Tech. Edu. J. 13(4): 797-813, Autumn 2019

Technology of Education Journal

(TEJ)

Homepage: jte.sru.ac.ir

ORIGINAL RESEARCH PAPER

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

M Letafat Abrebekouh, A Rrasouli Konari*, M Shamsi

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

ABSTRACT

Receive: 27 September 2018
Reviewed: 16 December 2018
Revise: 23 January 2019
Accept: 03 February 2019

KEYWORDS:
Persian Sign Language
Ontology
Deaf
Natural Language Processing
Correction Ambiguity

* Corresponding author
rasouli@qut.ac.ir

Background and Objectives: The mainstream of society has isolated deaf people due to their inability to communicate with the deaf community. Currently, 1,400 deaf people are born in the country every year, most of whom have a hearing family. For this reason, parents in the early years of a child's life, due to unfamiliarity with sign language, cannot teach them a complete model of sign language. As the name implies, sign language uses gestures, or more precisely hand, eye, and standing postures to convey meaning. If, for example, the movement of the mouth or the general state of the face changes, the meaning of the word produced can also change, or in other words, another word is produced. The same principle applies to the movements of the fingers and hands. 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.

Methods: 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.

Findings: 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.

Conclusion: Sign language interpreter system is designed to connect deaf people with other people in the community. In addition, the sign language translator system can be used to teach sign language. Implementing the Persian sign language translator system has challenges due to its unstructured as well as the problem of ambiguity of some words in Persian, as well as the problems in Persian sign language and the limitations in the number of known movements of this language. The paper addressed the challenges to an acceptable degree by presenting a new architecture. In the new architecture, two modules for recognizing and correcting puns and converting sentences into the world of the deaf have been used. Both modules use their own designed ontology.

NUMBER OF REFERENCES	NUMBER OF FIGURES	NUMBER OF TABLES
30	20	1

مقاله پژوهشی

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

محمد لطافت ابربکوه، عبدالرضا رسولی کناری*، محبوبه شمسی

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

چکیده

پیشینه و اهداف: جریان اصلی جامعه به دلیل عدم توانایی در برقراری ارتباط با جامعه ناشنوایی باعث گوشه گیری این افراد شده است. در حال حاضر سالانه ۱۴۰۰ ناشنوای عمیق در کشور متولد می شود که بیشتر این کودکان دارای خانواده شنوا هستند. به همین دلیل والدین در سال های ابتدایی زندگی کودک، به دلیل عدم آشنایی با زبان اشاره نمی توانند الگوی کاملی از زبان اشاره را به آن ها آموزش دهند. همانطور که از نام «زبان اشاره» پیداست، در زبان اشاره از اشارات و یا به عبارت دقیق تر 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 دست ها نیز صادق است. مهم‌ترین عامل ارتباط بین انسان‌ها، توانمندی در برقراری ارتباط از روش‌های مختلف مانند گفتار، نوشتار و زبان‌اشاره می‌باشد. ناتوانی افراد کم‌شنوا و ناشنوا در برقراری ارتباط صمیمانه، هم‌چنین نبود آموزش مرسوم با توجه به نیازهای جامعه، باعث کنارگذاری این افراد شده است. سیستم های 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 عملا 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.

روش‌ها: 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 هم چنین ماژول تبدیل متن ورودی به دنیای ناشنوایان، با استفاده از آنتولوژی طراحی شده استفاده می‌کند.

یافته‌ها: استفاده از این معماری، نرخ کارایی سیستم مترجم متن فارسی به زبان اشاره ی فارسی را تا ۹۵٪ بهبود بخشیده و باعث افزایش دقت سیستم مترجم تا سه برابر شده است.

نتیجه‌گیری: سیستم مترجم زبان اشاره به منظور برقراری ارتباط افراد ناشنوا با سایر افراد جامعه ایجاد شده است. به علاوه از سیستم مترجم زبان اشاره می‌توان به منظور آموزش زبان اشاره نیز بهره برد. پیاده‌سازی سیستم مترجم زبان اشاره فارسی به دلیل غیرساخت یافته بودن و هم‌چنین مشکل چندمعنایی و ابهام برخی کلمات در زبان فارسی، هم‌چنین مشکلات موجود در زبان اشاره فارسی و محدودیت‌هایی که در تعداد حرکات شناخته شده این زبان وجود دارد، دارای چالش‌های می‌باشد، که در این مقاله با ارائه یک معماری جدید چالش‌ها 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 می‌کند.

دریافت: ۵ مهر ۱۳۹۷
داوری: ۲۵ آذر ۱۳۹۷
اصلاح: ۳ بهمن ۱۳۹۷
پذیرش: ۱۴ بهمن ۱۳۹۷

واژگان کلیدی:
زبان اشاره فارسی
آنتولوژی
ناشنوایان
پردازش زبان طبیعی
تصحیح ابهام

*نویسنده مسئول
rasouli@qut.ac.ir

مقدمه

جریان اصلی جامعه به دلیل عدم توانایی در برقراری ارتباط با جامعه ناشنوایی باعث گوشه گیری این افراد شده است. در حال حاضر سالانه ۱۴۰۰ ناشنوای عمیق در کشور متولد می شود که بیشتر این کودکان دارای خانواده شنوا هستند. به همین دلیل والدین در سال های ابتدایی زندگی کودک، به دلیل عدم آشنایی با زبان اشاره نمی توانند الگوی کاملی از زبان اشاره را به آن ها آموزش دهند. همانطور که از نام «زبان اشاره» پیداست، در زبان اشاره از اشارات و یا به عبارت دقیق تر 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 دست ها نیز صادق است[۲۲]. 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، دسته بندی و بازیابی و پردازش آن توجه بیش از پیش به پردازش زبان طبیعی توسط رایانه افزایش یافته است[۲۳]

زبان اشاره

زبان اشاره نوعی از زبان است که در آن ناشنوایان به جای استفاده از سیستم صوتی و یا به عبارتی گفتار، از دست ها برای سخن گفتن بهره می‌بریم. در واقع این دست‌ها هستند که به جای زبان، سخن می‌گویند[۱]. زبان اشاره به ندرت توسط شنوا قابل درک می‌باشد، و این یک مانع بزرگ ارتباطی می‌باشد[۱۲].

امروزه مکتب های آموزشی مختلفی برای ناشنوایان و کم‌شنوایان وجود دارد. گروهی بر این باورند که تضعیف زبان اشاره و تاکید بر روش های 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[۲۴].زبان اشاره، یک زبان جهانی نیست، و از هر کشوری به کشور دیگر متفاوت است[۱۳]. زبان اشاره فارسی از گویشی به گویش دیگر متفاوت است. زبان اشاره فارسی علی رغم افزایش تعداد ناشنوایان و استفاده از زبان اشاره، این زبان پیشرفت آن چنانی 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[۳۰]. 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[۲].

پردازش زبان طبیعی

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

از کاربردهای نوشتاری می توان به موارد زیر اشاره کرد:

استخراج اطلاعات خاص از یک متن

ترجمه یک متن به زبانی دیگر

خلاصه سازی متون

یافتن مستنداتی خاص در یک پایگاه داده نوشتاری

از کاربردهای گفتاری می‌توان به موارد ذیل اشاره کرد:

سیستم های پرسش و پاسخ انسان و رایانه

سرویس های خودکار ارتباط با مشتری از طریق تلفن

سیستم های آموزشی به دانش آموزان

سیستم های کنترلی توسط صدا

از دیدگاه دیگر می توان پردازش زبان طبیعی را به سطوح زیر تقسیم بندی کرد[۲۵]:

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 آن ها 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.

جایگاه زبان فارسی

جایگاه زبان فارسی در میان زبان های دیگر را از سه جنبه ی وراثتی) تاریخی (، ناحیه ای و رده شناختی می‌توان بررسی کرد[۳]: 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 می باشد. 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.[۲۷] 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 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[۲۶].

مسائل و مشکلات پردازش زبان فارسی

برخلاف زبان انگلیسی که در آن هم حروف و هم لغات کاملا 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 پیچیده تر 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 زبان های 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 لزوما امکان‌پذیر نبوده و مستلزم آزمایش‌های فراوان برای اثبات صحت آن خواهد بود. [۲۵]

چالش های 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[۲۶].

