مترجم خودکار متن فارسی به زبان اشاره فارسی و میزان درک جملات در خروجی این سیستم توسط مترجمین زبان اشاره را با استفاده از مازول تبدیل خودکار متن ورودی به دنیای ناشنوایان و بدون استفاده از این مازول و دقت ترجمه خودکار سیستم را مشاهده می‌کنید. این ارزیابی بر مبنای مجموع جملات تست ارزیابی شده است. در صورت عدم برطرف کردن جناس و هم‌چنین عدم تبدیل جملات به دنیای ناشنوایان تنها 10٪ جملات برای افراد ناشنوایی که از زبان اشاره استاندارد استفاده می‌کنند، قابل درک خواهد بود. هم‌چنین با استفاده از سیستم مترجم که مشکل جناس کلمات را برطرف می‌کند، هم‌چنین تبدیل کلماتی که در دنیای ناشنوایان اشاره معادل ندارد با کلمه هم‌معنی خود، 90٪ باقی‌مانده جملات نیز توسط افراد ناشنوا قابل درک خواهد بود. به طور کلی در نمودار 18 تاثیر هریک از مازول‌های سیستم پیشنهادی در کارایی و دقت انتقال جملات انتخابی به ناشنوایانی که از زبان اشاره استاندارد استفاده می‌کنند مورد بررسی قرار گرفته است. همان‌گونه که در نمودار شکل 18 مشاهده می‌کنید



شکل 18. تاثیر مازول‌ها

Fig. 18. The impact of modules

برای نمونه اگر تنها در سیستم مترجم از مازول برطرف کردن جناس استفاده می‌شد، فقط 10 جمله بدون نیاز به مازول تبدیل و تنها به برطرف شدن مشکل جناس آن‌ها توسط ناشنوایان قابل استفاده می‌باشند، اما همان‌گونه که در نمودار مشاهده می‌کنید ستون نارنجی رنگ بیانگر این موضوع است که با استفاده از مازول برطرف کردن جناس در مجموع، 13 جمله برای ناشنوایان قابل درک می‌باشند، این نشان می‌دهد که 3 جمله بدون استفاده از هیچ یک از مازول‌ها قابل درک برای ناشنوایان می‌باشد. در استفاده از هر دو مازول، ستون آبی رنگ در نمودار بیانگر این موضوع است

ذکر است که 15 جمله تنها با استفاده از مازول تبدیل جمله توسط ناشنوایان قابل درک خواهد بود و 3 جمله نیازمند مازولی علاوه بر مازول تبدیل نیز می‌باشد. 3 جمله نیز بدون استفاده از هیچ‌گونه مازولی توسط ناشنوایان قابل درک می‌باشد. به طور کلی برای ارزیابی کارایی سیستم با استفاده از مازول تبدیل جملات ورودی از گفتار فارسی به همتای خود در دنیای ناشنوایان، هم‌چنین میزان تاثیر آن در دقت پردازش متن از روش‌های دقت و هم‌چنین صحت و در نهایت متد کارایی استفاده شده است.

$$\text{Precision} = \frac{15}{18} = 0.83 \quad (5)$$

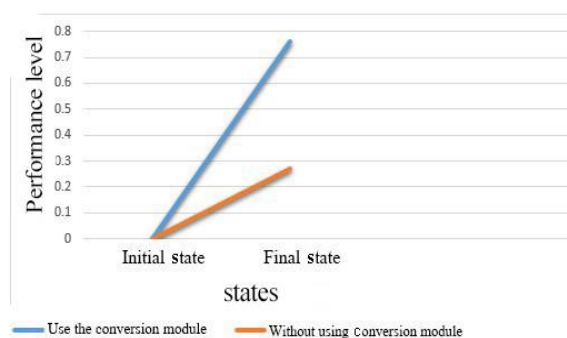
$$\text{Recall} = \frac{15}{21} = 0.71 \quad (6)$$

$$\text{F-measure} = \frac{2 * 0.83 * 0.71}{0.83 + 0.71} = 0.76 \quad (7)$$

بر اساس این محاسبات استفاده از مازول تبدیل جملات ورودی به دنیای ناشنوایان کارایی را به 76٪ افزایش می‌دهد، این در حالی است که در صورت عدم استفاده از این مازول همان‌گونه که در رابطه زیر مشاهده می‌کنید، کارایی 22٪ می‌باشد.

