ارائهٔ راه‌کاری قاعده‌مند جهت تبدیل خودکار درخت تجزیهٔ نحوی وابستگی به درخت تجزیهٔ نحوی ساخت‌سازه‌ای برای زبان فارسی

فاطمه سلطان‌زاده[1]، محمد بحرانی[2] و محرم اسلامی[3]

[1و2] گروه زبان‌شناسی رایانشی ، مرکز زبان‌ها و زبان‌شناسی، دانشگاه صنعتی شریف، تهران، ایران

[3] گروه زبان و ادبیات فارسی، دانشکدهٔ علوم انسانی، دانشگاه زنجان، زنجان ایران

چکیده

هدف مقالهٔ حاضر ارائهٔ روشی خودکار جهت تبدیل درخت تجزیهٔ نحوی وابستگی به درخت تجزیهٔ نحوی مبتنی بر دستور ساخت‌سازه‌ای معادل در زبان فارسی است. برای نیل بدین هدف، الگوریتمی قاعده‌مند جهت این نوع تبدیل طراحی شده است. به‌منظور دستیابی به درخت تجزیهٔ نحوی ساخت‌سازه‌ای مناسب برای جملات زبان فارسی، دستور زبان فارسی مورد مطالعه قرار گرفته و قوانین خاص زبان فارسی استخراج و به الگوریتم مذکور افزوده شده است. نتایج حاصل از ارزیابی پژوهش، بیانگر این است که الگوریتم طراحی‌شده برای تبدیل ساختارها به F1 معادل با ۹۶/۰۵ درصد دست یافته است؛ سپس، الگوریتم مذکور بر روی پیکرهٔ وابستگی زبان فارسی با حدود سی‌هزار جمله اعمال و دادگان درخت نحوی ساخت‌سازه‌ای معادل حاصل شده است. علاوه‌بر این، تجزیه‌گر نحوی استنفورد بر روی پیکرهٔ ساخت‌سازه‌ای حاصل آموزش داده شده و تجزیه‌گر نحوی ترکیبی زبان فارسی با F1 معادل با ۸۶/۰۱ درصد به دست آمده است.

واژگان کلیدی: تبدیل، دستور وابستگی، دستور ساخت‌سازه‌ای ، پردازش زبان طبیعی، دادگان درخت نحوی، زبان فارسی.

۱- مقدّمه

در علوم زبان‌شناسی و زبان‌شناسی رایانشی[1] حجم عظیمی از متون ساخت‌یافتهٔ زبان را پیکره گویند. به عبارت دیگر، پیکرهٔ زبانی حجم زیادی از داده‌های زبانی است که براساس معیارهای مشخص برای هدف معینی جمع‌آوری و ذخیره شده‌اند؛ به‌طوری‌که نمایندهٔ زبان یا گویش مورد مطالعه باشند (اتکینز و دیگران، ۱۹۹۲). پیکره‌های زبانی، به‌منظور تحلیل‌های آماری، صحت‌سنجی فرضیه‌ها و بررسی رخداد یا صحت قواعد زبانی در حوزه‌های مشخص به کار می‌روند. بسیاری از پژوهش‌های زبان‌شناختی و تصمیم‌گیری‌ها در برنامه‌ریزی زبانی، تنها با استفاده از یک پیکرهٔ زبانی امکان‌پذیر است؛ بنابراین، پیکرهٔ زبانی پایه‌ای‌ترین و یکی از مهم‌ترین دادگان‌های ورودی در کاربردهای پردازش زبان طبیعی به‌شمار می‌رود (دبیرخانه شورای عالی اطلاع‌رسانی، ۱۳۸۸).

دادگان درخت نحوی[2] یک پیکرهٔ زبانی است که هرجملهٔ آن به‌صورت نحوی تجزیه شده است. در این پیکره‌ها ساختار نحوی هر جمله به‌طورمعمول به‌صورت درخت تجزیهٔ نحوی آن جمله نمایش داده می‌شود؛ به همین دلیل به این پیکره‌ها دادگان درخت نحوی می‌گویند.

اغلب دادگان‌های درخت نحوی براساس دو نظریهٔ دستوری وابستگی[3] و زایشی[4] شکل گرفته‌اند. در مورد نخست می‌توان به دادگان درخت نحوی وابستگی پراگ برای زبان چکی[5] (بوهموا و دیگران، ۲۰۰۳)، دادگان وابستگی زبان فارسی[6] (رسولی و دیگران، ۲۰۱۳)، پیکره وابستگی فارسی اوپسالا[7] (سراجی و دیگران، ۲۰۱۲) و پیکرهٔ وابستگی Dep-PerTreeBank (قیومی و کوهن، ۲۰۱۴) اشاره کرد. در مورد

[1] Computational Linguistics

[2] Treebank
[3] Dependency Grammar
[4] Generative Grammar
[5] Prague Dependency Treebank
[6] Persian Dependency Treebank
[7] Uppsala Persian Dependency Treebank

دوم نیز یکی از شناخته‌شده‌ترین موارد، دادگان درخت نحوی پن برای زبان انگلیسی[1] است (تیلور و دیگران، ۲۰۰۳). در مورد دادگان درخت نحوی بر مبنای دستور زایشی برای زبان فارسی، می‌توان به دادگان درخت نحوی فارسی Pertreebank اشاره کرد که در چارچوب دستور ساخت‌سازه‌ای هسته‌بنیان[2] تهیه شده و شامل حدود هزار جملهٔ برچسب‌خورده با اطلاعات نحوی است (قیومی، ۲۰۱۲ a). در ادامه به تفاوت‌های میان این دو دستور اشاره شده است.

دستور وابستگی و دستور زایشی هر دو از جمله نظریه های صورت‌گرا محسوب می‌شوند و هر دو نظریه می‌کوشند ساخت جمله را براساس پیوندهای گوناگونی که میان اجزا وجود دارد، نمایش دهند. در دستور زایشی فعل و فاعل به‌عنوان ارکان اصلی جمله قلمداد و جمله به دو بخش نهاد و گزاره تقسیم می‌شود؛ اما در دستور وابستگی این مفاهیم به کل کنار گذاشته می‌شوند و از آن‌ها به‌عنوان مفاهیمی یاد می‌شود که براساس ساخت اطلاعاتی جمله شکل گرفته‌اند و ربطی به ساختار نحوی زبان ندارند (طبیب‌زاده، ۱۳۸۳).

دستور وابستگی علی‌رغم داشتن ویژگی‌هایی مانند امکان نمایش بی‌ترتیبی زبان، دارای کاستی‌هایی در مقایسه با نظریهٔ دستور زایشی است. برای مثال همان‌گونه که می‌دانیم، یکی از اهداف پژوهش‌های زبان‌شناسی معاصر، ایجاد دستوری است که قادر به زایش جملات دستوری یک زبان باشد؛ اما دستور وابستگی قادر به برآورده‌کردن این هدف نیست. درحقیقت گرچه دستور وابستگی برای تجزیهٔ جملات بسیار کاربردی و اقتصادی است، اما یک جملهٔ تجزیه‌شده با استفاده از این دستور قادر به بازسازی مجدد نیست. از دیگر کاستی‌های این دستور در مقایسه با دستور زایشی، این است که در انگارهٔ وابستگی، بازنمایی ساخت‌های همپایه به‌درستی انجام نمی‌شود (اشناید، ۱۹۹۸).

بحث در تفاوت‌های دو نظریهٔ دستوری وابستگی و زایشی بسیار است و در ظرف مقالهٔ حاضر نمی‌گنجد (گایفمن، ۱۹۶۵)، اما علی‌رغم تمامی تفاوت‌های این دو دستور می‌بایست این نکته را در نظر داشت که این دو نظریه، دو دیدگاه متخاصم و مقابل نیستند و می‌توانند دو دیدگاه مکمل تلقی شوند. درحقیقت هر یک از آنها دارای ویژگی‌ها و مزایایی است که کاستی‌های نظریهٔ دیگر را جبران می‌کند. علاوه‌بر بحث نظری، در کاربردهای مختلف پردازش زبان طبیعی نیز سامانه‌های طراحی‌شده براساس دو نظریه می‌توانند کارایی‌های متفاوتی داشته باشند. آزمایش‌ها نشان داده‌اند که تجزیه‌گرهای[3] مبتنی بر این دو نظریه هر کدام به نوبهٔ خود می‌توانند کارا و موثر باشند. درحقیقت تجزیه‌گرهای نحوی مبتنی بر هر یک از این دو نظریه دارای توانایی‌های متفاوتی در تجزیهٔ ساخت‌های متمایز زبانی هستند (نیور و مک‌دونالود، ۲۰۰۸). با توجه به سودمندی هر یک از این تجزیه‌گرها در جایگاه خود، در روش‌های جدیدتر سعی بر آن است که مدل‌های ترکیبی از دو تجزیه‌گر مبتنی بر دستور زایشی و وابستگی ارائه شود؛ بدین ترتیب می‌توان از توانمندی‌های هر دو تجزیه‌گر در کنار هم بهره جست و بر کاستی‌های موجود چیره شد (کلاین و منینگ، ۲۰۰۳؛ رن و دیگران، ۲۰۱۳؛ فرکاس و بوهنت، ۲۰۱۲، لروکس و دیگران، ۲۰۱۲).

با توجه به موارد ذکرشده و به‌دلیل نبود دادگان درخت نحوی مبتنی بر دستور زایشی با حجم مناسب، پژوهش حاضر سعی بر آن داشت که دادگان درخت نحوی زایشی زبان فارسی را به‌طور خودکار و با استفاده از دادگان وابستگی زبان فارسی (رسولی و دیگران، ۲۰۱۳) تولید کند و سپس تجزیه‌گر نحوی حاصل از ترکیب این دو دستور را به‌کمک دادگان حاصل برای زبان فارسی به دست آورد.

همان‌گونه که می‌دانیم دستور زایشی یک مکتب بزرگ در زبان‌شناسی تلقی می‌شود و شامل انواع نظریه‌های زبانی چون دستور ساخت‌سازه‌ای[4]، دستور ساخت‌سازه‌ای هسته بنیان، دستور درخت الحاقی[5]، دستور واژی نقشی[6] و غیره است. ذکر این نکته حائز اهمیت است که بنیان نظری و زبان‌شناختی پژوهش حاضر بر دستور ساخت‌سازه‌ای استوار است.

در ادامه به مرور کارهای پیشین در راستای تبدیل دادگان‌های درخت نحوی به یکدیگر پرداخته و سپس راه‌کار مورد استفاده در این مقاله برای زبان فارسی را شرح می‌دهیم.

۲- مروری بر کارهای پیشین

امروزه تهیهٔ پیکره‌هایی با چندین نوع بازنمایی، بسیار متداول شده است. برای مثال می‌توان از پیکرهٔ HUTB[7] (پالمر و دیگران، ۲۰۰۹) نام برد که دارای سه نوع بازنمایی متفاوت

[1] Penn English Treebank
[2] Head-driven Phrase Structure Grammar (HPSG)

[3] Parser
[4] Phrase Structure Grammar
[5] Tree Adjoining Grammar
[6] Lexical Functional Grammar
[7] Hindi-Urdu Treebank

برای زبان هندی-اردو است؛ دو بازنمایی نحوی وابستگی-زایشی و یک بازنمایی معنایی. با توجه به اینکه درخت‌های نحوی مبتنی بر دو دیدگاه دستوری وابستگی و زایشی دارای این قابلیت هستند که توسط الگوریتم‌هایی به‌طور خودکار به یکدیگر تبدیل‌شوند؛ می‌توان با داشتن پیکره‌ای با یک شیوهٔ بازنمایی، پیکره‌ای با شیوهٔ بازنمایی دیگر را تولید کرد (کلاین، ۲۰۰۹). در ادامه به چندین پژوهش انجام‌یافته در راستای تبدیل درخت‌های تجزیهٔ نحوی مبتنی بر دستور وابستگی به زایشی به تفکیک روش‌های آماری[1] و قاعده‌مند[2] اشاره خواهد شد؛ سپس به مرور چندین پژوهش دیگر در راستای عکس این نوع تبدیل یعنی تبدیل درخت‌های تجزیهٔ نحوی مبتنی بر دستور زایشی به وابستگی خواهیم پرداخت.

۲-۱- تبدیل درخت تجزیهٔ نحوی وابستگی به زایشی

۲-۱-۱- روش‌های آماری

روش‌های آماری تبدیل درخت تجزیهٔ وابستگی به درخت تجزیهٔ زایشی در چندین پژوهش مورد استفاده قرارگرفته‌اند (کلاین، ۲۰۰۹؛ ونگ و دیگران، ۲۰۱۰). این پژوهش‌ها نیازمند یک پیکرهٔ زایشی هستند. در روش معرفی‌شده توسط کلاین[3] (کلاین، ۲۰۰۹) برای ساخت یک درخت تجزیهٔ نحوی زایشی از درخت تجزیهٔ نحوی وابستگی، در دادگان درخت نحوی بزرگ زایشی جستجو صورت می‌گیرد و به هر زیردرخت تجزیهٔ وابستگی مجموعه‌ای از زیردرخت‌های تجزیهٔ زایشی محتمل تخصیص داده می‌شود و سپس محتمل‌ترین درخت به‌عنوان درخت تجزیهٔ زایشی معادل برگزیده خواهد شد.