نبود نکات گرامری تعریف شده همانند آن چه که در زبان انگلیسی وجود دارد . این مساله در ریشه‌یابی و پیش‌پردازش متن تاثیرگذار 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.

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 آن جا 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[۲۵].

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[۲۵]. برای مقابله با این گونه 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.

آنتولوژی

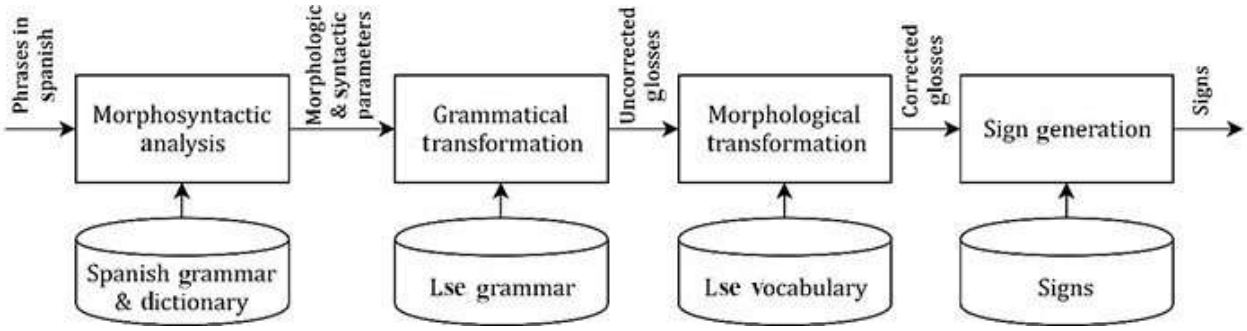
آنتولوژی ریشه در فلسفه دارد و مبدا آن را ارسطو می‌دانند. در فلسفه، آنتولوژی شاخه‌ای از علم است که به بررسی موجودات و روابط میان آن ها می‌پردازد. مفهوم آنتولوژی در وب معنایی کمی متفاوت از فلسفه است. آنتولوژی از دو واژه Onto به معنی هستی و Logia به معنی مطالعه به وجود آمده است و در کل معنی هستی‌شناسی دارد[۴]. آنتولوژی در وب معنایی واژه‌ها و ارتباط بین آن ها در دامنه‌ای که استفاده می‌شود را نشان می‌دهد. عناصر اصلی تشکیل دهنده‌ی آنتولوژی عبارتند از: مفاهیم، ارتباط بین آن ها و خصوصیات آن ها.

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 مفهوم سازی باعث بحث‌های زیادی شده‌اند، نکته‌ی اساسی در مورد این تعریف از آنتولوژی موارد زیر هستند[۵].

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 به‌کار 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 تامین 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



شکل ۱: معماری تبدیل زبان اسپانیایی به زبان اشاره اسپانیایی

Fig. 1: Architecture translation system spanish to spanish sign language

مشخص می‌کند که با این کار اسناد مربوطه توسط ماشین قابل پردازش می‌شود و اشتراک گذاری اطلاعات بین عامل‌ها را تسهیل می‌کند.در واقع می توان گفت[۶]:

Vocabulary + Structure = Taxonomy

Taxonomy + Relationship, Constraints and Rules = Ontology

Ontology + Instance = Knowledgebase

عناصر مختلف آنتولوژی شامل موارد زیر است[۲۸]:

نمونه‌ها: اشیای ابتدای یا نمونه های اولیه که مبنای ساخت آنتولوژی می باشد.

کلاس ها: مجموعه ها، مفاهیم، انواع اشیا یا انواع چیزهای ابتدای تشکیل دهنده نمونه.

خاصیت ها: جنبه ها، ویژگی ها، خصوصیات یا پارامترهایی که آن اشیا(کلاس مربوط به اشیا) می توانند داشته باشند.

ارتباطات: روش هایی که در آن کلاس ها و نمونه هایی می توانند با یکدیگر ارتباط داشته باشند.

جملات تابعی: ساختارهای پیچیده ای که از یک ارتباط مشخص شکل می گیرند، که می توانند به جای یک کلمه یا کلمه ی خاص یک مورد استفاده قرار گیرند.

قیدها: توضیحاتی که به صورت رسمی بیان می شوند تا مشخص کنند که چه چیزی باید صحیح باشد تا این که یک حکم به عنوان ورودی مورد پذیرش قرار گیرد.

پیشینه پژوهش

در سال ۲۰۰۳، 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. 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[۷].

در سال ۲۰۰۹، 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 ۱ آورده شده است.

چهار ماژول اصلی در این معماری عبارتند از:

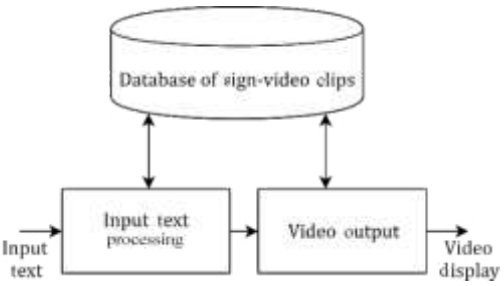
ماژول آنالیز مورفولوژیکی: این ماژول وظیفه استخراج روابط و وابستگی‌های نحوی را دارد.

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 ناشنوایان را دارا می‌باشد.

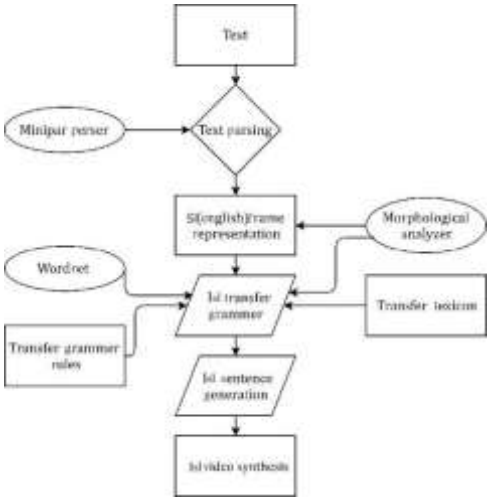
ماژول تولید اشاره: وظیفه نمایش جمله ساخته شده در مراحل قبل را دارد.

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 به‌صورت آزمایشی به سیستم داده شده بود بررسی شد. در مجموع ۹۲ جمله که شامل ۵۶۱ کلمه بود توسط سیستم مترجم مورد آزمایش قرار گرفت[۸].



شکل ۲: معماری تبدیل متن به زبان اشاره بنگلادشی

Fig. 2: Text to basl translation's architecture



شکل ۳: معماری سیستم مترجم Ingit

Fig. 3: Ingit translation system's architecture

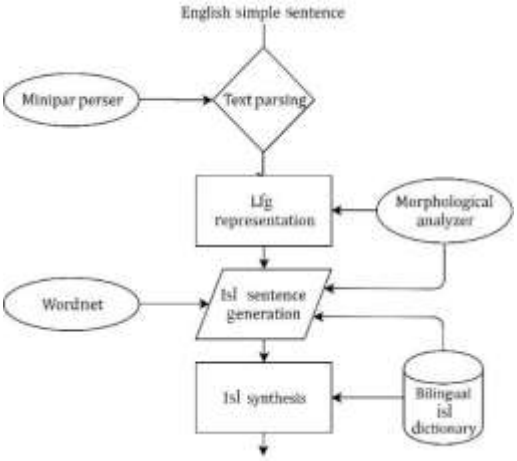
در سال ۲۰۰۹، عده‌ای از دانشجویان بنگلادشی یک سیستم مترجم برای تبدیل متن بنگلادشی به زبان اشاره بنگال ایجاد کردند. آن‌ها برای کاهش موانع ارتباطی بین افرادی که دارای اختلال شنوایی هستند با افراد عادی جامعه، نرم‌افزار مترجمی را توسعه دادند که متن ورودی را به زبان بنگال دریافت کرده و به طور خودکار به زبان 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 می شود که این کلمات، عمدتا ساده هستند و از کتاب های درسی کودکان جمع آوری شده اند. روش کار این مترجم اینگونه است که کلمه یا جمله ساده در بخش مربوطه وارد شده و پردازش‌ها و جابجایی‌های 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 ۲ آورده شده است.معماری این مترجم شامل سه ماژول اصلی است که عبارتند از:

ماژول پردازش متن ورودی

ماژول پایگاه‌داده‌هایی از کلیپ‌های تصویری

ماژول خروجی ویدئویی

این سیستم برای نمایش خروجی برنامه که زبان اشاره بنگال می‌باشد از ویدئوهای از قبل تهیه شده استفاده می‌کند. این مترجم به دنبال



شکل ۴: معماری سیستم تبدیل متن هندی به زبان اشاره هندی

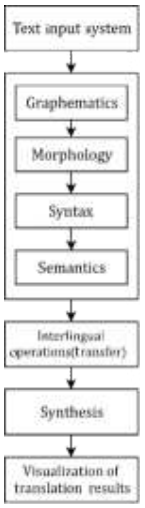
Fig. 4: Text to isl translation system's architecture

استانداردسازی Basl در تمام مناطق بنگلادش می‌باشد

در سال ۲۰۱۰، 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 ۳ نمایش داده شده است[۹].