$$\text{F-measure} = \frac{2 * 0.14 * 0.5}{0.14 + 0.5} = 0.22 \quad (8)$$

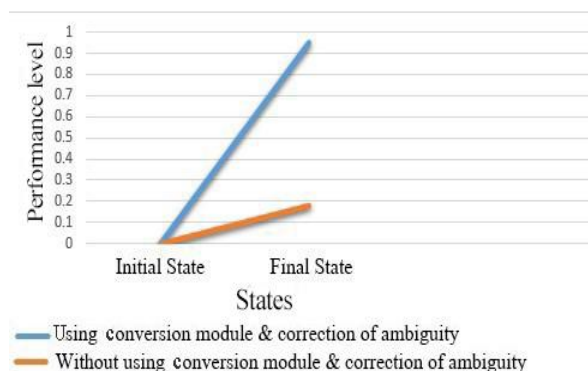
بنابراین استفاده از مازول تبدیل جملات ورودی به دنیای ناشنوایان به تنهایی کارایی سیستم مترجم متن فارسی به زبان اشاره فارسی را تقریباً 3.5 برابر می‌کند. این در حالی است که در صورت عدم استفاده از این مازول سیستم مترجم خودکار متن فارسی به زبان اشاره فارسی کلمات را به دنیای ناشنوایان تبدیل نمی‌کند و سیستم مترجم برای نمایش این کلمات با مشکل مواجه می‌شود. همان‌گونه که در شکل 17 نیز مشاهده می‌کنید.



شکل 17. میزان کارایی سیستم با استفاده از مازول تبدیل

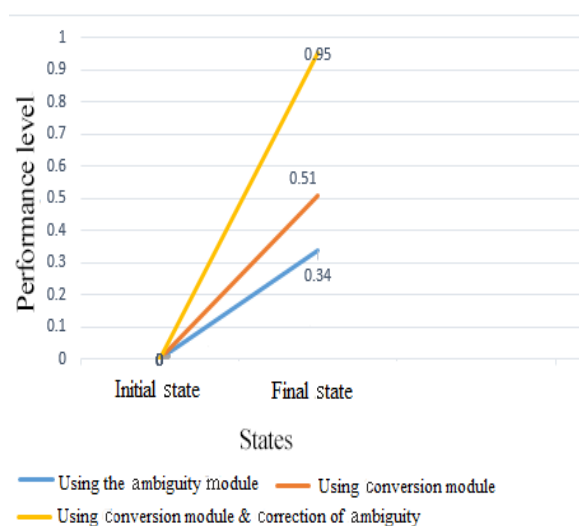
Fig. 17. System performance using convert module

میزان کارایی سیستم مترجم با استفاده از هریک از ماژول‌های تبدیل جملات ورودی به دنیای ناشنویان و همچنین تشخیص و رفع جناس کلمات و در نهایت ترکیب این دو ماژول می‌پردازیم.



شکل 19. میزان کارایی سیستم با استفاده از هر دو ماژول

Fig. 19. System performance using both modules



شکل 20. مقایسه میزان کارایی با استفاده از هریک از ماژول‌ها

Fig. 20. Comparative performance of any of the modules

همان‌طور که در شکل 20 مشاهده می‌کنید، استفاده از ماژول تشخیص و رفع جناس به تنهایی کارایی سیستم را به 34% افزایش می‌دهد، این درحالی است که استفاده از ماژول تبدیل جملات ورودی از گفتار فارسی به همتای خود در دنیای ناشنویان کارایی سیستم را تا 51% افزایش می‌دهد و در نهایت استفاده از ترکیب این دو ماژول باعث افزایش 95% کارایی سیستم مترجم می‌شود.

که 28 جمله از مجموع تست با استفاده از دو ماژول اصلاح شدند، و توسط ناشنویان قابل درک می‌باشند. ستون نارنجی رنگ در آن نشان می‌دهد که با اصلاح آن 28 جمله در مجموع تمام 31 جمله مجموع تست برای ناشنویان قابل درک می‌باشند. به طور کلی اگر سیستم مترجم متن فارسی به زبان اشاره فارسی هم از ماژول تبدیل جمله از گفتار فارسی به همتای خود در دنیای ناشنویان، و همچنین ماژول تشخیص و رفع ابهام و جناس کلمات استفاده شود، را مورد بررسی قرار می‌دهیم. به طور کلی 31 جمله در مجموع تست قرار دارد، که 28 جمله آن توسط ناشنویان قابل درک نمی‌باشد و نیازمند استفاده از ماژول تبدیل و یا ماژول رفع جناس هستند. لازم به ذکر است که 3 جمله نیازمند استفاده از هر دو ماژول رفع جناس و ماژول تبدیل به صورت همزمان خواهد بود. به طور کلی برای ارزیابی کارایی سیستم با استفاده از هر دو ماژول هم‌چنین میزان تاثیر آن در دقت پردازش متن از متدهای دقت و هم‌چنین صحت و در نهایت مدت کارایی استفاده شده است.