در پژوهشی دیگر (لی و دیگران، ۲۰۱۰) برای تبدیل درخت تجزیهٔ نحوی زایشی به درخت تجزیهٔ وابستگی معادل، با استفاده از روش آنتروپی بیشینه[4] هسته‌های روابط وابستگی در عبارات نحوی زایشی یافته شده و بدین‌ترتیب درخت وابستگی معادل به‌دست آمده است.

به گفتهٔ بات[5] (بات و دیگران، ۲۰۱۲) اکثر روش‌های آماری در تبدیل درخت وابستگی به زایشی شامل سه مرحله هستند. در مرحلهٔ نخست که مرحلهٔ آموزش[6] است، ورودی به‌صورت مجموعه‌ای از زوج‌های درخت تجزیهٔ وابستگی و درخت تجزیهٔ زایشی $\{(DS_i, PS_i)\}$ دریافت می‌شود. خروجی مرحلهٔ نخست یک مدل است. در مرحلهٔ دوم که همان مرحلهٔ آزمایش[7] است، مدل حاصل و درخت وابستگی مورد نظر (DS_t) دریافت شده و خروجی به شکل درخت تجزیهٔ زایشی $(PS_t^{(0)})$ تولید می‌شود. در مرحلهٔ سوم نیز $PS_t^{(0)}$ با داده‌های استاندارد یعنی PS_t مقایسه شده و نتایج ارزیابی گزارش خواهد شد.

بات و همکاران ادعا می‌کنند که روش‌های آماری معرفی‌شده در تبدیل درخت وابستگی به زایشی این مشکل را دارند که اگر در مرحلهٔ آموزش، داده‌های کافی وجود نداشته باشد، در آن صورت الگوریتم ممکن است، کارایی لازم را نداشته باشد؛ به همین سبب در ادامه به شرح روش‌های قاعده‌مند در این نوع تبدیلات خواهیم پرداخت.

۲-۱-۲- روش‌های قاعده‌مند

روش‌های قاعده‌مند تبدیل درخت تجزیهٔ وابستگی به زایشی در چندین پژوهش مورد استفاده قرار گرفته‌اند. برای مثال الدزبال[8] و همکاران با استفاده از جداولی، تمام روابط وابستگی را معرفی و سپس رابطهٔ زایشی معادل با هر رابطهٔ وابستگی را تعریف کرده‌اند (الدزبال و دیگران، ۲۰۰۸). نکتهٔ مهم در روش‌های قاعده‌مند این است که برخلاف روش‌های آماری که عملیات جستجو را در پیکره‌های بزرگ نحوی انجام می‌دهند، این روش‌ها از قوانینی آماده جهت تبدیل انواع درخت‌ها به یکدیگر بهره می‌گیرند.

در پژوهشی (ژیا و پالمر، ۲۰۰۱) سه الگوریتم متفاوت برای این نوع تبدیل معرفی و کارایی آن‌ها با یکدیگر مقایسه شده است. هر سه الگوریتم از روش‌های قاعده‌مند برای کار خود بهره می‌گیرند. در این الگوریتم‌ها قوانینی برای تبدیل درخت تجزیهٔ وابستگی به درخت تجزیهٔ زایشی تعریف شده است. با کمک این قوانین هر زیردرخت از درخت وابستگی به زیردرخت زایشی معادل با خود تبدیل و سپس کل درخت زایشی با اتصال مجموعهٔ زیردرخت‌ها به یکدیگر ساخته می‌شود. در پژوهشی دیگر (ژیا و دیگران، ۲۰۰۹) الگوریتمی جدید در راستای این نوع تبدیل طراحی شده است. این پژوهش برای تبدیل درخت وابستگی به درخت زایشی، از روشی قاعده‌مند در کنار اطلاعات آماری استفاده می‌کند. نتایج ارزیابی این روش ترکیبی، حاکی از این بوده است که

[1] Statistical Approach
[2] Rule-based Approach
[3] A. Klein
[4] Maximum Entropy
[5] R. Bhatt
[6] Train
[7] Test
[8] I. Aldezabal

الگوریتم جدید در مقایسه با سه الگوریتم پژوهش پیشین (ژیا و پالمر، ۲۰۰۱) به نتایج بهتری دست یافته است. بدیهی است که استفاده از روش قاعده‌مند به‌تنهایی در تبدیل درخت تجزیهٔ وابستگی به درخت تجزیهٔ زایشی خود دارای دشواری‌هایی است؛ چون استخراج قوانین زبانی از ساختارهای نحوی پیچیدهٔ یک زبان کاری بس دشوار و نیازمند به دانش افراد خبره و آشنا با نحو زبان است. همچنین استخراج کامل و جامع قوانین نحوی یک زبان با توجه به پیچیدگی‌های مختلف زبانی امری ناممکن است.

۲-۲- تبدیل درخت تجزیهٔ نحوی زایشی به وابستگی

در این بخش به اختصار چند نمونه از پژوهش‌های مرتبط با تبدیل درخت تجزیهٔ زایشی به وابستگی را مرور می‌کنیم.

در (بات، ۲۰۰۳) و (دوم و دیگران، ۲۰۰۴) درخت تجزیه نحوی زایشی برای زبان آلمانی به درخت تجزیه وابستگی معادل تبدیل شده است. پژوهش‌های دیگر نیز تلاش‌هایی ارزنده را در تبدیل درخت زایشی به وابستگی به ثمر رسانده‌اند و توانستند دادگان درخت نحوی زایشی پن را با دقتی بالا به دادگان درخت وابستگی معادل خود تبدیل کنند (ژبراتسکی و کوچروا، ۲۰۰۲؛ گلباخ و دیگران، ۲۰۰۵؛ ژوهانسون و نوگاس، ۲۰۰۷؛ لی و دیگران، ۲۰۱۰؛ کلینز، ۱۹۹۷).

در تبدیل درخت تجزیهٔ نحوی زایشی به وابستگی برای زبان فارسی می‌توان به پژوهش (قیومی و کوهن، ۲۰۱۴) اشاره کرد. در پژوهش نام‌برده الگوریتمی برای تبدیل درخت تجزیهٔ زایشی بر مبنای دستور ساخت‌سازه‌ای هسته‌بنیان به درخت وابستگی معادل معرفی شده و دادگان وابستگی Dep-PerTreeBankبرای زبان فارسی حاصل شده است.

در پژوهشی انجام‌شده برای تبدیل درخت تجزیهٔ زایشی به درخت وابستگی برای زبان عربی از پیکرهٔ PATB[1] استفاده شده است (الاباس و رامسی، ۲۰۱۲). این پیکره دارای حجمی بالغ بر پنج‌هزار جمله با میانگین طول ۲۸ کلمه است. لازم به ذکر است که جملات این پیکره براساس دستور زایشی از نوع ساخت‌سازه‌ای تجزیهٔ شده‌اند. جملات پیکرهٔ PATB به‌عنوان ورودی سامانهٔ پژوهش مذکور دریافت شده و توسط الگوریتمی به درخت تجزیهٔ وابستگی معادل تبدیل شده‌اند. روند این الگوریتم بدین شکل است که هرگاه فرآیند به گره برگ در درخت تجزیهٔ زایشی رسید، آن را بدون هیچ فرزندی به‌صورت مستقیم باز می‌گرداند. در غیر این صورت، گرهٔ هسته را تشخیص داده و زیردرختی را انتخاب می‌کند که شامل گرهٔ هسته باشد و سپس آن را به درخت وابستگی تبدیل می‌کند. بدین شکل تمامی زیردرخت‌ها را به زیردرخت‌های وابستگی تبدیل می‌کند. درنهایت نیز آنها را به یکدیگر پیوند داده و به گرهٔ هسته متصل می‌کند. بدین ترتیب کار را ادامه داده و درخت تجزیهٔ وابستگی کل جمله را به‌دست می‌آورد.

در ادامه به شرح راه‌کار قاعده‌مندی در جهت تبدیل درخت تجزیهٔ نحوی وابستگی به درخت تجزیهٔ نحوی زایشی از نوع ساخت‌سازه‌ای برای زبان فارسی خواهیم پرداخت؛ اما پیش از آن به‌اختصار دادگان مورد استفاده به‌عنوان ورودی سامانه را معرفی می‌کنیم.

۳- دادگان مورد استفاده

در پژوهش حاضر به‌عنوان ورودی سامانه از پیکرهٔ وابستگی زبان فارسی (رسولی و دیگران، ۲۰۱۳) استفاده شده است. پیکرهٔ وابستگی زبان فارسی شامل حدود سی‌هزار جمله است که توسط گروه پژوهشی دادگان و تحت نظارت شورای عالی اطلاع رسانی تهیه شده است. در این پروژه، ابتدا جملاتی به‌صورت تصادفی از وب جمع‌آوری و سپس برچسب‌زنی شده‌اند. پس از آن سه کار به‌صورت هم‌زمان انجام شده است؛ برچسب‌زنی جملات جدید در سطح صرف و نحو، ویرایش جملات برچسب‌زنی‌شده و پالایش آن از خطاها و هم‌چنین مطالعه در مورد نحوهٔ ارتقای این پیکره از سطح نحو به معنا. در این پیکره، ارتباطات بین اجزای جمله به‌صورت روابط هسته و وابسته نمایش داده شده است. هم‌چنین برچسب ویژگی‌های صرفی- نحوی هر کلمه مانند شخص، شمار، وجه، نمود و حالت (برای فعل) به کلمات موجود در این پیکره افزوده شده است. لازم به ذکر است که در طراحی این پیکره ۴۴ نوع رابطه وابستگی در نظر گرفته شده است.

تعداد انواع برچسب مقولهٔ نحوی دانه‌درشت[2] و دانه‌ریز[3] این پیکره به‌ترتیب ۱۷ و ۳۲ نوع است (رسولی و دیگران، ۱۳۹۳). در جدول (۱) اطلاعاتی از پیکرهٔ وابستگی شامل تعداد کل واژه‌ها، تعداد واژه‌های منحصربه‌فرد، تعداد کل جملات، میانگین طول جملات و غیره نمایش داده شده است. برای مثال در این پیکره جملهٔ "وی در اواخر عمر به‌علت کار

[1] Penn Arabic Treebank

[2] Coarse-Grained Part-of-Speech Tags

[3] Fine-Grained Part-of-Speech Tags

و مطالعهٔ زیاد نابینا شد." به مانند شکل (۱) و به قالب **CoNLL-X** (بوکلز و مارسی، ۲۰۰۶) نمایش داده شده است. همان‌طور که می‌دانیم، در این نوع بازنمایی، روابط وابستگی میان اجزای جمله به رابطهٔ میان هسته و عنصر وابسته آن اشاره دارند.

1	وی	وی	PR	SEPER	person=3|attachment=ISO|number=SING|senID=23687	12	SBJ	_	_
2	در	در	PREP	PREP	attachment=ISO|senID=23687	12	ADV	_	_
3	اواخر	آخر	N	IANM	attachment=ISO|number=PLUR|senID=23687	2	POSDEP	_	_
4	عمر	عمر	N	IANM	attachment=ISO|number=SING|senID=23687	3	MOZ	_	_
5	به	به	PREP	PREP	attachment=ISO|senID=23687	12	ADV	_	_
6	علت	علت	N	IANM	attachment=ISO|number=SING|senID=23687	5	POSDEP	_	_
7	کار	کار	N	IANM	attachment=ISO|number=SING|senID=23687	6	MOZ	_	_
8	و	و	CONJ	CONJ	attachment=ISO|senID=23687	7	NCONJ	_	_
9	مطالعهٔ	مطالعه	N	IANM	attachment=ISO|number=SING|senID=23687	8	POSDEP	_	_
10	زیاد	زیاد	ADJ	AJP	attachment=ISO|senID=23687	7	NPOSTMOD	_	_
11	نابینا	نابینا	ADJ	AJP	attachment=ISO|senID=23687	12	MOS	_	_
12	شد	کرد#کن	V	PASS	person=3|attachment=ISO|number=SING|tma=GS|senID=23687	0	ROOT	_	_
13	.	.	PUNC	PUNC	attachment=ISO|senID=23687	12	PUNC	_	_

(شکل- ۱): نمایش ساختار نحوی مبتنی بر دستور وابستگی برای جملهٔ "وی در اواخر عمر به‌علت کار و مطالعهٔ زیاد نابینا شد. "بنابر قالب CoNLL-X.