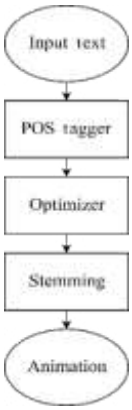
به منظور ارزیابی این سیستم از کارشناسان زبان‌شناسی برای درک مفهوم خروجی استفاده شده است. نتایج اعلام شده از سوی این کارشناسان در مجموع قابل قبول بوده است.

در سال ۲۰۱۰، 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 ناشنوای هند 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 ۴ آن را مشاهده می‌کنید.



شکل ۵: معماری سیستم تبدیل زبان روسی به زبان اشاره روسی

Fig. 5: Russian text to rsa translation system's architecture



شکل ۶: معماری تبدیل زبان مالایی به زبان اشاره مالایی

Fig. 6: Malayalam language to sign language architecture

در سال ۲۰۱۳، 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 ۵ 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

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 ۲۰۱۴، در آزمایشگاه دانشگاه Cochin 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 ریشه یاب می‌باشد که وظیفه یافتن ریشه کلمه را دارد. در نهایت ماژول انیمیشن خروجی را نمایش می‌دهد. معماری این سیستم در شکل ۶ آورده شده است.

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[۱۰].

به طور کلی در جدول ۱ خلاصه ای از تمام سیستم های ارائه شده، همراه با مزایا و معایب هریک بیان شده است.

روش پژوهش

برای ایجاد یک سیستم مترجم از زبانی به زبان دیگر دو روش وجود دارد[۱۴]:

روش تبدیل مستقیم زبان مبدا به زبان مقصد: این روش نیازمند اطلاعات کامل از هر دوزبان مبدا و مقصد می‌باشد، به علاوه این روش زمانی استفاده می‌شود که کلمات در هر دو زبان مبدا و مقصد نظیر به نظیر وجود داشته باشد.

روش تبدیل مبتنی بر معماری: در این روش از روش مستقیم نیز برای برخی کلمات استفاده می‌شود اما به طور کلی سیستم تبدیل دارای چارچوبی است که تمام کلمات و جملات زبان مبدا پس از طی کردن آن چارچوب به زبان مقصد تبدیل می‌شود.

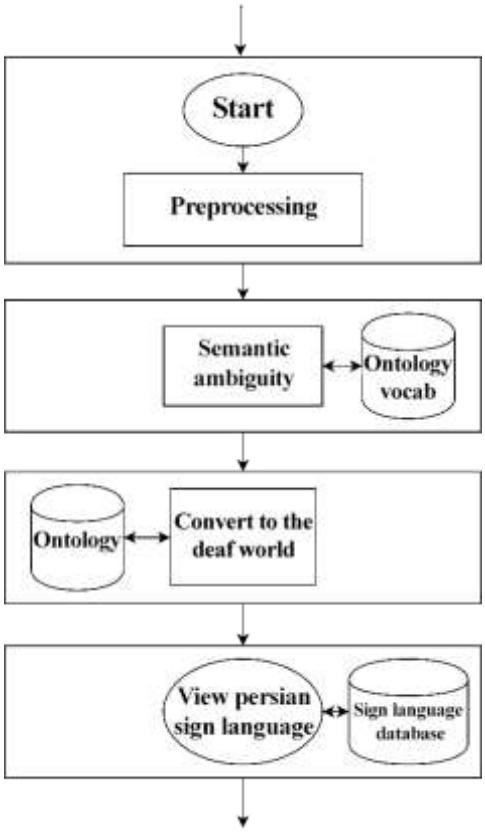
برای آن دسته از سیستم‌های مترجمی که دامنه و حوزه مشخصی دارد و دارای کاربرد عمومی نیستند، و زبان مبدا و مقصد به خوبی شناسایی شده است، به عبارت دیگر هردو زبان مبدا و مقصد استاندارد هستند روش تبدیل مستقیم مناسب می‌باشد. اما برای کاربرد عمومی، روش مبتنی بر معماری مناسب است. برای افزایش کارایی سیستم مترجم متن فارسی به زبان اشاره فارسی به دلیل چالش‌های ذاتی زبان فارسی،

جدول ۱: خلاصه از سیستمهای پیشین

Table 1: A summary of past work

Title	Years	Method	Advantages	Disadvantages
South African sign language machine translation system	2003	• Pos tagging • Syntactic tree	• Use animation marking	• Failure to identify the relationship between words • Lack of identification semantic ambiguity
Automatic translation system to Spanish	2009	• Feeling analysis	• Use animation • High accuracy	• Lack of identification semantic ambiguity
An English to Indian machine translation system	2010	• Convert text to sign language	• Proper architecture • Detect the relationship between words	• Memory overhead • Lack of identification semantic ambiguity
A prototype Malayalam to sign language automatic	2014	• Pos tagging • Animation	• Pos tagging • Use animation • Word stemming	• Lack of identification semantic ambiguity

از قبیل غیرساخت‌یافته بودن و هم‌چنین مواجع با چندمعنای و وجود ابهام در لغات فارسی، از روش مبتنی بر معماری استفاده می‌کنیم. معماری پیشنهادی برگرفته از معماری‌های استفاده شده در سایر زبان‌ها می‌باشد با این تفاوت که به معماری‌های پیشین باید بخش‌های مربوط به آنتولوژی اضافه شود تا قابلیت رفع چالش‌های موجود در زبان فارسی را داشته باشد. مراحل کار معماری موردنظر در شکل ۷ ارائه شده است. این معماری از چهار گام اصلی برای تبدیل جمله ورودی در زبان فارسی



شکل ۷: معماری پیشنهادی

Fig. 7: Proposed architecture

به زبان اشاره فارسی استفاده می‌کند؛

ماژول پیش‌پردازش

ماژول بررسی ابهام و جناس

ماژول نمایش اشارات

ماژول پیش‌پردازش

ماژول پیش‌پردازش که اولین ماژول در معماری پیشنهادی سیستم ترجمه متن فارسی به زبان اشاره فارسی است خود دارای ماژول‌های زیر می‌باشد، که عبارتند از:

نرمال سازی

تجزیه متن به اجزا

تجزیه متن به جملات

استانداردسازی

ریشه یابی

برچسب گذاری

تجزیه متن

هریک از این ماژول‌ها که در ادامه در مورد آن‌ها صحبت خواهیم کرد به صورت گام به گام انجام می‌شود، یعنی به طور مثال زمانی که ماژول تجزیه متن در حال انجام است ماژول قبلی آن، که نرمالسازی است قطعا انجام شده است و نتیجه این ماژول در حال تجزیه شدن می‌باشد.