$$\text{Precision} = \frac{28}{28} = 1 \quad (9)$$

$$\text{Recall} = \frac{28}{31} = 0.90 \quad (10)$$

$$\text{F-measure} = \frac{2 \times 0.9 \times 1}{0.9 + 1} = 0.95 \quad (11)$$

براساس این محاسبات استفاده از هر دو ماژول تبدیل جملات ورودی به دنیای ناشنویان و ماژول رفع جناس کارایی را به 95% افزایش می‌دهد، این درحالی است که در صورت عدم استفاده از این ماژول همان‌گونه که در رابطه زیر مشاهده می‌کنید، کارایی 18% می‌باشد.

$$\text{F-measure} = \frac{2 \times 0.97 \times 1}{0.97 + 1} = 0.18 \quad (12)$$

بنابراین استفاده از هر دو ماژول تبدیل جملات ورودی به دنیای ناشنویان و هم‌چنین ماژول رفع جناس کلمات کارایی سیستم مترجم متن فارسی به زبان اشاره فارسی را تقریباً 5 برابر می‌کند. در نمودار شکل 19 میزان کارایی سیستم مترجم با استفاده از ماژول تبدیل جملات ورودی به دنیای ناشنویان هم‌چنین بدون استفاده از این ماژول را مشاهده می‌کنید.

بنابراین استفاده از این دو ماژول به صورت همزمان در انتقال مفهوم و هم‌چنین کاهش شکاف بین افراد ناشنوا با دنیای اطراف خود می‌تواند بسیار مفید باشد. در نمودار شکل 20 به مقایسه

4. نتیجه گیری

سیستم مترجم زبان اشاره به منظور برقراری ارتباط افراد ناشنوا با سایر افراد جامعه ایجاد شده است. به علاوه از سیستم مترجم زبان اشاره می توان به منظور آموزش زبان اشاره نیز بهره برد. پیاده سازی سیستم مترجم زبان اشاره فارسی به دلیل غیرساخت یافته بودن و همچنین مشکل چندمعنایی و ابهام برخی کلمات در زبان فارسی، هم چنین مشکلات موجود در زبان اشاره فارسی و محدودیت هایی که در تعداد حرکات شناخته شده این زبان وجود دارد، دارای چالش های می باشد، که در این مقاله با ارائه یک معماری جدید چالش ها تا حدود قابل قبولی برطرف شده است. در معماری جدید از دو ماژول تشخیص و تصحیح جناس و تبدیل جملات به دنیای ناشنوایان استفاده شده است. هر دو ماژول از آنتولوژی طراحی شده، مربوط به خود استفاده می کند. با توجه به نتایج حاصل از ارزیابی سیستم مترجم زبان اشاره با استفاده از معماری جدید 95٪ می باشد.