(جدول- ۱): فراوانی واژه‌ها و جملات در پیکرهٔ وابستگی زبان فارسی (رسولی و دیگران، ۱۳۹۳)

تعداد کل جملات	۲۹۹۸۲
میانگین طول جملات	۱۶,۶۱
تعداد کل واژه ها	۴۹۸۰۸۱
تعداد واژه‌های منحصر به فرد	۳۷۶۱۸
تعداد بن‌واژه‌های منحصر به‌فرد	۲۲۰۶۴

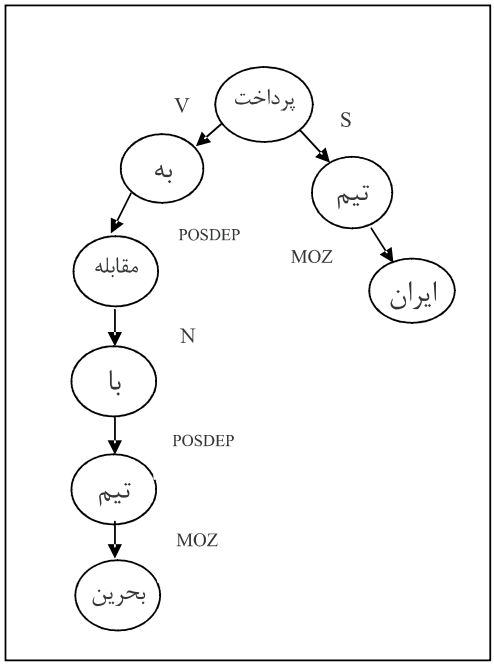
۴- تبدیل درخت تجزیهٔ نحوی وابستگی به ساخت‌سازه‌ای در زبان فارسی

در ابتدا ورودی سامانه به شکل یک جمله به همراه درخت تجزیهٔ وابستگی آن از دادگان درخت نحوی وابستگی زبان فارسی دریافت و تمامی روابط وابستگی موجود، از آن استخراج می‌شود؛ سپس الگوریتم تبدیل درخت تجزیهٔ وابستگی به درخت ساخت‌سازه‌ای آغاز به کار و هر زیردرخت وابستگی را به زیردرخت ساخت‌سازه‌ای معادل تبدیل می‌کند؛ پس از آن، زیردرخت‌های ساخت‌سازه‌ای حاصل را به یکدیگر متصل کرده و کل درخت تجزیهٔ نحوی ساخت‌سازه‌ای را پدید می‌آورد. در این مرحله قوانین تبدیل درخت وابستگی به ساخت‌سازه‌ای برای ساختارهای گوناگون زبان فارسی به الگوریتم افزوده شده و درنهایت درخت تجزیهٔ نحوی ساخت‌سازه‌ای زبان فارسی به‌عنوان خروجی سامانه بازگردانده می‌شود. در ادامه با ارائه مثال و با جزئیات بیشتری به شرح این مراحل خواهیم پرداخت. در ابتدا الگوریتم پایه جهت تبدیل درخت تجزیهٔ وابستگی به درخت ساخت‌سازه‌ای معادل را شرح می‌دهیم و سپس به قوانین خاص زبان فارسی جهت این نوع تبدیل اشاره خواهیم کرد.

۴-۱- الگوریتم پایه

جملهٔ "تیم ایران به مقابله با تیم بحرین پرداخت." را در نظر بگیرید. این ورودی حاوی اطلاعاتی همچون برچسب مقولهٔ نحوی دانه‌درشت و دانه‌ریز هر کلمه، عنصر هستهٔ هر کلمه، نوع وابستگی به آن و غیره است. در این مرحله روابط وابستگی موجود در جملهٔ ورودی استخراج و درخت وابستگی آن برای سامانه تعریف می‌شود. برای این منظور در هنگام پیاده‌سازی، برای هر کلمهٔ ورودی یک شیء از نوع گره تعریف

می‌شود. هر گره حاوی اطلاعاتی همچون شمارهٔ گره، کلمهٔ متناظر با گره، برچسب مقولهٔ نحوی کلمه متناظر با گره، گرهٔ هسته، نوع رابطهٔ وابستگی با گرهٔ هسته، گره‌های فرزند و غیره است. بدین شکل با تعریف اشیایی از جنس گره و مشخص‌کردن روابط بین آن‌ها گراف متناظر با جملهٔ ورودی مطابق با شکل (۲) تشکیل می‌شود.



(شکل- ۲): درخت تجزیهٔ وابستگی برای جملهٔ "تیم ایران به مقابله با تیم بحرین پرداخت."

پس از تعریف روابط وابستگی برای سامانه، الگوریتم تبدیل درخت تجزیهٔ نحوی وابستگی به ساخت‌سازه‌ای آغاز به کار و هر رابطهٔ وابستگی در درخت تجزیهٔ وابستگی را به رابطهٔ معادل در درخت ساخت‌سازه‌ای تبدیل می‌کند. درحقیقت الگوریتم از ریشهٔ درخت وابستگی یعنی "پرداخت" کار خود را آغاز می‌کند. برای این کار پشته‌ای[1] را تعریف کرده و شمارهٔ ریشه را به پشته وارد می‌کند؛ سپس به سمت گره‌های فرزند "پرداخت" در درخت وابستگی پایین می‌رود، آن‌ها را به‌ترتیب و از سمت چپ به راست به پشته وارد می‌کند. برای مثال در جملهٔ بالا پس از گره "پرداخت"، گره‌های "به" و "تیم" را به‌ترتیب وارد پشته می‌کند. حال نوبت به تحلیل عنصر بالای پشته یعنی "تیم" می‌رسد. الگوریتم در درخت پایین رفته و فرزندان این گره را مورد بررسی قرار می‌دهد. بدین شکل به گره "ایران" می‌رسد و شمارهٔ گره "ایران" را وارد پشته می‌کند. این گره یک گرهٔ برگ و مطلوب عملیات جستجو است. همان‌طور که می‌بینید عملیات جستجو به روش الگوریتم جستجوی نخست عمق[2] و به‌منظور رسیدن به گره برگی که تا به حال ملاقات نشده (در این مثال گره "ایران") انجام می‌شود. الگوریتم پس از رسیدن به گره برگ "ایران" گروه اسمی متناظر با آن را تشکیل داده و زیردرخت ساخت‌سازه‌ای معادل آن را مطابق با گام نخست شکل ۳ می‌سازد. درحقیقت این کار با توجه به ورودی سامانه که درخت وابستگی به همراه برچسب مقوله نحوی کلمه است، صورت می‌گیرد به این معنا که گره پدر یک کلمه با برچسب مقولهٔ نحوی اسم، گرهی از نوع گروه اسمی است.

الگوریتم پس از این مرحله گره "ایران" را از پشته خارج می‌کند و آن را به فهرست گره‌های بررسی‌شده می‌افزاید. در مرحلهٔ بعد عنصر بالای پشته گره "تیم" است. این گره تنها دارای یک فرزند به نام "ایران" است که در مرحلهٔ پیش بررسی شد، پس تمامی فرزندان آن مورد بررسی قرار گرفته‌اند. حال نوبت به تشکیل زیردرخت ساخت‌سازه‌ای با ریشهٔ "تیم" فرا می‌رسد. برای این کار ابتدا گرهٔ ساخت‌سازه‌ای کلمهٔ "تیم" به گره ساخت‌سازه‌ای متناظر با برچسب مقولهٔ نحوی آن متصل می‌شود و سپس کل درخت، از اتصال گره ساخت‌سازه‌ای متناظر با برچسب مقولهٔ نحوی کلمه و زیردرخت متناظر با "ایران" (تشکیل‌شده در مرحلهٔ قبل) حاصل می‌شود. بدین دلیل که کلمه "ایران" در جمله، پس از کلمه "تیم" آمده است، این گره در سمت چپ زیردرخت "تیم ایران" قرار می‌گیرد. لازم به ذکر است که هستهٔ اصلی گروه اسمی "تیم ایران" اسم "تیم" است، به همین دلیل این کلمه در سطح سوم و زیر گره اسم (N) قرار می‌گیرد. همان‌طور که می‌دانیم، اسم "ایران" هستهٔ گروه اسمی "تیم ایران" نیست؛ بلکه وابسته به اسم "تیم" از نوع متمم[3] اسمی یا همان مضافٌالیه است؛ و از آنجایی‌که وابسته‌ها به‌صورت گروه ظاهر می‌شوند، گروه اسمی "ایران" (NP) در سطح دوم و در زیر گروه اسمی بزرگ‌تر "تیم ایران" قرار می‌گیرد و چون فاقد وابسته است در سطح پایین‌تر به‌صورت واژگانی (N) نمود پیدا می‌کند. درخت حاصل از این فرآیند در مرحلهٔ دوم از شکل (۳) ترسیم شده است؛ سپس گره "تیم" به فهرست گره‌های بررسی‌شده افزوده‌شده و از پشته خارج می‌شود.

[1] Stack

[2] Depth First Search

[3] Complement

پس از اینکه نخستین فرزند گره "پرداخت" بررسی و از پشته خارج شد، الگوریتم به سراغ فرزند دیگر آن یعنی گره "به" رفته و فرایند جستجوی گره برگ را برای آن تکرار می‌کند. لازم به ذکر است که در این زمان عنصر بالای پشته همان گره "به" است. الگوریتم فرزند گره "به" یعنی گره "مقابله" را وارد پشته می‌کند و به سمت فرزند آن در درخت پایین می‌رود. الگوریتم کار را به همین شکل ادامه داده و فرزند هر گره را وارد پشته می‌کند و در انتها به گره "بحرین" می‌رسد. با توجه به اینکه این گره یک گره برگ است، الگوریتم زیردرخت ساخت‌سازه‌ای آن را می‌سازد؛ سپس کل زیردرخت را به همین شکل از پایین به بالا درست می‌کند. به‌دلیل مشابهت این فرایند با فرایند توصیف‌شده در مرحلهٔ قبل از ذکر جزییات آن صرف‌نظر می‌کنیم. درنهایت زیردرخت ساخت‌سازه‌ای متناظر با گره وابستگی "به" مطابق با مرحلهٔ سوم از شکل (۳) تشکیل می‌شود.

تا به این‌جا زیردرخت‌های ساخت‌سازه‌ای تمامی فرزندان گره "پرداخت" تشکیل شدند. حال زمان تشکیل زیردرخت ساخت‌سازه‌ای گره "پرداخت" فرا می‌رسد. در ابتدا گره‌هایی در درخت ساخت‌سازه‌ای برای کلمه "پرداخت" و برچسب مقولهٔ نحوی آن شکل می‌گیرد و سپس با اتصال این قطعه و همچنین زیردرخت به‌دست‌آمده در مرحلهٔ ۳ از شکل (۳)، به گره "VP" کل زیردرخت حاصل می‌شود. این زیردرخت در شکل (۴) ترسیم شده است.

با توجه به اینکه یکی از وابسته‌های فعل "پرداخت" با رابطهٔ فاعلی به آن متصل شده است، در این مرحله می‌بایست گرهی از نوع جمله یعنی "S" به وجود آید و به‌ترتیب زیردرخت "تیم ایران" به‌عنوان گروه اسمی در نقش فاعل و زیردرخت "VP"به‌عنوان گروه فعلی به آن متصل شوند. بدین ترتیب درخت تجزیهٔ ساخت‌سازه‌ای معادل با جملهٔ "تیم ایران به مقابله با تیم بحرین پرداخت" مطابق با شکل ۴ حاصل می‌شود. شایان ذکر است که الگوریتم پایه به‌تنهایی برای هیجده نوع رابطهٔ وابستگی از ۴۴ نوع رابطهٔ وابستگی به‌کار می‌آید. این روابط در جدول (۲) آورده شده‌اند.

۴-۲- استخراج قوانین خاص زبان فارسی

همان‌طور که پیشتر اشاره شد، الگوریتم تشریح‌شده تا به اینجا تنها برای برخی از ساختارهای زبانی کارا است. درحقیقت این الگوریتم، تنها می‌تواند به‌عنوان یک پایه و زیرساخت برای سایر ساختارها تلقی شود. برای سایر روابط وابستگی با توجه به ذات این روابط و کاستی‌های الگوریتم در تبدیل آنها، دستور زبان فارسی مطالعه شد و قوانین تبدیل زیردرخت وابستگی به ساخت‌سازه‌ای برای این روابط و یا ترکیب این روابط به دست آمد.

(جدول- ۲): روابط وابستگی بی‌نیاز از قوانین تبدیل خاص زبان فارسی

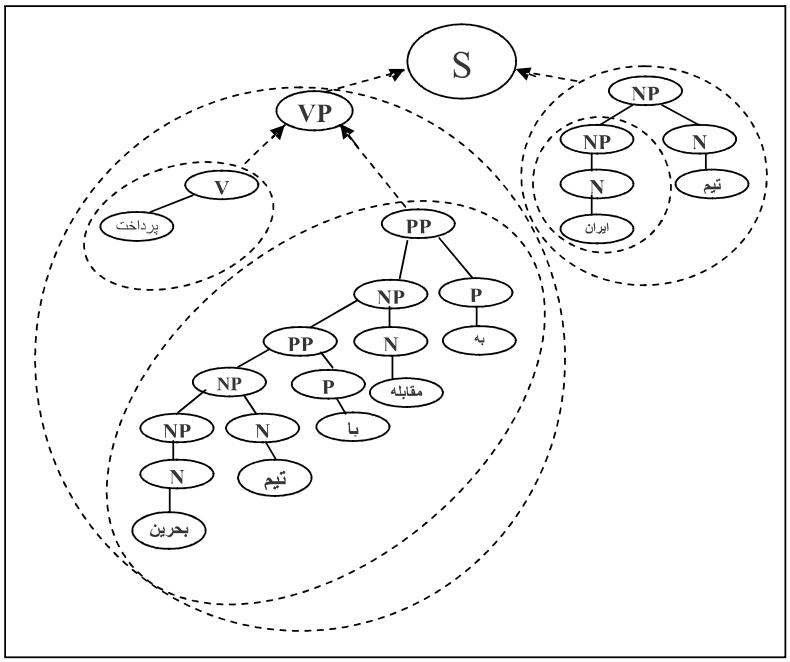
توضیح	اختصار
متمم حرف اضافه‌ای صفت	AJPP
بدل	APP
قید	ADV
وابستهٔ پیشین صفت	APREMOD
وابستهٔ پسین صفت	APOSTMOD
حرف اضافهٔ تفضیلی	COMPPP
مسند	MOS
مضافٌ الیه	MOZ
متمم نشانهٔ اضافه‌ای صفت	NEZ
صفت پیشین اسم	NPREMOD
صفت پسین اسم	NPOSTMOD
حرف اضافهٔ اسم	NPP
مفعول	OBJ
مفعول دوم	OBJ2
ریشهٔ جمله	ROOT
فاعل	SBJ
تمیز	TAM
مفعول حرف اضافه ای	VPP