نرمال سازی

در اولین گام باید متون برای استفاده در گام های بعدی به شکلی استاندارد درآیند[۱۶]. از آن جا که متون مختلف ممکن است بسیار به هم شبیه باشند اما به دلیل تفاوت‌های ساده ظاهری از نظر ماشین متفاوت بیایند؛ به همین دلیل سعی شده است این تفاوت‌های ساده ی ظاهری برطرف گردد. برای رسیدن به این هدف، قبل از مقایسه متون، پیش‌پردازش‌هایی روی آن ها انجام می‌شود. طبیعتا هر چه این پیش‌پردازش‌ها قوی‌تر باشد، نتایج حاصل از مقایسه متون قابل اطمینان تر خواهد بود[۱۱]. لازم به ذکر است که از آن جایی که زبان

فارسی جزو زبان‌های غیرساخت یافته است با مشکلات بسیار بیشتری نسبت به سایر زبان‌ها مواجه خواهیم شد. متون غیرساخت‌یافته، متونی هستند که پیش فرض خاصی در مورد قالب آن ها نداریم و آن ها را به صورت مجموعه‌ای مرتب از جملات در نظر می‌گیریم.

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[۲۳]:

بررسی همزه و انواع مختلف املاهای موجود و اصلاح هر کدام (به عنوان مثال تبدیل ؤ به و ، ئ به ی ، أ به ا ، إ به ا و ...)

حذف تشدید از واژه‌ها

تبدیل ارقام عربی و انگلیسی به معادل فارسی

اصلاح اعراب و حذف فتحه، کسره و ضمه و همچنین تنوین‌ها

حذف نیم‌فاصله‌های تکراری

حذف نویسه‌ی «ـ» که برای کشش نویسه های چسبان مورد استفاده قرار می گیرد. مانند تبدیل«بـــر» به «بر»

چسباندن پسوندهای «تر»، «ترین» و ... به آخر واژه‌ها

تبدیل «ۀ» به «ه‌ی»

حذف فاصله‌ها و نیم‌فاصله‌های اضافه به کار رفته در متن

تجزیه متن به اجزا

Token کردن ابزاری برای شکستن متن ورودی براساس واحدهای با معنی مانند کلمه است [۱۱]. نمادهای معناداری مانند Space، Tab و... لازمه ایجاد این ابزار در زبان فارسی می‌باشد؛ این کار به منظور شناسایی واحدهای مستقل معنایی می‌باشد. سپس براساس انتخاب هرکدام از این واحدها متن بر اساس آن‌ها شکسته خواهد شد[۱۸].

.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.

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.

ریشه یابی

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 تک‌واژها نگاه می‌کنیم و به کلمات اطراف آن لغت کاری نداریم در فارسی به این کار ریشه‌یابی لغوی می‌گویند[۱۹].

برچسب گذاری

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

تجزیه متن

این ماژول از دستورات گرامری زبان بهره می‌برد. ماژول پارسر باعث تشخیص گروه‌های تشکیل دهنده جملات ورودی می‌باشد. تجزیه و تحلیل جمله ورودی و شکستن آن به اجزای تشکیل دهنده‌ی 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 تشکیل دهنده متن ورودی می باشد. ابزار پیش پردازش متن در سیستم مترجم، چندین جمله در قالب یک متن را دریافت می کند و خروجی آن نرمال شده کلمات تشکیل دهنده آن جملات می باشد[۲۰].

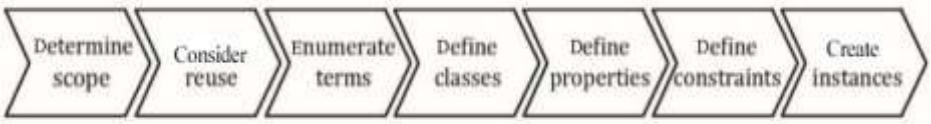
ماژول تشخیص و تصحیح جناس کلمات

برای برطرف کردن ابهام و جناس بین کلمات از آنتولوژی استفاده می‌شود. به طور کلی دو روش خودکار و دستی برای ایجاد آنتولوژی استفاده می شود. روش های خودکار از روش های متفاوتی برای استخراج آنتولوژی از متون موجود استفاده می کنند و بیش تر این روش ها از تکنیک های پردازش زبان طبیعی استفاده می کنند .اکنون روش خودکاری که آنتولوژی قابل قبولی تولید کند، معرفی نشده است. بنابراین برای ایجاد آنتولوژی از روش های دستی و یا نیمه خودکار استفاده می شودمستقل از روش ایجاد آنتولوژی می توان از متدولوژی خاصی برای ایجاد آنتولوژی استفاده کرد. متدولوژی استانداردی برای ایجاد آنتولوژی وجود ندارد و بسته به کاربرد آنتولوژی از متدولوژی های متفاوتی

استفاده می شود. در این بخش یکی از این متدولوژی ها که روند آن در شکل ۸ نشان داده شده است، مطرح می شود. در روند ایجاد آنتولوژی به صورت بازگشتی عمل می شود، یعنی در هر مرحله ممکن است نیاز به ایجاد اصلاحات و بازگشت به مرحله ی قبلی باشد.

تشخیص محدوده

مراحل مختلف توسعه ی یک آنتولوژی را با تعریف دامنه و حوزه ی آن



شکل ۸ : متدلوژی ایجاد آنتولوژی
Fig.8 : Methodology for ontology creation

آغاز می کنیم. این کار شامل پاسخ گویی به سوالات پایه زیر است[۲۹]:
آنتولوژی در حال توسعه، چه دامنه ای را پوشش می دهد؟
برای چه منظوری از این آنتولوژی استفاده می کنیم؟
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 حوزه ی مدل کمک خواهند کرد. به طور کلی الگوریتم مراحل تعیین و تثبیت دامنه مورد نظر برای آنتولوژی در شکل ۹ 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.

استفاده مجدد

در این مرحله مطالعه ی جامعی بر روی آنتولوژی موجود انجام و مشخص می شود که از کدام یک از آنتولوژی های موجود می توان برای ایجاد آنتولوژی مورد نظر کمک گرفت. استفاده ی مجدد این مزیت را دارد که صحت آنتولوژی های قبلی در عمل بررسی شده و از ابزارهایی که بر اساس آنها کار می کنند نیز می توان استفاده کرد[۲۱].

تشخیص واژه ها

تهیه لیستی از تمامی واژه هایی که تمایل داریم برای آن ها جمله بسازیم یا در یک کاربرد درباره ی آنها توضیح دهیم نیز کاری مفید محسوب می گردد:
تمایل داریم در مورد چه واژگانی صحبت کنیم؟
این واژه ها چه ویژگی هایی دارند؟
درباره این واژه ها چه می خواهیم بگوییم؟
دو گام بعدی شامل توسعه سلسله مراتب کلاس ها و تعریف ویژگی های مفاهیم بسیار نزدیک به هم و در هم تنیده هستند. به گونه ای که تمایز

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

شکل۹ : الگوریتم تعیین دامنه آنتولوژی
Fig. 9: Algorithm determine the domain

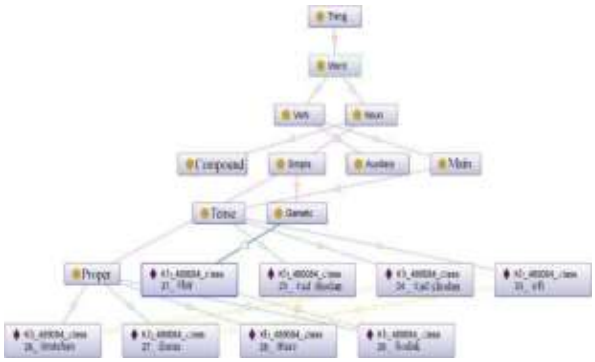
میان آن ها و قائل به تقدم و تاخر شدن میان آن ها بسیار سخت است. به‌طور معمول با ایجاد تعریف برای
تعدادی محدود از مفاهیم در سلسله مراتب آغاز می کنیم و سپس با توصیف ویژگی های آن ها (مفاهیم ایجاد شده) کار را ادامه می دهیم. این دوگام، مهم ترین گام ها در فرآیند طراحی آنتولوژی هستند.