مراجع

- Federation for Information Processing. Paper Presented at The Ifip Conference On Human-Computer Interaction.
- [9] T.dasgupta and A.basu. (2010). An english to indian sign language machine translation system. dept. of computer science and engineer-ing, indian institute of technology, kharagpur 721302.
- [10] Jestin joy, Kannan balakrishnan. (2014). A prototype malayalam to sign language automatic translator. paper presented at the ifip conference on human-computer interaction.
- [11] Ralf herbich and Thore graepel. (2010). Natural language processing. 676
- [12] Cheok ming jin, Zaid omar. (2016). A mobile application of american sign language translation via image processing algorithms. 978-1-5090-0931-2, ieee, bali, Indonesia
- [13] Verónica lópez-ludeña, rubén san-segundo, Juan manuel montero, Ricardo córdoba, Javier ferreiros, José manuel pardo. (2018). Automatic categorization for improving spanish into spanish sign language machine translation. computer speech & language 26.3 (2018): 149-167.
- [14] Tirthankar dasgupta, s. d., Anupam basu. (2010). Prototype machine translation system from text-to-indian sign. asian federation of natural language processing.
- [15] Vincenzo lombardo, Cristina battaglino, Rossana damiano. (2011). An avatar-based interface for the italian sign language. 978-1-61284-709-2, ieee, seoul, south Korea.
- [16] Ali daud, wahab khan, dunren che. (2016) Urdu language processing: a survey. department of computer science and software engineering iiui slamabad Pakistan
- [17] R. san-segundo, J. m. montero, R. córdoba, V. sama, F. fernández, L. f. d'haro, V. lópez-ludena, D. sañchez, A. garcía. (2012) Design, development and field evaluation of a spanish into sign language translation system. departamento de ingeniería electrónica, etsi telecomunicación universidad politécnica de Madrid Spain
- [18] Tiago maritan ugulino de Araújo, Felipe lacet silva ferreira, Danilo assis nobre dos Santos silva, felipe herminio lemos, gutenberg pessoa neto. (2013) automatic generation of brazilian sign language windows for digital tv systems. digital video applications lab (lavid)federal university of paraíba Brazil
- [19] Matthew j. denny, Arthur spirling. (2016) text preprocessing for unsupervised learning: why it matters, when it misleads, and what to do about it. published by
- [1] W. Stokoe. (1976). Dictionary of American Sign Language On Linguistic Principles, Linstok Press. Linstok Press.
- [2] Sara siavashi. (2005). Persian language and necessity of revision in education and assessment of Iranian deaf. First Conference of Iranian Linguistics Society. Tehran
- [3] Mehrnoush Shamsfard. (2005). Challenges and Open Problems in Persian Text Processing
- [4] Aleahmad. (2007). An introduction to semantic web. University of Tehran.
- [5] Michael C. Daconta, Leo J. Obrst, Kevin Smith. (2006). The semantic web: a guide to the future of xml, web services and knowledge management.
- [6] Amiri and Abdoli. (2009). Design and implementation of course arak university. faculty of engineering
- [7] Lynette van Zijl, dean barker. (2003). South african sign language machine translation system. paper presented at the proceedings of the 2nd international conference on computer graphics, virtual Reality, visualisation and interaction in Africa.
- [8] Sandra baldassarri, Eva cerezo and Francisco Royo-Santas. (2009). Automatic Translation System to Spanish Sign Language with A Virtual Interpreter. International

- [25] Mehrnoush bazrafkan. (2014). A review of persian language processing problems using computer systems. national conference on computer engineering and information technology management. [in Persian]
- [26] Publication of the university of science and technology. (2009). system analysis automatic find keywords farsi. [in Persian]
- [27] Mehrnoush shamsfard. (2006). Persian text processing: past achievements, challenges ahead. The second workshop on farsi and computer. [in Persian]
- [28] Dokhani. (2014) A review of protege software. iranian electronic science and technology magazine. scientific information management organization of iran, volume 1, issue 7 [In Persian]
- [29] Ala akramifard. (2012) ontology representation. international university of ferdowsi [In Persian]
- [30] Vivek kumar verma and Sumit srivastava. (2018). Toward machine translation linguistic issues of indian sign language. part of the advances in intelligent systems and computing book series (AISC, volume 664)
- cambridge university press on behalf of the society for political methodology.
- [20] Kazuma hashimoto, Caiming xiong, Yoshimasa tsuruoka, and Richard socher. (2017) A joint many-task model: growing a neural network for multiple nlp tasks. artificial intelligence (cs.ai)
- [21] Grigoris antoniou and Frank van harmelen. (2008) A semantic web primer.
- [22] Golnaz khaleghi. (2010). Deaf and sign language, limitations and features. [in Persian]
- [23] Ahmad astiri. (2012). Natural language processing tools, the fourth annual workshop of web technology Lab, ferdows university of mashhad. [in Persian]
- [24] Sara siavashi. (2006). Deaf education, defending the reference language in education and eliminating discrimination. Journal of Information And Study Of Disabilities And Disabilities. 1, 35-37 [in Persian]

How to cite this paper:

Mohammad Letafat Abrebekouh, Abdolreza Rrasouli Konari, Mahboobeh Shamsi, , (2019). Introduce detection and correction of ambiguity of words system education to increase the efficiency of translator from persian text to persian sign language using ontology. *Journal of Technology of Education*, 14(1), 59-77.

DOI: 10.22061/jte.2019.3369.1858

URL: http://jte.sru.ac.ir/?_action=showPDF&article=997