پس از حصول قوانین تبدیل، این قوانین به الگوریتم بالا افزوده شد. در تهیه و پیاده‌سازی قوانین تبدیل درخت تجزیهٔ وابستگی به ساخت‌سازه‌ای، تلاش شد که روش ارائه‌شده، ساختارهای نحوی گوناگونی را پوشش دهد. در ادامه به‌صورت خلاصه به برخی از این قوانین اشاره می‌کنیم:

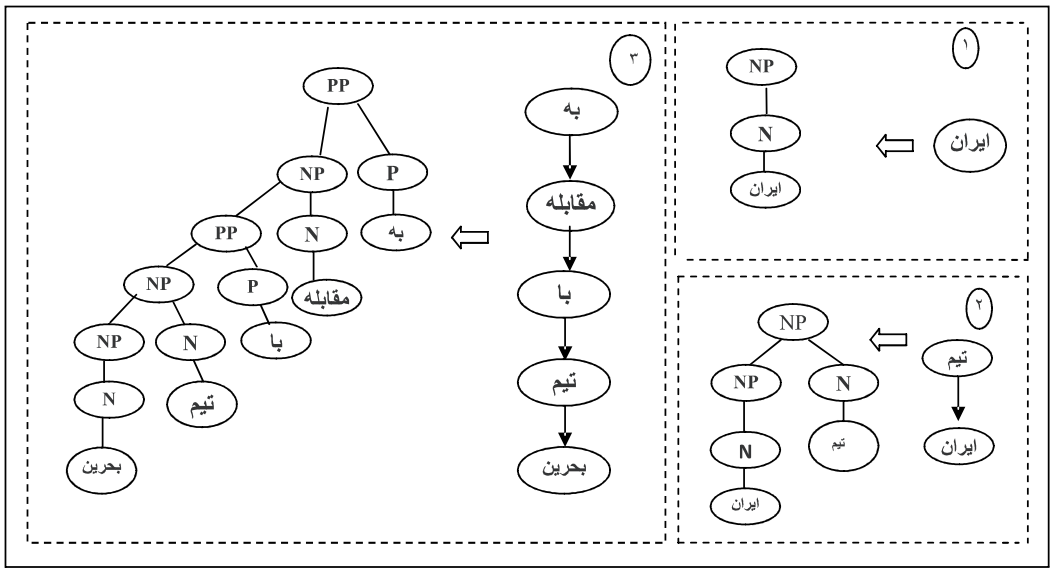
۱. قانون تبدیل ممیز:

اعداد اصلی به همراه ممیز و یا بدون آن به‌عنوان وابستهٔ پیشین اسم مورد استفاده قرار می‌گیرند (مشکوه‌الدینی، ۱۳۸۹). به‌عنوان مثال گروه اسمی "سه ماه کار" را در جملهٔ "او ترجیح می‌دهد بعد از سه ماه کار استراحت کند"، در نظر بگیرید که در آن درخت وابستگی به شکلی به درخت ساخت‌سازه‌ای تبدیل می‌شود که در آن عبارت "سه ماه" در درون گروه اسمی یک گروه صفت عددی[1] را تشکیل می‌دهد و هستهٔ "کار" را توصیف می‌کند.

[1] Numeral Phrase



(شکل- ۳): روند تبدیل درخت تجزیهٔ نحوی وابستگی به ساخت‌سازه‌ای برای جملهٔ "تیم ایران به مقابله با تیم بحرین پرداخت."



(شکل- ۴): درخت تجزیهٔ نحوی ساخت‌سازه‌ای برای جملهٔ "تیم ایران به مقابله با تیم بحرین پرداخت."

۲. قانون تبدیل ناهمپایگی برای بند متممی فعل:

طبیب‌زاده متمم بندی را متمم مستقلی در ساخت ظرفیتی[1] برخی افعال در نظر می‌گیرد و دلایلی را در تأیید نظر خود دارد که خارج از بحث ماست (طبیب‌زاده، ۱۳۹۰). در این نوع نگاه، جمله‌وارهٔ وابستهٔ "انسان خلیفهٔ خدا باشد" در جملهٔ "اسلام می‌خواهد که انسان خلیفهٔ خدا باشد"، به‌عنوان بند متممی فعل "می‌خواهد" تلقی می‌شود. در درخت وابستگی این جمله، فعل "باشد" وابستهٔ فعل "می‌خواهد" از نوع بند متممی است. در درخت ساخت‌سازه‌ای این جمله، جمله‌وارهٔ پیرو "انسان خلیفهٔ خدا باشد" در درون جمله‌وارهٔ اصلی قرار می‌گیرد. در پژوهش

[1] Valency Structure

حاضر قانون تبدیلی برای رابطهٔ وابستگی بند متممی فعل به رابطهٔ ساخت‌سازه‌ای معادل طراحی شده است.

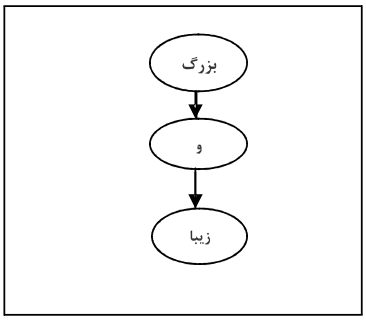
۳. قانون تبدیل حرف نشانهٔ "را":

"را" واژهٔ دستوری پسین است که تنها به‌دنبال گروه اسمی ممکن است ظاهر شود (مشکوه‌الدینی، ۱۳۸۹). برای مثال هستهٔ "خواندن" در جملهٔ "من کتاب را خواندم" علاوه‌بر فاعل از نظر وابستگی لزوماً یک وابسته دارد که در اینجا "کتاب را" متمم اجباری آن محسوب می‌شود؛ لذا در این پژوهش قانونی برای تبدیل حرف نشانهٔ "را" در نظر گرفتیم.

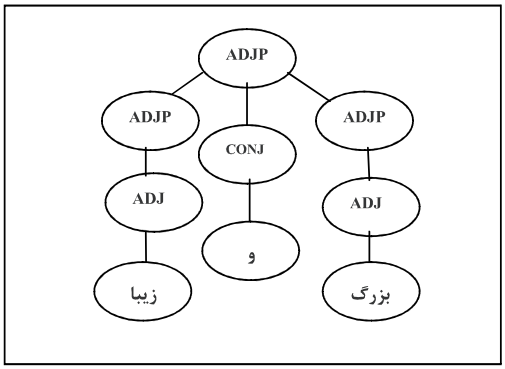
۴. قوانین تبدیل همپایگی:

پیوستن مقوله‌ها یا جمله‌ها به یکدیگر را با واژه‌هایی مانند "و"، "یا"، "اما"، "لیکن" و غیره همپایه‌سازی گویند (قریب، ۱۳۶۸). یکی از مهم‌ترین ویژگی گروه‌های همپایه این است که همواره دو یا چند گروه یکسان همپایه می‌شوند (طبیب‌زاده، ۱۳۹۰). در پژوهش حاضر قوانین تبدیل روابط وابستگی به ساخت‌سازه‌ای برای ساخت‌های همپایگی اسم، فعل، صفت، قید و جمله پیاده‌سازی شده‌است. به‌عنوان مثال برای همپایگی گروه صفتی "بزرگ و زیبا" نمودار وابستگی نمایش داده‌شده در شکل (۵) به نمودار ساخت‌سازه‌ای شکل (۶) تبدیل شده است.

همان‌طور که می‌دانید در پژوهش حاضر از برچسب مقولهٔ نحوی کلمات و روابط وابستگی بین اجزای جمله در پیکرهٔ وابستگی زبان فارسی به‌عنوان ورودی سامانه استفاده شده است. بنابراین برچسب‌دهی مقولهٔ نحوی کلمه و نحوهٔ تحلیل ساخت‌های همپایه براساس ورودی سامانه که پیکرهٔ وابستگی زبان فارسی (رسولی و دیگران، ۲۰۱۳) است صورت گرفته که متفاوت با پژوهش (قیومی و کوهن، ۲۰۱۴) است.



(شکل- ۵): درخت تجزیهٔ وابستگی برای گروه صفتی "بزرگ و زیبا"



(شکل- ۶): درخت تجزیهٔ ساخت‌سازه‌ای برای گروه صفتی "بزرگ و زیبا"

همچنین برای گروه‌های همپایگی پیچیده‌تر که عناصر همپایه خود دارای وابسته‌های گوناگونی باشند و یا یک گروه همپایگی در درون گروه همپایگی دیگر قرار داشته باشد نیز الگوریتمی طراحی و به سامانه افزوده شده است. برای مثال جملهٔ "اسلام مردم را به تمدن عقلانی، عقل ابزاری، تقسیم کار، آبادی زمین و پیشرفت مدنی و اقتصادی تشویق می‌کند" را در نظر بگیرید. در این مثال همپایگی گروه‌های اسمی "تمدن عقلانی، عقل ابزاری، تقسیم کار، آبادی زمین و پیشرفت مدنی و اقتصادی" وجود دارد که در این گروه‌ها، وابسته‌ها از نوع متمم (مضافٌ‌الیه) و آویزه[1] (صفت در این مثال) هستند. همچنین در درون این گروه اسمی یک گروه صفتی همپایه (مدنی و اقتصادی) نیز قرار دارد. لازم به ذکر است که برای پیاده‌سازی قوانین تبدیل هر کدام از گروه‌های اسمی، فعلی، صفتی، قیدی و همچنین همپایگی جملات تابع مجزایی طراحی شده که روند کلی آنها بسیار مشابه یکدیگر است. طرز کار این توابع را با ذکر مثال گروه همپایگی اسمی "تمدن عقلانی، عقل ابزاری، تقسیم کار، آبادی زمین و پیشرفت مدنی و اقتصادی" بیان می‌کنیم.

در ابتدا برای یافتن عناصر همپایه در جمله جستجو صورت می‌گیرد و اسامی "تمدن"، "عقل"، "تقسیم"، "آبادی" و "پیشرفت" به‌عنوان نتایج جستجو بازگردانده و ذخیره می‌شوند. پس از اینکه کل گروه اسمی همپایه یافت شد، از وابسته‌های هر عنصر همپایه صرف نظر شده و عناصر همپایه به‌ترتیب وارد یک پشته می‌شوند، به‌طوری‌که نخستین عنصر همپایه (در این مثال "تمدن") در بالای پشته قرار می‌گیرد. پس از قراردان عناصر در پشتهٔ مورد نظر

[1] Adjunct

ابتدا عنصر بالای پشته از پشته خارج می‌شود و به همراه وابسته‌های خود یک گروه اسمی را تشکیل می‌دهد. برای مثال گروه اسمی "تمدن عقلانی" تشکیل می‌شود. به همین ترتیب برای سایر عناصر همپایه که در پشته قرار دارند نیز گروه اسمی متناظر تشکیل می‌شود. بدین طریق گروه‌های اسمی "عقل ابزاری"، "تقسیم کار"، "آبادی زمین" و "پیشرفت مدنی و اقتصادی" به‌ترتیب تشکیل خواهند شد. ذکر نکته‌ای که در اینجا حائز اهمیت است و آن اینکه در هنگام تشکیل گروه اسمی"پیشرفت مدنی و اقتصادی"، حضور گروه صفتی همپایه "مدنی و اقتصادی" در درون این گروه اسمی تشخیص داده می‌شود. برای تشکیل این گروه نیز تابع همپایگی صفت فراخوانده می‌شود. روال تشکیل این گروه همپایه نیز مانند تابع همپایگی اسمی است، با این تفاوت که این‌بار صفت و روابط صفتی در تشکیل این گروه دخالت دارند؛ پس از تشکیل گروه صفتی همپایهٔ "مدنی و اقتصادی"، گروه اسمی "پیشرفت مدنی و اقتصادی" نیز شکل می‌گیرد و بدین صورت کل گروه اسمی "تمدن عقلانی، عقل ابزاری، تقسیم کار، آبادی زمین و پیشرفت مدنی و اقتصادی" تشکیل می‌شود. درخت ساخت‌سازه‌ای حاصل از روال توصیف‌شده برای این گروه اسمی به قالب دادگان درخت نحوی پن در شکل (۷) نمایش داده شده است.

۵- قانون تبدیل جابه‌جایی[1] بند متممی صفت

بنابر گفته طبیب‌زاده متمم بندی صفت، یکی از انواع متمم‌های صفت است. صفت‌هایی چون"آگاه"، "شاهد"، "منکر" و "موافق"، متمم بندی می‌گیرند (طبیب‌زاده، ۱۳۹۰). جملهٔ "او آگاه است که مادرش بیمار است" را در نظر بگیرید. در این جمله صفت "آگاه" دارای یک وابسته از نوع متمم بندی است. همان‌طور که می‌بینید جملهٔ وابسته به جایگاه پس از فعل جملهٔ اصلی انتقال یافته و به‌صورت گسسته از هستهٔ خود ظاهر شده است. لازم به ذکر است که در پژوهش حاضر، تشخیص جابه‌جایی بند متممی صفت به‌کمک رابطهٔ وابستگی متناظر و همچنین ترتیب کلمات موجود در جمله صورت می‌گیرد. به این شکل که اگر مکان قرارگرفتن بند متممی صفت پس از فعل جمله اصلی باشد، در این صورت جابه‌جایی بند متممی صفت رخ داده است.