تشخیص کلاس ها و ساختار آن ها

با استفاده از واژه های به دست آمده کلاس ها و زیرکلاس های آن ها را مشخص می کنیم. برای این کار می توان 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 تعاریف مفاهیم عام در دامنه آغاز می کند. سپس با ایجاد زیرکلاس‌های خاص تر از این مفاهیم، فرآیند توسعه را ادامه می دهد.
رهیافت پایین به بالا: فرآیند توسعه پایین به بالا با تعریف کلاس های بسیار خاص آغاز می کند. برگ های سلسله مراتب کلاس ها سپس با گروه بندی این کلاس ها در قالب مفاهیم عام تر، فرآیند توسعه را ادامه می دهد.
رهیافت ترکیبی: فرآیند توسعه ترکیبی، ترکیبی از روش های بالا به پایین و پایین به بالا محسوب می گردد. در این روش، ابتدا مفاهیم برجسته و حائز اهمیت تعریف می شوند. سپس با عمومی سازی (ترکیب) و یا خصوصی سازی (تجزیه) این مفاهیم، فرآیند توسعه ادامه می یابد.
هیچ یک از این سه روش، به صورت ذاتی از مابقی بهتر نیست. رهیافتی که جهت توسعه آنتولوژی اتخاذ می گردد، کاملا به دیدگاه شخصی طراح، از دامنه وابسته است. اگر یک توسعه 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 می رسد که رهیافت ترکیبی برای بسیاری از توسعه دهندگان آنتولوژی ها راه حل ساده تری محسوب می گردد و تمایل بیشتری به استفاده از این روش وجود دارد. جدا از این که کدام یک از این رهیافت ها را مورد استفاده قرار می دهیم، معمولا با تعریف کلاس ها آغاز می کنیم. از لیستی که در گام سوم ایجاد کرده ایم واژه هایی را انتخاب می کنیم که اشیاء مستقل را توصیف می کنند. این واژه ها «کلاس» را مشخص کرده و نقاط لنگرگاه در سلسله مراتب کلاس ها را تشکیل می دهند. به عبارت دیگر کلاس ها مجموعه ای از نمونه ها است که به توصیف دقیق موجودیت ها اشاره دارد. به طور کلی الگوریتم مراحل ساخت و ایجاد آنتولوژی در شکل ۱۰ آورده شده است؛
یکی از مهم ترین مراحل ایجاد یک آنتولوژی شناخت دامنه و حوزه کاری آن می باشد. با شناخت دامنه ایجاد آنتولوژی می توان واژه های محدود

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 ۱۰: الگوریتم ایجاد آنتولوژی
Fig. 10: Ontology construction algorithm



شکل ۱۱: بخشی از آنتولوژی طراحی شده
Fig. 11: Part of the ontology is designed

به آن حوزه را شناسایی کرد. پس از شناسایی واژه ها، با استفاده از یکی از رهیافت های ذکر شده، وخواص شناسایی شده برای آن ها، کلاس بندی انجام می شود. در گام بعدی اعضای هریک از کلاس براساس میزان ارجاع به آن واژه شماره گذاری می شود.

در شکل ۱۱ آنتولوژی طراحی شده در مقیاس کوچک را مشاهده می کنید.

این آنتولوژی با استفاده از ابزار طراحی پرونج ایجاد شده است. در این طراحی از رهیافت ترکیبی استفاده شده است. کلاس ها و زیرکلاس ها را مشاهده می کنید که در ساختار سلسله مراتبی قرار گرفته اند. در سطح آخر و برگ این ساختار نمونه ها قرار گرفته اند که در این جا برای مثال در سطح آخر کلمه «شیر» و کلمه کودک را مشاهده می کنید که با یکدیگر ارتباط داشته و به دلیل این که در ویژگی های کلمه شیر آورده شده است که این کلمه دارای ابهام و جناس می باشد، به منظور رفع جناس ان از کلمه هم معنی یا به اصطلاح معنی دوم این کلمه استفاده می شود. حال آنتولوژی در دامنه مورد نظر ایجاد شده است و آماده استفاده می باشد. برای تشخیص و تصحیح جناس با استفاده از آنتولوژی و جایگزنی کلمات جناس دار با معادل ساده و بدون جناس خود مراحل الگوریتم شکل ۱۲ 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.

تبدیل به دنیای ناشنوایان

زبان اشاره فارسی دارای الفبای مخصوص به خود ناشنوایان می‌باشد و

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 ۱۲: الگوریتم شناسایی و تصحیح ابهام و جناس
Fig. 12: Semantic ambiguity recognition & correction algorithm

اشارات تمام حروف الفبا برای آنان موجود است، به علاوه در زبان اشاره حرکات درنظر گرفته شده برای کلمات محدود است و تمام کلمات اشاره مستقیم ندارد. زبان اشاره کامل و بدون نقصی نیست، همچنین از گویشی به گویش 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 این‌که 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 به روز 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 می باشند که اشاره نظیر داشته و مستقیما برای ناشنوایان قابل درک هستند. برای تبدیل کلمات به همتای خود در دنیای ناشنوایان از تکنیک استفاده از آنتولوژی استفاده می شود. آنتولوژی مربوط به ماژول تبدیل به مانند آنتولوژی مربوط به ماژول تشخیص و تصحیح جناس می باشد. برای استفاده از ماژول تبدیل متن

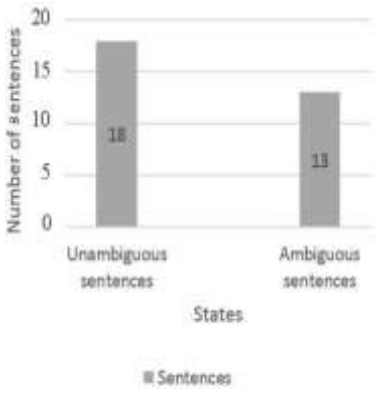
1. Algorithm translator
2. Input(text)
3. Output(text in deaf word)
4. Preprocessing input text
5. Tokenize text
6. Compare word in ontology
7. If (word in ontology)
    a. Replace word for word in deaf word
    b. View sign language
8. Else
    a. Parsing word for alphabet
    b. View alphabet sign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

شکل ۱۳: الگوریتم تبدیل گفتار فارسی به دنیای ناشنوایان
Fig. 13: Translator module algorithm

از گفتار فارسی به همتای خود در دنیای ناشنوایان، در ابتدا متن ورودی پیش‌پردازش می‌شود. خروجی این کار تجزیه متن ورودی به کلمات تشکیل‌دهنده 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 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 این‌بار به حروف تشکیل‌دهنده خود تجزیه می‌شوند. و در نهایت به‌صورت هجای حروف نمایش داده می‌شوند.
مراحل بیان شده برای انجام تبدیل متن ورودی به همتای خود در دنیای ناشنوایان را در الگوریتم شکل ۱۳ نیز مشاهده می‌کنید.

نمایش اشارات

برای نمایش زبان اشاره دو روش وجود دارد[۱۵]، یکی استفاده از فیلم‌های از قبل تهیه شده و دیگری استفاده از انیمیشن می‌باشد. هردو روش مشکلات و محاسن مربوط به خود را دارد. استفاده از فیلم‌های از قبل تهیه شده دارای سادگی در پیاده سازی 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 فضای ذخیره‌سازی دارای

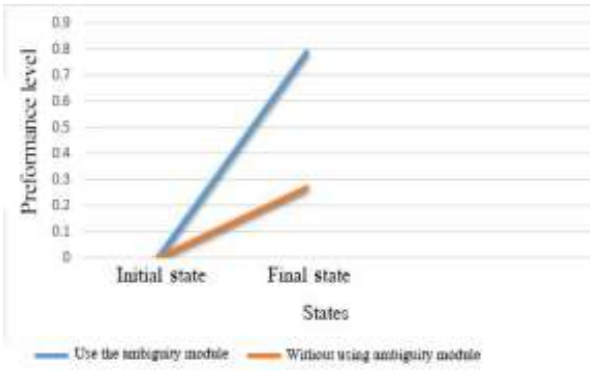


شکل ۱۴: جملات مجموعه تست برای ماژول تصحیح ابهام
Fig.14 : Test set sentences for ambiguity correction module

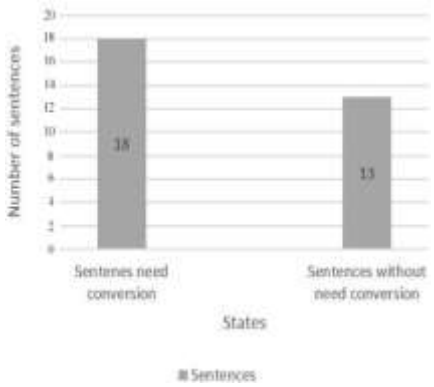
سرباز حافظه است. البته برای استفاده در یک دامنه محدود و کوچک استفاده از فیلم بسیار مناسب‌تر و بهتر است. استفاده از انیمیشن برای نمایش زبان اشاره دارای پویایی بیشتر می‌باشد و دچار سربار حافظه نیز نمی‌باشد، اما پیاده‌سازی آن پیچیده و دشوار می‌باشد[۱۷]. همچنین به دلیل نمایش آن توسط کاراکترهای انیمیشنی دقت پایین‌تری نسبت به حالت قبل دارد. این نوع نمایش برای استفاده در دامنه بزرگ مناسب‌تر است.