[1] Extraposition

باید به این نکته توجه داشت که در مکان اصلی بند متممی صفت نیز رد[2] به جا گذاشته می‌شود. در شکل ۸ درخت ساخت‌سازه‌ای برای جملهٔ "او آگاه است که مادرش بیمار است" نمایش داده شده است.

```
(NP
    (NP (N تمدن)
            (ADJP (ADJ عقلانی)))
        (PUNC ،)
    (NP (N عقل)
            (ADJP (ADJ ابزاری)))
        (PUNC ،)
    (NP (N تقسیم )
            (NP (N کار)))
    (PUNC ،)
            (NP (N آبادی)
                    (NP (N زمین)))
    (CONJ و)
            (NP (N پیشرفت)
                (ADJP
                    (ADJP (ADJ مدنی))
                    (CONJ و)
                    (ADJP (ADJ اقتصادی)))))
```

(شکل- ۷): نمایش درخت تجزیهٔ نحوی ساخت‌سازه‌ای برای گروه اسمی "تمدن عقلانی، عقل ابزاری، تقسیم کار، آبادی زمین و پیشرفت مدنی و اقتصادی" بنابر قالب دادگان درخت نحوی پن.

لازم به ذکر است که توضیح جامع سایر قوانین طراحی‌شده در ظرف مقالهٔ حاضر نمی‌گنجد؛ ولی برای آشنایی با این قوانین، در ادامه فهرست آنها آورده شده است:

۵. قانون تبدیل افعال وجهی
۶. قانون تبدیل ناهمپایگی برای بند موصولی و بند متممی اسم

[2] Trace

۷. قانون تبدیل ناهمپایگی برای بند افزودهٔ فعل
۸. قانون تبدیل قید جمله
۹. قانون تبدیل افعال کمکی
۱۰. قانون تبدیل افعال مرکب
۱۱. قانون تبدیل مصدرهای مرکب
۱۲. قانون تبدیل جابه‌جایی بند موصولی و بند متممی اسم
۱۳. قوانین تبدیل فاعل تهی
۱۴. قوانین تبدیل حذف پیوند

در این پژوهش الگوریتم تبدیل درخت وابستگی به ساخت‌سازه‌ای به شکلی توسعه داده شد که ساختارهایی را که حاصل مجموعه‌ای پیچیده از روابط وابستگی است، تحت پوشش خود قرار دهد. بدین‌ترتیب اگر الگوریتم با چند رابطهٔ وابستگی در یک زیردرخت وابستگی روبه‌رو شد، آنها را به روابط معادل در زیردرخت ساخت‌سازه‌ای تبدیل خواهد کرد. در ادامه به برخی از قوانین ترکیبی به‌صورت خلاصه و با ارائهٔ یک مثال اشاره خواهیم کرد.

- ساخت همپایه در درون ساخت همپایه‌ای دیگر:

مثال: [[اتومبیلی جلوی نانوایی پارک کرد] و [خانمی [با عجله بیرون آمد و یک عدد نان خرید]]].

توضیح: در این مثال دو جملهٔ "اتومبیلی جلوی نانوایی پارک کرد" و "خانمی با عجله بیرون آمد و یک عدد نان خرید" با یکدیگر همپایه شده‌اند. همچنین جملهٔ دوم از همپایه‌سازی دو گروه فعلی "با عجله بیرون آمد " و "یک عدد نان خرید" تشکیل شد ه است.

- ساخت ناهمپایه در درون ساخت ناهمپایهٔ دیگر:

مثال: فکر می‌کنم [ما اجازه نداشته‌باشیم [با دانش اندک خودمان نظر یک استاد را زیر سؤال ببریم]].

توضیح: در این مثال جملهٔ "ما اجازه نداشته باشیم با دانش اندک خودمان نظر یک استاد را زیر سؤال ببریم" وابسته‌ای از نوع بند متممی برای فعل "فکر می‌کنم" و "با دانش اندک خودمان نظر یک استاد را زیر سؤال ببریم" نیز وابسته‌ای از نوع بند متممی برای فعل "اجازه نداشته باشیم" است. همان‌گونه که ملاحظه می‌کنید در این جمله، یک بند متممی در درون بند متممی دیگر قرار گرفته است.

- جابه‌جایی بند موصولی وابستهٔ اسم:

مثال: وی اظهار داشت [کمیتهٔ داوری جایزهٔ صلح نوبل را احزابی برمی‌گمارند [که در پارلمان نروژ حضور دارند]].

توضیح: زیرساخت جمله بالا به این صورت است: "وی اظهار داشت (که) کمیتهٔ داوری جایزهٔ صلح نوبل را احزابی که در پارلمان نروژ حضور دارند، برمی‌گمارند". در این مثال، جملهٔ "که در پارلمان نروژ حضور دارند"، وابسته‌ای از نوع بند موصولی برای اسم "احزاب" و جملهٔ "کمیتهٔ داوری جایزهٔ صلح نوبل را احزابی برمی‌گمارند که در پارلمان نروژ حضور دارند"، وابستهٔ "اظهار داشت" است. همچنین همان‌گونه که ملاحظه می‌کنید، بند موصولی اسم از مکان اصلی خود جابه‌جا شده است.

- ساخت همپایه در درون بند متممی فعل:

مثال: دبیر شورای عالی امنیت ملی گفت [مردم در این انتخابات شکوه جمهوریت نظام اسلامی را به رخ جهانیان کشاندند و دشمنان جمهوری اسلامی با تمام توان به مقابله با آن برخاستند].

توضیح: در این مثال جملهٔ "مردم در این انتخابات شکوه جمهوریت نظام اسلامی را به رخ جهانیان کشاندند و دشمنان جمهوری اسلامی با تمام توان به مقابله با آن برخاستند" وابسته‌ای از نوع بند متممی برای فعل "گفت" است و خود از ساخت همپایه تشکیل شده‌است.

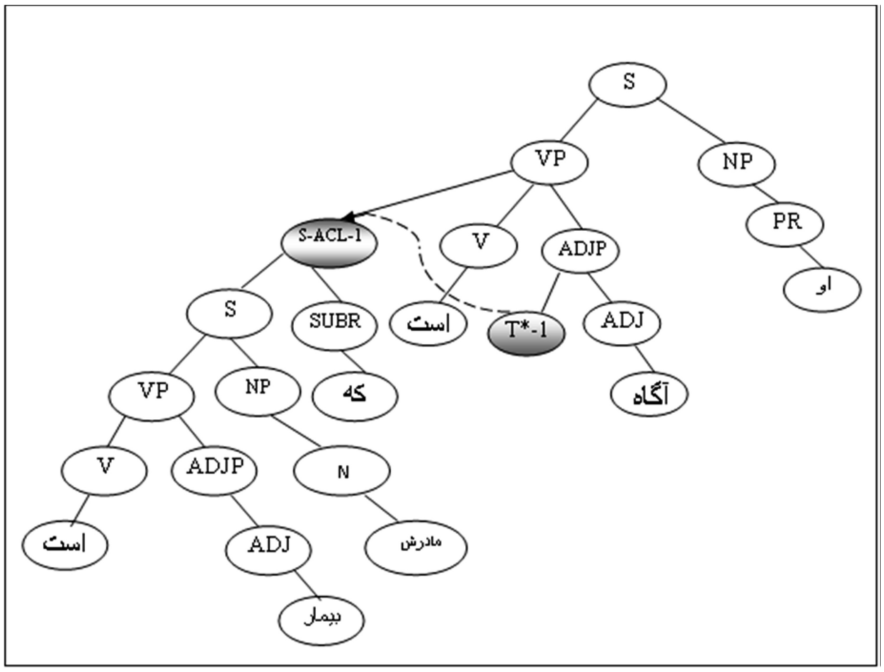
- ساخت ناهمپایه در درون ساخت همپایه:

مثال: حکیمی از رشد ۲۱ درصدی جابجایی مسافر از بنادر مسافری هرمزگان خبرداد و افزود [در این مدت پنج میلیون و ۷۷۱ هزار و ۳۷۷ مسافر از طریق بنادر مسافری استان جابه‌جا شدند].

توضیح: در این مثال گروه فعلی از همپایگی دو گروه فعلی "از رشد ۲۱ درصدی جابه‌جایی مسافر از بنادر مسافری هرمزگان خبرداد" و "افزود در این مدت پنج میلیون و ۷۷۱ هزار و ۳۷۷ مسافر از طریق بنادر مسافری استان جابه‌جا شدند" تشکیل شده است که گروه فعلی دوم دارای وابسته‌ای از نوع بند متممی "در این مدت پنج میلیون و ۷۷۱ هزار و ۳۷۷ مسافر از طریق بنادر مسافری استان جابه‌جا شدند" برای هستهٔ "افزود" است.

حضور هم‌زمان فعل وجهی و ساخت ناهمپایه در جمله‌وارهٔ اصلی:

مثال: نویسنده باید بنویسد که [خود آن را تجربه کرده‌است].



(شکل ۸) – درخت تجزیهٔ ساخت‌سازه‌ای برای جملهٔ "او آگاه است که مادرش بیمار است".

توضیح: در این مثال فعل وجهی "باید" و جمله‌وارهٔ وابستهٔ "خود آن را تجربه کرده است" در درون جمله‌وارهٔ اصلی جای گرفته‌اند و به‌عنوان وابسته‌هایی برای هستهٔ "بنویسد" تلقی می‌شوند.

- ترکیب فاعل تهی و بند متممی فعل:

مثال: این فیلم با "طاووس‌های بی‌پر" تفاوت دارد و نگاه ما در تولید این فیلم به گیشه است و [امیدوارم [فیلمی پرفروش بسازیم]] .

توضیح: در این مثال دو فاعل "من" و "ما" در جملهٔ "امیدوارم فیلمی پرفروش بسازیم" مستتر هستند و آنها را می‌توان با تصریف فعل بازیابی کرد. علاوه‌بر این، جملهٔ "فیلمی پرفروش بسازیم" به‌عنوان بند متممی فعل تلقی می‌شود.

- ترکیب فاعل تهی و بند افزودهٔ فعل:

مثال: عصایم شو تا [جاده، قدم‌هایمان را فال‌گوش بایستد].

توضیح: در این مثال فاعل "تو" در جملهٔ "عصایم شو تا جاده، قدم‌هایمان را فال‌گوش بایستد" مستتر است. همچنین در این مثال جمله "جاده، قدم‌هایمان را فال‌گوش بایستد"، وابسته به جملهٔ اصلی از نوع بند افزودهٔ فعل است.

- ساخت همپایه در درون بند افزودهٔ فعل:

مثال: [اگر [سخنرانی برای شما خیلی مشکل است] و [احساس بی‌قراری آزارتان می‌دهد]]، بهترین راه این است [که شما [یک نوار کاست صوتی را با مطالب مورد نظرتان پرکنید] و [سپس این خلاصهٔ سمینار را برای علاقه‌مندان بفرستید]].

توضیح: جملهٔ "اگر سخنرانی برای شما خیلی مشکل است، بهترین راه این است" را در نظر بگیرید. این جمله از یک ساخت ناهمپایه تشکیل شده است و در آن جملهٔ نخست پیرو و جملهٔ دوم پایه است. جملهٔ پیرو با جملهٔ دیگر یعنی "احساس بی‌قراری آزارتان می‌دهد" همپایه شده است و از طرف دیگر جملهٔ پایه، بند موصولیِ "که شما یک نوار کاست صوتی را با مطالب مورد نظرتان پرکنید و سپس این خلاصهٔ سمینار را برای علاقه‌مندان بفرستید" را گرفته است که یک ساخت همپایه است.

- ترکیب ناهمپایگی و همپایگی در مواردی که عنصر پیوند حذف شده است:

مثال: وی [[به رابطه در سینما اشاره کرد] و [گفت [[رابطه در بین اهالی سینما برتری پیدا کرده است] و

[حمایت از فیلم‌سازان به گروه خاصی اختصاص یافته است]]].

توضیح: این مثال در زیرساخت از دو جملهٔ "وی به رابطه در سینما اشاره کرد" و "وی گفت" تشکیل شده است که جملهٔ دوم یک وابسته به‌صورت یک ساخت همپایه یعنی "(که) رابطه در بین اهالی سینما برتری پیدا کرده است و حمایت از فیلم‌سازان به گروه خاصی اختصاص یافته است"، گرفته است.

همان‌گونه که ملاحظه می‌کنید در مثال‌های بالا ساخت‌های مختلفی با یکدیگر ترکیب شده‌اند. در مورد نحوهٔ پیاده‌سازی این قوانین نیز می‌بایست گفت که آنها با استفاده از شروط "اگر--> آنگاه" عمل می‌کنند؛ بدین معنا که اگر الگوریتم به ترکیب چند رابطهٔ وابستگی خاص (مانند مثال‌های بالا) رسید، آنگاه آن را به روابط ساخت‌سازه‌ای معادل آن تبدیل خواهد کرد. لازم به ذکر است که برای تمام مواردی که حاصل از ترکیب چند رابطهٔ وابستگی خاص هستند، به‌طور مجزا شروط "اگر--> آنگاه" مجزایی طراحی شده است. به‌منظور درک بهتر نحوهٔ کار الگوریتم، جملهٔ زیر را در نظر بگیرید:

"وی به رابطه در سینما اشاره کرد و گفت رابطه در بین اهالی سینما برتری پیدا کرده است و حمایت از فیلم‌سازان به گروه خاصی اختصاص یافته است."

در این مثال سامانه وجود ساخت همپایه را در جملهٔ اصلی تشخیص داده و تلاش می‌کند، با کمک شروط "اگر--> آنگاه"، دو گروه همپایه را به‌طور مستقل ساخته و با فراخوانی تابع همپایگی مورد نظر درنهایت آنها را با یکدیگر همپایه سازد. الگوریتم در هنگام تبدیل عبارت:

"گفت رابطه در بین اهالی سینما برتری پیدا کرده است و حمایت از فیلم‌سازان به گروه خاصی اختصاص یافته‌است"

نیز از شروط طراحی‌شده خاص استفاده کرده و وجود ساخت ناهمپایه و همچنین ساخت همپایه را در این جمله تشخیص داده و رابطهٔ ساخت‌سازه‌ای معادل را خواهد ساخت. الگوریتم در این قسمت نیز با استفاده از شروط "اگر--> آنگاه" عمل کرده، درخت مربوط به دو جملهٔ "رابطه در بین اهالی سینما برتری پیدا کرده است" و "حمایت از فیلم‌سازان به گروه خاصی اختصاص یافته است" را به‌طور مجزا ساخته و با فراخوانی تابع بازگشتی همپایگی جملات این دو جمله را با هم همپایه می‌سازد. بدین ترتیب با کمک شروط ترکیبی و همچنین فراخوانی توابع بازگشتی همپایگی، درخت ساخت‌سازه‌ای کل جمله به‌صورت مرحله‌به‌مرحله تشکیل خواهد شد.

لازم به ذکر است که قوانین تبدیل پیاده‌سازی‌شده در این پژوهش بسیاری از ساخت‌های نحوی رایج در زبان نوشتاری فارسی امروز را تحت پوشش خود قرار می‌دهند و اکثر ساخت‌های پوشش‌داده‌نشده به ساخت‌هایی تعلق دارند که در زبان نوشتاری امروز کمتر به کار می‌روند.

در ادامه به ساخت‌هایی اشاره خواهیم کرد که الگوریتم در حال حاضر آن‌ها را پوشش نمی‌دهد. این ساخت‌ها عبارتند از:

- ساخت‌های دارای منادا:

مثال: یا علی.

توضیح: در این مثال "یا" نقش نمای ندا و کلمهٔ "علی" منادا محسوب می‌شود.

- ساخت‌های دارای بند فعل وصفی[1]:

مثال: احمد به خانه رفته، ناهار می‌خورد.

توضیح: در این مثال فعل وصفی "رفته"، افزودهٔ فعل مرکزی جمله یعنی "می‌خورد" است. در اینجا نوع رابطهٔ وابستگی میان "رفته" و "می‌خورد" بند فعل وصفی است (رسولی و دیگران، ۱۳۹۳).

- ساخت‌های دارای شبه جمله[2]:

مثال: وی نمایش‌نامه‌های خود را چنان می‌پرداخت که گویی نوعی اپرای بدون موسیقی تصنیف می‌کند.

توضیح: در این مثال کلمهٔ "گویی" شبه‌جمله تلقی می‌شود. درحقیقت شبه‌جمله، کلمه یا مجموعه‌ای از کلمات است که بدون داشتن فعل معنای کاملی را بیان می‌کند (انوری و احمدی گیوی، ۱۳۸۵).

- ساخت‌هایی که فرآیند حذف فعل در آن‌ها صورت گرفته است:

مثال: داوود روزنامه می‌خواند و من کتاب (می‌خوانم).

توضیح: در این مثال فعل "می‌خوانم" به قرینه حذف شده‌است.

- ساخت‌هایی که گروه صفتی به خارج از گروه اسمی انتقال داده شده است:

مثال: وی پزشکی است بسیار حاذق.

توضیح: در این مثال گروه صفتی "بسیار حاذق" که وابستهٔ اسم "پزشک" است، به خارج از گروه اسمی

[1] Participle Clause
[2] Interjection

منتقل شده است. درحقیقت صورت اصلی جمله به این شکل است: "وی پزشکی بسیار حاذق است".

- ساخت‌های مبتداسازی‌شده[1]:

مثال: به کارمندان، دولت پاداش خواهد داد.

توضیح: در این مثال گروه حرف اضافه‌ای "به کارمندان"، برای تأکید در ابتدای جمله قرار گرفته است که در اصطلاح به این فرآیند مبتداسازی[2] گویند (غلامعلی‌زاده، ۱۳۹۰). درحقیقت صورت اصلی جمله به این شکل است: "دولت به کارمندان پاداش خواهد داد".

شایان ذکر است که برخی از این ساخت‌ها به‌شدت پیچیده هستند و طراحی الگوریتم تبدیل درخت وابستگی به ساخت‌سازه‌ای برای آن‌ها زمان‌بر و دشوار است. همان‌طور که پیش‌تر در بخش مروری بر کارهای پیشین اشاره شد در روش‌های قاعده‌مند به‌دلیل پیچیدگی‌های مختلف زبانی امکان استخراج کامل و جامع قوانین یک زبان وجود ندارد. با توجه به این‌که پژوهش حاضر در زمرهٔ نخستین کارهای انجام‌یافته در این حوزه برای زبان فارسی محسوب می‌شود، از پیاده‌سازی این قوانین صرف نظر شده و به کارهای آینده موکول شده است.

در پژوهش حاضر پس از طراحی الگوریتم تبدیل درخت تجزیهٔ وابستگی به درخت تجزیهٔ ساخت‌سازه‌ای، الگوریتم مشروح بر روی کل پیکرهٔ وابستگی زبان فارسی با ۲۹۹۸۲ جمله اعمال شد و دادگان درخت نحوی مبتنی بر دستور ساخت‌سازه‌ای با ۲۷۷۱۰ جمله به دست آمد. شایان ذکر است که تا پیش از این زبان فارسی فاقد چنین دادگانی با این‌چنین حجمی بوده و این دادگان یکی از مهم‌ترین دستاوردهای پژوهش حاضر به شمار می‌رود که نه‌تنها در پژوهش‌های نظری زبان‌شناسی، بلکه در حوزهٔ پردازش متن و گفتار زبان فارسی نیز بسیار قابل استفاده خواهد بود. به‌عنوان مثال، یکی از مهم‌ترین کاربردهای دادگان درخت نحوی، استخراج مدل‌های نحوی زبان است. همان‌طور که می‌دانیم این مدل‌ها در تجزیهٔ خودکار نحوی جملات زبان به کار می‌آیند. در ادامه به نحوهٔ استفاده از پیکرهٔ حاصل در تجزیه نحوی جملات زبان فارسی خواهیم پرداخت.

[1] Topicalized
[2] Topicalization

۶- تجزیه‌گر احتمالاتی واژگانی[3]

یک تجزیه‌گر زبان طبیعی برنامه‌ای است که ساختار نحوی جملات یک زبان را تولید می‌کند. تجزیه‌گرهای احتمالاتی برای تحلیل جملات ورودی از مجموعه جملات موجود در دادگان درخت نحوی استفاده می‌کنند و سعی بر آن دارند که تحلیلی مشابه را برای جمله ورودی ارائه کند. درحقیقت یک تجزیه‌گر احتمالاتی یک سامانه مبتنی بر یادگیری است که بر روی دادگان درخت نحوی آموزش داده شده و مدل نحوی یک زبان خاص را به دست می‌دهد و سپس از این مدل در تجزیهٔ هر جمله دلخواه استفاده می‌کند. در پژوهش حاضر به‌منظور دست‌یابی به چنین سامانه‌ای برای تجزیهٔ جملات زبان فارسی از تجزیه‌گر استنفورد[4] (کلاین و منینگ، ۲۰۰۳) استفاده شده‌است. این تجزیه‌گر پیش از این برای تجزیه نحوی جملات زبان‌های انگلیسی، فرانسوی، آلمانی، چینی و عربی مورد استفاده قرار گرفته است. لازم به ذکر است که تجزیه‌گر مذکور بر روی دادگان درخت نحوی فارسی Pertreebank نیز آموزش دیده شده است (قیومی، ۱۳۹۱؛ قیومی، ۲۰۱۲b).

در این مطالعه بر آن بودیم که از تجزیه‌گر مذکور برای استخراج مدل نحوی زبان فارسی استفاده کنیم. برای نیل بدین هدف، تجزیه‌گر مذکور بر روی دادگان درخت نحوی حاصل از پژوهش حاضر آموزش داده شده و تجزیه‌گر نحوی احتمالاتی برای زبان فارسی به دست آمده است.

تجزیه‌گر احتمالاتی واژگانی استنفورد درواقع ترکیبی از یک تجزیه‌گر مبتنی بر گرامر مستقل از متن احتمالاتی[5] و یک تجزیه‌گر وابستگی[6] است. هر یک از این دو تجزیه‌گر معیاری جهت یک تجزیه مناسب را ارائه می‌دهند. این تجزیه‌گر با استفاده از الگوریتمی که از هر دو معیار استفاده می‌کند و همچنین با استفاده از الگوریتم جستجوی **A*** در فضای حالت همهٔ تجزیه‌های ممکن، سعی می‌کند بهترین خروجی را ارائه دهد (هوانگ و چایانگ، ۲۰۰۵).

درواقع مدل مورد استفاده در این تجزیه‌گر به‌صورت دوتایی L (T, D) است که T مدل گرامر مستقل از متن احتمالاتی و D مدل وابستگی مورد نظر است. با فرض استقلال احتمال وقوع هر دو مدل داریم:

$$P(T,D) = P(T) * P(D) \quad (۱)$$

[3] Lexicalized Probabilistic Parser
[4] The Stanford Parser
[5] Probabilistic Context-Free Grammar (PCFG)
[6] Dependency Parser

که P (T) از حاصل‌ضرب احتمال هر یک از قواعد مورد استفاده در تجزیه جمله به دست می‌آید و مقدار P (D) نیز از حاصل‌ضرب احتمال حضور هریک از کلمات جمله حاصل می‌شود. به‌منظور درک درخت تجزیهٔ مدل ترکیبی[1] جمله "Factory payrolls fell in September." را در نظر بگیرید. درخت تجزیهٔ وابستگی، درخت تجزیهٔ مبتنی بر گرامر مستقل از متن احتمالاتی و درخت تجزیهٔ مدل ترکیبی به‌ترتیب در شکل‌های (۹)، (۱۰) و (۱۱) نمایش داده شده‌اند (کلاین و منینگ، ۲۰۰۳).

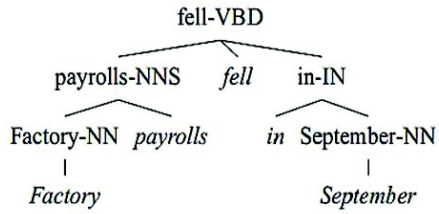
در مورد جملهٔ بالا مقدار P (T) به این شکل محاسبه می‌شود:

P (T) = P (S→NP VP). P (NP→NN NNS). P (VP→ VBD PP). P (PP→IN NN)

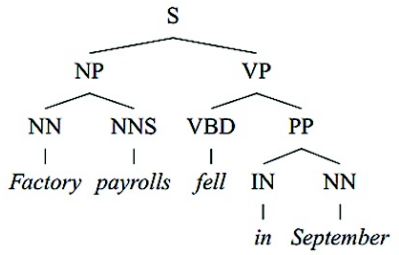
همچنین مقدار P (D) نیز از رابطهٔ زیر به‌دست می‌آید:

P (D) = P (Fell | payrolls, left). P (Factory | payrolls, right). P (START_SYMBOL | Factory, left). P (in | fell, right). P (September | in, left)

که در آن $P(w_i \mid w_j, dir)$ احتمال حضور کلمهٔ w_i در جهت dir (چپ یا راست) کلمهٔ w_j است (فیلی، ۱۳۸۹).

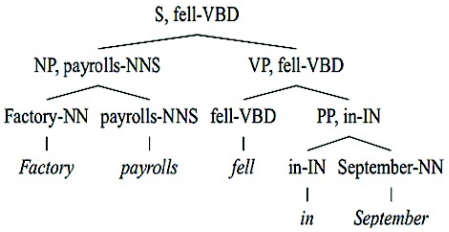


(شکل- ۹): درخت تجزیه وابستگی جمله "Factory payrolls fell in September."



(شکل- ۱۰): درخت تجزیه مبتنی بر گرامر مستقل از متن احتمالاتی جمله "Factory payrolls fell in September."

[1] Factored Model



(شکل- ۱۱): درخت تجزیه ترکیبی جمله "Factory payrolls fell in September."

همان‌طور که پیش‌تر بیان شد، در این پژوهش به‌منظور استخراج مدل نحوی زبان فارسی، تجزیه‌گر مذکور بر روی دادگان درخت نحوی ساخت‌سازه‌ای که شامل ۳۹۲۱۴۹ کلمه و ۲۷۷۱۰ جمله با میانگین طول ۱۴,۱۵ کلمه است، جمله آموزش داده شده و تجزیه‌گر نحوی خاص زبان فارسی به‌دست آمده است. نتایج حاصل از ارزیابی این تجزیه‌گر در بخش بعد شرح داده می‌شود.