نتایج و بحث

برای ارزیابی سیستم مترجم متن فارسی به زبان اشاره فارسی از ۳۱ 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 تائید مربیان و مترجمین زبان اشاره نیز رسیده است. همچنین خروجی سیستم مترجم خودکار متن فارسی به زبان اشاره فارسی و خروجی هریک از ماژول ها، توسط مترجمین زبان اشاره فارسی بررسی و کارایی و دقت انتقال مفاهیم و کلمات به دنیای ناشنوایان ارزیابی شده است که در ادامه بیان می شود.
به منظور ارزیابی ماژول تشخیص و تصحیح ابهام کلمات، ۱۳ جمله در مجموع تست وجود دارد، همچنین ۱۸ جمله برای درک توسط ناشنوایان نیازمند استفاده از ماژول تبدیل می‌باشند.
در نمودار شکل ۱۴ نحوه قرارگیری جملات دارای جناس را در جملات تشکیل‌دهنده مجموعه تست سیستم مترجم مشاهده می‌کنید.
همان‌گونه که مشاهده می‌کنید، ۴۲٪ جملاتی که برای تست سیستم مترجم درنظرگرفته شده‌اند دارای جناس می‌باشد. این جملات برای درک توسط سیستم مترجم و بالا بردن کارایی سیستم و گاهش 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



شکل ۱۵: میزان کارایی سیستم با استفاده از ماژول تصحیح ابهام
Fig. 15: System performance using ambiguity correction module



شکل ۱۶: جملات مجموعه تست برای ماژول تبدیل
Fig. 16: Sentence test Set for convert module

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 می کند، با این فرض تعداد جملات مجموع تست به ۱۶ جمله کاهش می یابد. در این مجموع ۱۳ جمله دارای جناس است که برای ترجمه به زبان اشاره فارسی و درک آن توسط ناشنوایان باید ابهام و جناس آن برطرف گردد. البته ۳ جمله برای انتقال مفهوم، علاوه بر ماژول تشخیص و رفع ابهام و جناس نیازمند ماژول دیگری نیز می باشد. بنابراین تنها ۱۰ جمله تنها با رع ابهام و جناس آن توسط ناشنوایان قابل درک می باشد. هم چنین ۳ جمله بدون جناس می باشد، که این ۳ جمله بدون استفاده از هیچ ماژول و پردازشی توسط ناشنوایان قابل درک می باشد. به طور کلی برای ارزیابی کارایی ماژول تشخیص و تصحیح ابهام و جناس در جملات، هم چنین میزان تاثیر آن در دقت پردازش متن از معیار دقت و صحت که برای ارزیابی سیستم های بازیابی اطلاعات است، هم‌چنین برای ارزیابی سیستم‌های که بصورت کیفی می‌باشند و برای ارزیابی نیازمند نتایج کمی هستیم از این نوع ارزیابی استفاده می کنیم. در نهایت به منظور ارتباط این دو معیار از میانگین هارمونیک آن دو که معیار F نامیده می شود بهره می بریم که کارایی سیستم مترجم خودکار متن فارسی به زبان اشاره فارسی را بررسی نمائیم.

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

$$\text{Precision} = \frac{10}{17} = 0.77 \qquad (۱)$$

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

$$\text{Recall} = \frac{13}{16} = 0.81 \qquad (۲)$$

با بهره گیری از این دو معیار به ارزیابی کارایی استفاده از ماژول تشخیص و رفع جناس در جملات موجود در مجموعه تست با استفاده از متدF می‌پردازیم:

$$\text{F-measure} = \frac{2*0.77*0.81}{0.77+0.81} = 0.79 \qquad (۳)$$

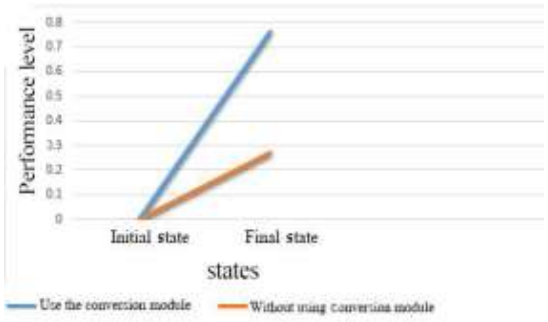
براساس این محاسبات استفاده از ماژول تشخیص و تصحیح ابهام و جناس در متن کارایی سیستم مترجم متن فارسی به زبان اشاره فارسی به صورت خودکار را به ۷۹٪ افزایش می دهد، یعنی در صورت استفاده از این ماژول دقت سیستم مترجم برای انتقال مفاهیم به دنیای ناشنوایان و درک مفاهیم توسط مترجمین زبان اشاره به صورت خودکار تا ۷۹٪ افزایش می یابد. این درحالی است که در صورت عدم استفاده از این ماژول همان گونه که در رابطه زیر مشاهده می کنید، کارایی ۲۷٪ می باشد.

$$\text{F-measure} = \frac{2*0.1875*0.5}{0.1875+0.5} = 0.27 \qquad (۴)$$

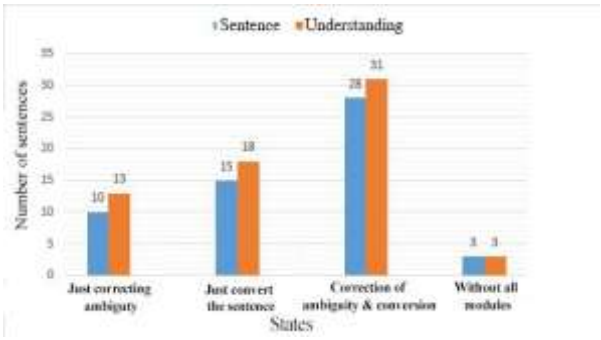
بنابراین استفاده از ماژول جناس به تنهایی کارایی سیستم مترجم متن فارسی به زبان اشاره فارسی به صورت خودکار را تقریبا ۳ برابر افزایش می دهد. در نمودار شکل ۱۵ میزان کارایی سیستم مترجم به منظور انتقال خودکار مفاهیم به ناشنوایان و مترجمین زبان اشاره با استفاده از ماژول تشخیص و رفع ابهام و جناس هم‌چنین بدون استفاده از این ماژول را مشاهده می کنید.