۷- نتایج آزمون

همان‌طورکه پیش‌تر اشاره شد تا پیش از پژوهش حاضر دادگان درخت نحوی مبتنی بر دستور ساخت‌سازه‌ای با حجم کافی برای زبان فارسی موجود نبود. به همین سبب برای ارزیابی سامانه پیاده‌سازی‌شده در این پژوهش، مشکل فقدان دادگان استاندارد وجود داشت؛ لذا برای رفع این مشکل یکصد جمله به‌صورت تصادفی از پیکره انتخاب شد و این جملات توسط نیروی انسانی و به‌صورت دستی، تجزیهٔ نحوی و برچسب‌دهی و بدین طریق دادهٔ استاندارد برای ارزیابی تهیه شد. علاوه‌بر این، برای ارزیابی سامانه از تجزیه‌گر استنفورد نیز استفاده شد و با ارزیابی این تجزیه‌گر که حاصل از آموزش بر روی دادگان درخت نحوی استخراج شده است، درحقیقت سامانه طراحی‌شده در این کار مورد آزمون قرار گرفت.

در پژوهش حاضر برای ارزیابی سامانه از معیاری به نام PARSEVAL استفاده شده است که برای مقایسهٔ کارایی تجزیه‌گرهای نحوی به‌کار می‌رود و زیردرخت‌هایی از کل درخت نحوی زایشی را با یکدیگر مقایسه می‌کند (منینگ و شوتزه، ۱۹۹۹). PARSEVAL به منظور ارزیابی سامانه سه معیار را محاسبه می‌کند: ۱- دقت برچسب‌دار[2]، ۲- پوشش برچسب‌دار[3] و ۳- میانگین توافقی دقت و پوشش برچسب‌دار[4] (معیار F1). در ادامه به شرح این سه معیار می‌پردازیم:

[2] Labeled Precision (LP)
[3] Labeled Recall (LR)
[4] F-measure

دقت برچسب‌دار به تعداد قلاب‌های صحیح موجود در درخت تجزیه‌شدهٔ خروجی نسبت به کل قلاب‌های درخت خروجی اشاره دارد، پوشش برچسب‌دار بیانگر تعداد قلاب‌های صحیح موجود در درخت تجزیه‌شدهٔ خروجی نسبت به کل قلاب‌های موجود در درخت استاندارد است. معیار سوم در ارزیابی سامانه نیز میانگین توافقی دقت و پوشش برچسب‌دار (معیار F1) است که می‌توان آن را توسط فرمول زیر بیان کرد:

$$F1 = \frac{2*(LP)*(LR)}{LP+LR} \qquad (۲)$$

در این فرمول LP و LR به ترتیب دقت و پوشش برچسب‌دار می‌باشند.

با توجه به معیارهای بالا، به‌منظور ارزیابی سامانه الگوریتمی پیاده‌سازی شد که جملات تجزیه‌شدهٔ دادگان درخت نحوی ساخت‌سازه‌ای حاصل و جملات تجزیه‌شدهٔ استاندارد را یک به یک از هر سطر خوانده و آن‌ها را با یکدیگر مقایسه می‌کند و سپس نتایج حاصل از ارزیابی سامانه را گزارش می‌دهد.

درحقیقت، ارزیابی سامانه در دو بخش انجام گرفت. در بخش نخست سنجش کارایی سیستم بر روی ۱۰۰ جمله با میانگین طول ۱۵/۸۳ کلمه انجام گرفت. در این بخش تأثیر استفاده از قوانین خاص زبان فارسی در تبدیل درخت تجزیهٔ وابستگی به درخت تجزیهٔ ساخت‌سازه‌ای بررسی شد. بدین منظور ابتدا کارایی الگوریتم پایه در تبدیل درخت وابستگی به ساخت‌سازه‌ای یکصد جمله سنجیده شد؛ سپس کارایی الگوریتم با استفاده از کل قوانین تبدیل و همچنین بخشی از قوانین تبدیل درخت وابستگی به ساخت‌سازه‌ای (کل قوانین تبدیل به جز قوانین تبدیل جملات دارای بند متممی صفت، بند موصولی و متممی اسم، بند افزودهٔ فعل، قید جمله و عناصر جابه‌جاشده) بر روی همان یکصد جملهٔ مورد آزمون قرار گرفت و در انتها نتایج حاصل از این سه حالت با یکدیگر مقایسه شد. درحقیقت هدف از این مقایسه، بررسی تأثیر استفاده از قوانین زبان فارسی در کارآمدی الگوریتم طراحی شده بود. در جدول ۳ نتایج حاصل از ارزیابی بخش نخست گزارش شده است.

(جدول- ۳): نتایج حاصل از ارزیابی تبدیل ۱۰۰ جمله

	الگوریتم پایه	بخشی از قوانین	کل قوانین
دقت برچسب‌دار	۳۸/۳۶	۸۵/۸۸	**۹۵/۸۶**
پوشش برچسب‌دار	۳۴/۴۹	۸۵/۳۴	**۹۶/۲۴**
F1	۳۶/۳۲	۸۵/۶۱	**۹۶/۰۵**

با توجه به آنچه در جدول (۳) آمده است، F1 در پیاده‌سازی الگوریتم پایه برابر با ۳۶/۳۲ درصد، الگوریتم پایه به همراه کل قوانین تبدیل برابر با ۹۶/۰۵ درصد و در حالت حذف برخی قوانین تبدیل (قوانین تبدیل مناسب برای ساختارهای پیچیده) معادل با ۸۵/۶۱ درصد است. همان‌گونه که ملاحظه می‌کنید، الگوریتم پایه در موارد محدودی کارایی دارد. این الگوریتم برای هر کلمه، گروه نحوی متناظر با آن را ساخته و آن را به گروه نحوی متناظر با هستهٔ خود متصل می‌کند. این الگوریتم تنها قادر به تبدیل روابط نشان‌داده‌شده در جدول شمارهٔ ۲ است و از بازنمایی درخت تجزیهٔ نحوی جملاتی دارای پیچیدگی مختلف زبانی ناتوان است و به‌عنوان پایه‌ای برای الگوریتم کلی می‌تواند تلقی شود. با توجه به مقایسهٔ نتایج به‌دست‌آمده از ارزیابی الگوریتم در سه حالت ذکر شده می‌توان نتیجه گرفت که مطالعهٔ دستور زبان فارسی و طراحی و پیاده‌سازی قوانین خاص زبان فارسی در تبدیل روابط وابستگی به ساخت‌سازه‌ای، منجر شده است که کارایی سامانه به‌طور قابل ملاحظه‌ای افزایش یابد. با مقایسهٔ نتایج به‌دست آمده در حالت افزودن بخشی از قوانین و همچنین کل قوانین به سامانه، می‌توان دریافت که بررسی ساختارهای گوناگون و لحاظ‌کردن قوانین تبدیل ساختارهای پیچیده در الگوریتم (همچون ساختارهایی دارای عناصر جابه‌جاشده، بند افزودهٔ فعل، بند متممی صفت و اسم و همین‌طور قید جمله) به افزایش ۱۰/۴۴ درصدی کارایی سامانه منجر شده است.

در بخش دوم ارزیابی نیز تجزیه‌گر نحوی استنفورد بر روی کل پیکره حاصل از این پژوهش با ۲۷۷۱۰ جمله آموزش داده شد و درنهایت مورد ارزیابی قرار گرفت. درحقیقت با ارزیابی این تجزیه‌گر، الگوریتم ارایه‌شده مورد آزمون قرار گرفت؛ چون تجزیه‌گر مذکور با استفاده از دادگان حاصل از خروجی الگوریتم آموزش داده شده است. به‌یقین هرچه دقت تجزیه‌گر آماری آموزش داده شده بیشتر باشد، کیفیت داده آموزش نیز بالاتر خواهد بود. برای این منظور از روش ارزیابی متقاطع ده‌تایی[1] استفاده شد و دادگان درخت نحوی ساخت‌سازه‌ای حاصل به ده بخش تقسیم شد؛ نه بخش از دادگان به دادگان آموزش و یک بخش به دادگان آزمایش اختصاص یافت. نتایج حاصل از ارزیابی تجزیه‌گر نحوی مبتنی بر گرامر وابستگی، گرامر مستقل از متن احتمالاتی و مدل ترکیبی، براساس معیار F1 در جدول (۴) نمایش داده شده‌است. لازم به ذکر است که این آزمایش در

[1] 10-Fold Cross-Validation

سه حالت لحاظ‌کردن جملاتی با حداکثر طول بیست کلمه، چهل کلمه و همچنین لحاظ‌کردن کل جملات و بدون هیچ‌گونه محدودیت طول جمله انجام پذیرفته است. با توجه به این نتایج و همچنین نتایج حاصل از پژوهش‌های پیشین (صادق‌زاده و دیگران، ۱۳۹۳)، می‌توان دریافت که کاهش طول جملات منجر به افزایش کارایی تجزیه‌گر نحوی خواهد شد.

شایان ذکر است که در هنگام آموزش تجزیه‌گر نحوی استنفورد بر روی دادگان، برچسب‌زن مقولهٔ نحوی کلمه[1] نیز آموزش داده می‌شود. درواقع ابتدا برچسب‌زن مقولهٔ نحوی کلمه استنفورد، برچسب مقولهٔ نحوی کلمه را تعیین می‌کند و سپس تجزیه‌گر نحوی استنفورد عملیات تجزیه نحوی جمله را به انجام می‌رساند. متأسفانه برچسب‌زن مقولهٔ نحوی کلمه استنفورد، دارای مشکلات جدّی است و همین امر تجزیهٔ نحوی جمله را به‌شدّت تحت تأثیر قرار می‌دهد (فیلی، ۱۳۸۹). با توجه به این امر و به‌منظور ارزیابی دقیق‌تر، برچسب مقولهٔ نحوی کلمات موجود در جمله به‌عنوان ورودی تجزیه‌گر نحوی استنفورد داده شد و سپس تجزیه‌گر نحوی ترکیبی آموزش‌داده‌شده بر روی دادگان مورد آزمون قرار گرفت. بدیهی است که استفاده از این برچسب‌ها موجب کاهش ابهام در تجزیه شده و به بهبود عملکرد تجزیه‌گر کمک خواهد کرد (صادق‌زاده و دیگران، ۱۳۹۳). در جدول (۵) نتایج ارزیابی مدل ترکیبی در دو حالت ورودی با برچسب مقولهٔ نحوی کلمات و بدون آنها نمایش داده شده است. همان‌طورکه ملاحظه می‌کنید، تجزیه‌گر نحوی در حالت ورودی با برچسب مقولهٔ نحوی کلمات به F1 معادل با ۸۶/۰۱ درصد دست یافته‌است که نسبت به حالت قبل حدود ۱۱ درصد افزایش داشته است.

(جدول ۴)- نتایج حاصل از ارزیابی تجزیه‌گر نحوی استنفورد

حداکثر طول جملات	گرامر وابستگی	گرامر مستقل از متن احتمالاتی	مدل ترکیبی
۲۰	۸۴/۱۸	۶۶/۷۵	۷۷/۵۱
۴۰	۸۲/۷	۶۳/۶۱	۷۵/۸
بدون محدودیت	۸۲/۴	۶۳/۵۸	۷۵/۰۳

[1] Parts of Speech Tagger

(جدول- ۵): نتایج مقایسه‌ای حاصل از ارزیابی تجزیه‌گر نحوی مدل ترکیبی با برچسب مقولهٔ نحوی کلمات و بدون آن

	بدون برچسب مقولهٔ نحوی کلمات	با برچسب مقولهٔ نحوی کلمات
دقت برچسب‌دار	۷۵/۳۴	۸۶/۴۷
پوشش برچسب‌دار	۷۴/۷۳	۸۵/۵۵
F1	۷۵/۰۳	۸۶/۰۱

۸- نتیجه‌گیری و کارهای آتی

در این مقاله سعی بر آن بود الگوریتمی قاعده‌مند برای تبدیل درخت تجزیهٔ نحوی وابستگی به درخت تجزیه نحوی ساخت‌سازه‌ای ارائه شود. به‌منظور اینکه الگوریتم ارائه‌شده در مقاله حاضر دامنه وسیعی از انواع جملات زبان فارسی را تحت پوشش قرار دهد، ساختارهای نحوی پیچیدهٔ زبان فارسی بررسی و قوانین خاص تبدیل روابط وابستگی به روابط ساخت‌سازه‌ای برای زبان فارسی استخراج شد؛ سپس قوانین تبدیل حاصل به الگوریتم مذکور اضافه شد و بدین ترتیب درخت نحوی ساخت‌سازه‌ای معادل با درخت وابستگی آغازین به‌دست آمد. در مرحله بعد الگوریتم طراحی‌شده بر روی دادگان وابستگی زبان فارسی با حدود سی‌هزار جمله اعمال و برای نخستین‌بار دادگان درخت نحوی مبتنی بر دستور ساخت‌سازه‌ای با حجمی در حدود ۲۸۰۰۰ جمله برای زبان فارسی حاصل شد. با توجه به اینکه دقت و پوشش برچسب‌دار روش پیشنهادی بالاست (به‌ترتیب ۹۵,۸۶ و ۹۶,۲۴ درصد) و همچنین در نظر داشتن این نکته که ورودی الگوریتم پیکرهٔ وابستگی زبان فارسی است که نمونه‌ای جامع از متون فارسی امروز به شمار می‌رود، می‌توان دریافت که الگوریتم طراحی‌شده بسیاری از ساخت‌های رایج در متون نوشتاری فارسی امروز را تحت پوشش قرار می‌دهد.