به‌علاوه در جملات در نظر گرفته شده ۱۸ جمله وجود دارد که قابل درک برای ناشنوایان نیستند و نیازمند تبدیل برخی کلمات تشکیل‌دهنده آن جملات به معادل آن‌ها در دنیای ناشنوایان داریم. این تبدیل کلمات به این دلیل انجام می‌شود که تمام کلمات در دنیای ناشنوایان، معادل نداشته، قابل نمایش توسط زبان اشاره نمی‌باشد. در شکل ۱۶ نحوه قرارگیری جملات نیازمند تبدیل را در مجموعه تست برای ارزیابی سیستم مترجم را مشاهده می‌کنید؛

همان‌طور که مشاهده می‌کنید ۵۸٪ جملاتی که برای ارزیابی سیستم در نظر گرفته شده است، نیازمند تبدیل کلمات تشکیل‌دهنده آن‌ها به



شکل ۱۷: میزان کارایی سیستم با استفاده از ماژول تبدیل
Fig .17 : System performance using convert module



شکل ۱۸: تاثیر ماژول‌ها
Fig. 18 : Modules' impact

معادل خود در دنیای ناشنوایان می‌باشد. برای مثال در جمله «پنج موقع را برای دعا و حاجت خواستن غنیمت شمارید» چندین کلمه این جمله در دنیای ناشنوایان ناشناخته می‌باشد و نیازمند تبدیل آن کلمات به معادل خود در دنیای ناشنوایان 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 تشکیل‌دهنده این جمله در دنیای ناشنوایان شناخته شده باشد و توسط زبان اشاره قابل نمایش گردد. اگر این که در سیستم مترجم متن فارسی به زبان اشاره فارسی تنها از ماژول تبدیل جمله گفتار فارسی به همتای خود در دنیای ناشنوایان استفاده کند، را مورد بررسی قرار می دهیم. به طور کلی ۲۱ جمله در مجموع تست قرار دارد، که ۱۸ جمله آن توسط ناشنوایان قابل درک نمی باشد و نیازمند استفاده از ماژول تبدیل هستند. لازم به ذکر است که ۱۵ جمله تنها با استفاده از ماژول تبدیل جمله توسط ناشنوایان قابل درک خواهد بود و ۳ جمله نیازمند ماژولی علاوه بر ماژول تبدیل نیز می باشد. ۳ جمله نیز بدون استفاده از هیچ گونه ماژولی توسط ناشنوایان قابل درک می باشد. به طور کلی برای ارزیابی کارایی سیستم با استفاده از ماژول تبدیل جملات ورودی از گفتار فارسی به همتای خود در دنیای ناشنوایان، هم چنین میزان تاثیر آن در دقت پردازش متن از روش‌های دقت و هم چنین صحت و در نهایت متد کارایی استفاده شده است.

$$\text{Precision} = \frac{15}{18} = 0.83 \quad (۵)$$

$$\text{Recall} = \frac{15}{21} = 0.71 \quad (۶)$$

$$\text{F-measure} = \frac{2*0.83*0.71}{0.83+0.71} = 0.76 \quad (۷)$$

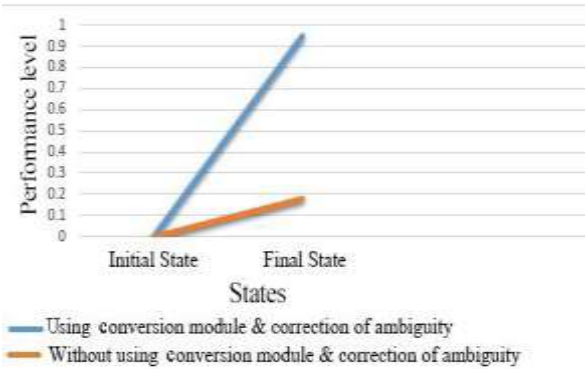
براساس این محاسبات استفاده از ماژول تبدیل جملات ورودی به دنیا ی ناشنوایان کارایی را به ۷۶٪ افزایش می دهد، این درحالی است که در صورت عدم استفاده از این ماژول همان گونه که در رابطه زیر مشاهده می کنید، کارایی ۲۲٪ می باشد.

$$\text{F-measure} = \frac{2*0.14*0.5}{0.14+0.5} = 0.22 \quad (۸)$$

بنابراین استفاده از ماژول تبدیل جملات ورودی به دنیای ناشنوایان به تنهایی کارایی سیستم مترجم متن فارسی به زبان اشاره فارسی را تقریبا ۳,۵ برابر می کند. این در حالی است که در صورت عدم استفاده از این ماژول سیستم مترجم خودکار متن فارسی به زبان اشاره فارسی کلمات را به دنیای ناشنوایان تبدیل نمی کند و سیستم مترجم برای نمایش این کلمات با مشکل مواجه می شود. همان گونه که در شکل ۱۷ نیز مشاهده می کنید.

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 ماژول تبدیل خودکار متن ورودی به دنیای ناشنوایان و بدون استفاده از این ماژول و دقت ترجمه خودکار سیستم را مشاهده می کنید. این ارزیابی برمبنای مجموع جملات تست ارزیابی شده است. در صورت عدم برطرف کردن جناس و هم‌چنین عدم تبدیل جملات به دنیای ناشنوایان تنها ۱۰٪ جملات برای افراد ناشنوایی که از زبان اشاره استاندارد استفاده می‌کنند، قابل درک خواهد بود. هم چنین 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، ۹۰٪ باقی‌مانده جملات نیز توسط افراد ناشنوا قابل درک خواهد بود. به طور کلی در نمودار ۱۸ تاثیر هریک از ماژول‌های 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 ۱۸ مشاهده می‌کنید

برای نمونه اگر تنها در سیستم مترجم از ماژول برطرف کردن جناس استفاده می‌شد، فقط ۱۰ جمله بدون نیاز به ماژول تبدیل و تنها به برطرف شدن مشکل جناس آن‌ها توسط ناشنوایان قابل استفاده می باشند، اما همان‌گونه که در نمودار مشاهده می‌کنید ستون نارنجی رنگ بیانگر این موضوع است که با استفاده از ماژول برطرف کردن جناس در مجموع، ۱۳ جمله برای ناشنوایان قابل درک می‌باشند،



شکل ۱۹: میزان کارایی سیستم با استفاده از هر دو ماژول
Fig. 19 : System performance using both modules

این نشان می‌دهد که ۳ جمله بدون استفاده از هیچ یک از ماژول‌ها قابل درک برای ناشنوایان می‌باشد. در استفاده از هردو ماژول، ستون آبی رنگ در نمودار بیانگر این موضوع است که ۲۸ جمله از مجموع تست با استفاده از دو ماژول اصلاح شدند، و توسط ناشنوایان قابل درک می باشند. ستون نارنجی رنگ در آن نشان می‌دهد که با اصلاح آن ۲۸ جمله در مجموع تمام ۳۱ جمله مجموع تست برای ناشنوایان 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 ناشنوایان، و هم چنین ماژول تشخیص و رفع ابهام و جناس کلمات استفاده شود، را مورد بررسی قرار می دهیم. به طور کلی ۳۱ جمله در مجموع تست قرار دارد، که ۲۸ جمله آن توسط ناشنوایان قابل درک نمی باشد و نیازمند استفاده از ماژول تبدیل و یا ماژول رفع جناس هستند. لازم به ذکر است که ۳ 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 هم چنین میزان تاثیر آن در دقت پردازش متن از متد های دقت و هم چنین صحت و در نهایت متد کارایی استفاده شده است.

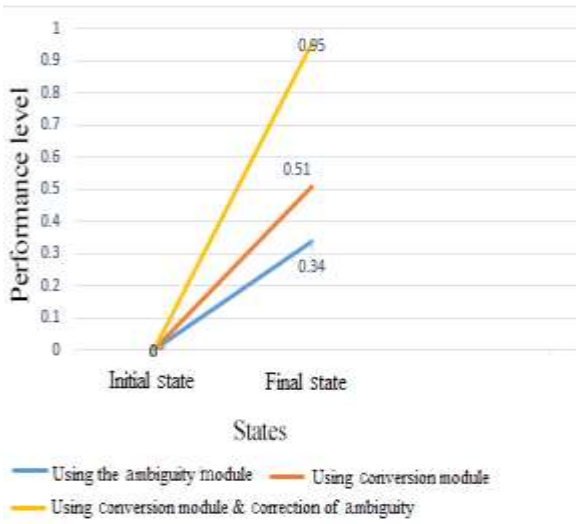
$$\text{Precision} = \frac{28}{28} = 1 \qquad (۹)$$

$$\text{Recall} = \frac{28}{31} = 0.90 \qquad (۱۰)$$

$$\text{F-measure} = \frac{2*0.9*1}{0.9+1} = 0.95 \qquad (۱۱)$$

براساس این محاسبات استفاده از هردو ماژول تبدیل جملات ورودی به دنیای ناشنوایان و ماژول رفع جناس کارایی را به ۹۵٪ افزایش می دهد، این درحالی است که در صورت عدم استفاده از این ماژول همان گونه که در رابطه زیر مشاهده می کنید، کارایی ۱۸٪ می باشد.

$$\text{F-measure} = \frac{2*0.97*1}{0.97+1} = 0.18 \qquad (۱۲)$$



شکل ۲۰: مقایسه میزان کارایی با استفاده از هریک از ماژول
Fig. 20: Compare performance by using any of the modules

بنابراین استفاده از هر دو ماژول تبدیل جملات ورودی به دنیای ناشنوایان و هم چنین ماژول رفع جناس کلمات کارایی سیستم مترجم متن فارسی به زبان اشاره فارسی را تقریبا ۵ برابر می کند. در نمودار شکل ۱۹ میزان کارایی سیستم مترجم با استفاده از ماژول تبدیل جملات ورودی به دنیای ناشنوایان هم چنین بدون استفاده از این ماژول را مشاهده می کنید.

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 ۲۰ به مقایسه میزان کارایی سیستم مترجم با استفاده از هریک ازماژول های تبدیل جملات ورودی به دنیای ناشنوایان و هم چنین تشخیص و رفع جناس کلمات و در نهایت ترکیب این دو ماژول می پردازیم.

همان طور که در شکل ۲۰ مشاهده می کنید، استفاده از ماژول تشخیص و رفع جناس به تنهایی کارایی سیستم را به ۳۴٪ افزایش می دهد، این درحالی است که استفاده از ماژول تبدیل جملات ورودی از گفتار فارسی به همتای خود در دنیای ناشنوایان کارایی سیستم را تا ۵۱٪ افزایش می دهد و در نهایت استفاده از ترکیب این دو ماژول باعث افزایش ۹۵٪ کارایی سیستم مترجم می شود.

نتیجه‌گیری

سیستم مترجم زبان اشاره به منظور برقراری ارتباط افراد ناشنوا با سایر افراد جامعه ایجاد شده است. به علاوه از سیستم مترجم زبان اشاره می‌توان به منظور آموزش زبان اشاره نیز بهره برد. پیاده‌سازی سیستم مترجم زبان اشاره فارسی به دلیل غیرساخت یافته بودن و هم‌چنین مشکل چندمعنایی و ابهام برخی کلمات در زبان فارسی، هم‌چنین مشکلات موجود در زبان اشاره فارسی و محدودیت‌هایی که در تعداد حرکات شناخته شده این زبان وجود دارد، دارای چالش‌های می‌باشد، که در این مقاله با ارائه یک معماری جدید چالش‌ها 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 می‌کند. با توجه به نتایج حاصل از ارزیابی سیستم مترجم زبان اشاره با استفاده از معماری جدید ۹۵٪ می‌باشد.

مشارکت نویسندگان

تمام نویسندگان به نسبت سهم برابر در این پژوهش مشارکت داشتند.

تشکر و قدردانی

از تمام کسانی که ما را در انجام این پژوهش یاری رساندند تشکر و قدردانی داریم.

تعارض و منافع

«هیچ گونه تعارض منافع توسط نویسندگان بیان نشده است.»

منابع و مآخذ

[1] Stokoe W. Dictionary of American sign language on linguistic principles. US: Linstok Press; 1976.

[2] Siavashi S. *Persian language and necessity of revision in education and assessment of Iranian deaf*. Paper presented in the 1st Conference of Iranian Linguistics Society. Tehran; 2005. Persian.

[3] Shamsfard M. Challenges and Open Problems in Persian Text Processing; 2005. Persian.

[4] Aleahmad. An introduction to semantic web. Tehran: University of Tehran Press; 2007. Persian.

[5] Michael C, Leo D, Obrst J, Smith K. The semantic web: a guide to the future of xml, web services and knowledge management; 2006.

[6] Amiri d Abdoli. *Design and implementation of course arak university.* Arak: Faculty of Engineering; 2009. Persian.

[7] van Zijl L, Barker D. *South African sign language machine translation system.* Paper presented at the 2nd International Conference on Computer Graphics, Virtual Reality, Visualization and Interaction in Africa; 2003.

[8] Baldassarri S, Cerezo E., Royo-santas F. *Automatic translation system to Spanish sign language with a virtual interpreter.* Paper presented at the IFIL Conference on Human-Computer Interaction; 2009.

[9] Dasgupta T, Basu A. An English to Indian sign language machine translation system. Dept. of Computer Science and engineering, Indian Institute of Technology, Kharagpur; 2010.

[10] Joy J, Balakrishnan K. *A prototype Malayalam to sign language automatic translator.* Paper presented at the IFIL Conference on Human-Computer Interaction; 2014.

[11] Herbich R, Graepel T. *Natural language processing*. UK: Taylor 7 Francis; 2010.

[12] Ming Jin C, Omar Z. A mobile application of american sign language translation via image processing algorithms. IEEE, Bali, Indonesia; 2016.

[13] López-lude˜na V, San-segundo R, Manuel Montero J, Córdoba R, Ferreiros J, Manuel Pardo J. Automatic categorization for improving Spanish into Spanish sign language machine translation. *Computer Speech & Language*. 2018; 26(3): 149-167.

[14] Dasgupta T, Anupam Basu SD. Prototype machine translation system from text-to-Indian sign. Asian Federation of Natural Language Processing; 2010.

[15] Ombardo V, Battaglino C, Damiano R. An avatar–based interface for the italian sign language. IEEE, Seoul, South Korea; 2011.

[16] Daud A, Khan W, Che D. Urdu language processing: a survey. Department of Computer Science and Software Engineering, Islamabad Pakistan; 2016.

[17]San-segundo R, Montero JM, Co´rdoba R, Sama V, Ferna´ndez F, D'haro LF, Lo´pez-luden˜a V, Sa´nchez D, Carcı´a A. Design, development and field evaluation of a Spanish into sign language translation system. Departamento de Ingeniería Electrónica, Etsi Telecomunicación Universidad Politécnica de Madrid Spain; 2012.

[18] Maritan Ugulino de Araújo T, Lacet Silva Ferreira F, Assis Nobre Dos Santos Silva D, Hermínio Lemos F, Pessoa Neto G. Automatic generation of Brazilian sign language windows for digital TV systems. Digital Video Applications Lab (LAVID), Federal University of Paraiba Brazil; 2013.

[19] Matthew J, Spirling DA. *Text preprocessing for unsupervised learning: why it matters, when it misleads, and what to do about.* Cambridge: Cambridge University Press; 2016.

[20] Hashimoto K, Xiong C, Tsuruoka Y, Socher R. A joint many-task model: Growing a neural network for multiple NLP tasks. Artificial Intelligence (cs.ai); 2017.

[21] Antoniou G, van Harmelen F. *A semantic web primer.* US: The MIT Press; 2008.

[22] Khaleghi G Deaf and sign language, limitations and features; 2010. Persian.

[23] Sstiri A. Natural language processing tools, the fourth annual workshop of web technology Lab, Ferdowsi University of Mashhad; 2012. Persian.

[24] Siavashi S. Deaf education, defending the reference language in education and eliminating discrimination. Journal of Information and Study of Disabilities and Disabilities. 2006; 1: 35-37. Persian.

[25] Bazrafkan M. *A review of Persian language processing problems using computer systems*. Paper presented in the National Conference on Computer Engineering and Information Technology Management; 2014. Persian.

Science and Technology Magazine. 2014; 1(7). Persian.

[29] Akramifard A. (2012) *Ontology representation*. Mashhad: International University of Ferdowsi Press. Persian.

[30] Kumar Verma V, Srivastava S. (2018). *Toward machine translation linguistic issues of Indian sign language*. Switzerland: Springer; 2018.

[26] *System analysis automatic find keywords Farsi.* Tehrna: IUST Press; 2009. Persian.

[27] Shamsfard M. (2006). *Persian text processing: Past achievements, challenges ahead.* The second workshop on Farsi and computer; 2006. Persian.

[28] Dokhani. A review of protégé Software. Iranian *Electronic*

Citation: (Vancoure): Letafat M, Rrasouli Konari A, Shamsi M. [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]. *Tech. Edu. J.* 2019; 13(4): 797-813.

http://dx.doi.org/10.22061/jte.2019.3369.1858

COPYRIGHTS

©2019 The author(s). This is an open access article distributed under the terms of the Creative Commons Attribution (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, as long as the original authors and source are cited. No permission is required from the authors or the publishers.