درنهایت تجزیه‌گر احتمالاتی استنفورد بر روی دادگان حاصل از پژوهش حاضر آموزش داده شد و تجزیه‌گر نحوی خاص زبان فارسی به دست آمد.

به‌عنوان کارهای آینده می‌توان الگوریتم معرفی‌شده در پژوهش حاضر را برای ساخت‌های پوشش داده نشده نیز بسط داد؛ چون الگوریتم تبدیل درخت تجزیهٔ وابستگی به درخت تجزیهٔ ساخت‌سازه‌ای، در تبدیل ۲۲۷۲ جمله (۷,۵۷ درصد از کل پیکرهٔ وابستگی) با میانگین طول ۴۹,۳۴ کلمه ناموفق بوده‌است و سیستم طراحی‌شده قادر به

حصول درخت تجزیهٔ ساخت‌سازه‌ای برای کل جمله نبوده‌است و در برخی موارد تنها بخشی از جمله به‌درستی تجزیه‌شده است. برخی از این جملات به‌شدت پیچیده و طولانی هستند؛ چون حاصل از عناصر محذوف و جابه‌جاشده بسیار در درون چندین جملهٔ تودرتو می‌باشند. بدیهی است که طراحی سامانه‌ای برای تبدیل جمله‌ای برای مثال با ۱۹۳ کلمه به‌شدت دشوار است و از لحاظ صرف وقت، هزینه و پیچیدگی سامانه به‌صرفه نیست؛ ولی برای برخی از ساخت‌ها مانند ساخت‌های دارای منادا، بند وصفی فعل و دیگر ساخت‌های نه‌چندان پیچیده الگوریتم حاضر را می‌توان بسط داد و به‌کارایی سامانه کنونی افزود و سپس الگوریتم حاصل را بار دیگر بر روی کل پیکره وابستگی اعمال کرد.

۹- مراجع

انوری، حسن و احمدی گیوی، حسن، "دستور زبان فارسی ۲". تهران، انتشارات فاطمی، ۱۳۸۵.

دبیرخانه شورای عالی اطلاع رسانی،"فاز اول طرح جامع پیکره زبان فارسی با موضوع فاز اول مطالعاتی ایجاد پیکره متنی زبان فارسی، بررسی ابعاد و تفاوت‌های پیکره‌های برچسب داده‌ای و پیکره‌های خام در زبان فارسی"، کد زیر پروژه: پیک متن فارسی-۲- ح، ۱۳۸۸، ویراست اول.

رسولی، محمد صادق؛ کوهستانی، منوچهر؛ مولودی، امیرسعید؛ "پیکرهٔ نحوی زبان فارسی – پژوهشی بر اساس دستور وابستگی"، تهران، ایران: دبیرخانهٔ شورای عالی اطلاع‌رسانی، ۱۳۹۳.

صادق‌زاده، محمد باقر؛ رزازی، محمدرضا؛ قیومی، مسعود؛ "بررسی روش‌های موثر بر عملکرد تجزیه‌گر آماری زبان فارسی"، سومین همایش ملی زبان‌شناسی رایانشی، تهران، دانشگاه شریف، ۱۳۹۳.

طبیب‌زاده، امید، "دستور زبان فارسی براساس نظریه گروه‌های خودگردان در دستور وابستگی". چاپ اول، تهران، ایران: نشر مرکز، ۱۳۹۰.

طبیب‌زاده، امید، "وابسته‌های فعل در زبان فارسی براساس نظریه وابستگی"، دستور، ویژه نامه فرهنگستان، ۱۳۸۳، شماره ۱، صص ۱۳-۲۸.

غلامعلی‌زاده، خسرو، "ساخت زبان فارسی". تهران، انتشارات احیا کتاب، چاپ ششم، ۱۳۹۰.

فیلی، هشام، "استفاده از تجزیه‌گرهای احتمالاتی زبان طبیعی جهت بهبود ترجمه افعال گروهی انگلیسی به فارسی"، پردازش علایم و داده‌ها، ۱۳۸۹، شماره ۱ (پیاپی ۱۳)، صص ۶۵-۷۶.

قریب، عبدالعظیم، "دستور زبان فارسی". تهران، ایران: انتشارات اشراقی، ۱۳۶۸.

قیومی، مسعود، "معرفی دادگان درختی و تجزیه‌گر خودکار فارسی"، ارائه شده در هشتمین همایش زبان‌شناسی ایران، تهران، دانشگاه علامه‌طباطبایی، ۱۳۹۱.

مشکوه الدینی، مهدی، "دستور زبان فارسی برپایه نظریه گشتاری". چاپ دوم، مشهد، انتشارات دانشگاه فردوسی مشهد، ۱۳۸۹.

Alabbas, M. and Ramsay, A., “Arabic Treebank: from Phrase-Structure Trees to Dependency Trees”, in META-RESEARCH Workshop on Advanced Treebanking, Language Resources and Evaluation Conference, Istanbul, Tutkey, 2012, pp. 61-68.

Aldezabal, I., Aranzabe, M. J., Diaz de Ilarraza, A., and Fernández, K., “From Dependencies to Constituents in the Reference Corpus for the processing of Basque”, in SEPLN, Madril, 2008, pp. 147-154.

Atkins, S., Clear, J. H., and Ostler, N., “Corpus design criteria”, Journal of Literary and Linguistic Computing, Gordon Dixon. Oxford: Oxford University Press, 1992, vol.7, pp. 1-16.

Bhatt R., and Xia, F., “Challenges in Converting between Treebanks: a Case Study from the HUTB”, in Proceedings of META-RESEARCH Workshop on Advanced Treebanking Held in conjunction with LREC-2012, Istanbul, Turkey, 2012, pp. 53-60.

Böhmova, A., Hajic, J., Hajicova, E., and Hladka, B., “The prague dependency treebank: Three-level annotation scenario”, in Treebanks: Building and using syntactically annotated corpora, Dordrencht: Kluwer Academic Publishers, 2003, pp. 103-127.

Bohnet, B., “Mapping Phrase Structures to Dependency structures in the Case of Free Word Order Languages”, in Proceedings of The First International Conference on Meaning-Text Theory, Paris, , 2003, pp. 239-249.

Buchholz, S. and Marsi, E., “CoNLL-X shared task on multilingual dependency parsing”. In Proceeding of the Tenth Conforence on Computational Natural Language Learning (CoNLL), 2006, pp. 149-164.

Collins, M., “Three generative, lexicalised models for statistical parsing”, in Proceedings of the 35th ACL, 1997, pp.16-23.

Daum, M., Foth, K.A., Menzel, W., “Automatic Transformation of Phrase Treebanks to Dependency

Manning, C. and Schütze, H., "Foundations of Statistical Natural Language Processing", Cambridge: The MIT Press, 1999.

Nivre, J. and McDonald, R., "Integrating graph-based and transition-based dependency parsers", In Proceedings of ACL-08: HLT, 2008, pp. 950–958.

Palmer, M., Bhatt, R., Narasimhan, B., Rambow, O., Sharma, D. M., Xia, F., "Hindi Syntax: Annotating Dependency, Lexical Predicate-Argument Structure, and Phrase Structure", in Proceeding of ICON, Heydarabad, 2009.

Rasooli, M. S., Kouhestani, M., and Moloodi, A., "Development of a Persian Syntactic Dependency Treebank", in The 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL HLT), Atlanta, USA, 2013, pp. 306-314.

Ren, X., Chen, X., Kit, C., "Combine constituent and dependency parsing via reranking", in the Proceedings of the 23th Joint Conference on Artificial Intelligence, 2013, pp. 2155-2161.

Seraji M., Beáta M., and Joakim N., "Bootstrapping a Persian Dependency Treebank". Linguistic Issues in Language Technology 7(18), 2012, pp. 1-10.

Schneid, G., "A Linguistic Comparison Constituency, Dependency and Link Grammar", M. A. Thesis, University of Zurich ,1998.

Taylor, A., Marcus, M. and Santorini, B., "The Penn treebank: an overview". Treebanks: Building and using parsed corpora, Kluwer Academic Publishers, The Netherlands, 2003, pp. 5-22.

Wang, Z., and Zong, C., "Phrae Structure Parsing with Dependency Structure", in COLING 2010: posters, Beijing, China, 2010, pp. 1292-1300,

Xia, F., and Palmer, M., "Converting Dependency Structures to Phrase Structures", in the Proceedings of the 1st Human Language Technology Conference (HLT-2001), San Diego, CA, 2001.

Xia, F., Bhatt, R., Rambow, O., Palmer, M. and Sharma, D., "Towards a multi-representational Treebank", in the Proceedings of The 7th International Workshop on Treebanks and Linguistic Theories, TLT-7, Groningen, the Netherlands, 2009, pp. 159-170.

Žabortsky, Z. and Kučerová, I., "Transforming Penn Treebank Phrase Trees into (Praguian) Tectogrammatical Dependency Trees", The Prague Bulletin of Mathematical Linguistics, 2002, pp. 77-94.

Trees", in Proceedings LREC-04, Lisbon, Portugal, 2004, pp. 1149-1152.

Farkas, R. and Bohnet, B., "Stacking of Dependency and Phrase Structure Parsers", In Proceedings of COLING 2012: Technical Papers, COLING 2012, Mumbai, 2012, pp. 849–866.

Gaifman, H., "Dependency Systems and Phrase-Structure Systems", Information and Control, 1965, pp. 304–337.

Gelbukh, A., Torres, S., Calvo, H., "Transforming a constituency treebank into a dependency treebank", in Procesamiento de Lenguaje Natural, No. 34, Spain, 2005.

Ghayoomi, M., "Bootstrapping the development of an HPSG-based treebank for Persian". Linguistics Issues in Language Technology, CSLI Publications, 2012a, 7 (1).

Ghayoomi, M., "Word clustering for Persian statistical parsing", In Proceedings of 8th International Conference On NLP, Kanazawa, Japan, pp. 126-137, 2012b.

Ghayoomi, M. and Kuhn, J., "Converting an HPSG-based treebank into its parallel dependency-based treebank", In Proceedings of the 9th International Conference on Language Resources and Evaluation. May 26-31, 2014, Reykjavik, Iceland, pp. 802-809.

Huang, L. and Chiang, D. "Better k-best parsing", In Proceedings of the 9th International Workshop on Parsing Technologies, Vancouver, BC, 2005, pp. 53–64.

Johansson, R. and Nugues, P., "Extended constituent-to-dependency conversion for English", in the Proceedings of the 16th Nordic Conference of Computational Linguistics NODALIDA, 2007, pp.105-112.

Klein, A., "From dependency to constituency: Automatic generation of Penn Treebank trees from LFG f-structures", M.S. Thesis , University of Heidelberg, Germany, 2009.

Klein, D., and Manning. C. D., "Fast Exact Inference with a Factored Model for Natural Language Parsing", Advances in Neural Information Processing Systems 15 (NIPS 2002), Cambridge, MA: MIT Press, 2003, pp. 3-10.

Le Roux, J., Favre, B., Nasr, A., and Mirroshandel, S. A., "Generative Constituent Parsing and Discriminative Dependency Reranking: Experiments on English and French", In ACL Joint Workshop on Statistical Parsing and Semantic Processing of Morphologically Rich Languages, Jeju, South Korea, 2012.

Li, X., Wang, X., and Yao, L., "Conversion between Dependency Structures and Phrase Structures Using a Head Finder Algorithm", in International Conference on Natural Language Processing and Knowledge Engineering, Beijing , 2010, pp. 1-6.

فاطمه سلطان‌زاده، مدرک کارشناسی خود را در رشته مهندسی کامپیوتر، گرایش نرم‌افزار از دانشگاه شهید چمران اهواز دریافت کرد. وی تحصیلات خود را در مقطع کارشناسی ارشد در رشته زبان‌شناسی رایانشی ادامه داد و با کسب رتبهٔ نخست از دانشگاه صنعتی شریف فارغ‌التحصیل شد. زمینه‌های تخصصی مورد علاقهٔ ایشان پردازش زبان طبیعی، نحو و صرف زبان و زبان‌شناسی شناختی است.

نشانی رایانامهٔ ایشان عبارت است از:

fatemeh.slt@gmail.com

محمد بحرانی، درجهٔ دکترای خود را در رشتهٔ هوش مصنوعی در سال ۱۳۸۹ از دانشگاه صنعتی شریف و کارشناسی ارشد خود را از همان دانشگاه در سال ۱۳۸۲ دریافت کرده است. وی هم‌اکنون عضو هیئت علمی گروه زبان‌شناسی رایانشی در مرکز زبان‌ها و زبان‌شناسی دانشگاه صنعتی شریف است. زمینه‌های پژوهشی مورد علاقه ایشان عبارت است از: زبان‌شناسی رایانشی، پردازش زبان طبیعی، پردازش و بازشناسی گفتار و مدل‌سازی زبانی.

نشانی رایانامهٔ ایشان عبارت است از:

bahrani@sharif.edu

محرم اسلامی، در سال ۱۳۷۰ در رشتهٔ مترجمی زبان انگلیسی از دانشگاه علامه طباطبایی فارغ‌التحصیل شد؛ سپس در سال ۱۳۷۳ در رشتهٔ زبان‌شناسی همگانی دورهٔ کارشناسی ارشد خود را در دانشگاه تهران به پایان رساند و در سال ۱۳۷۹ در همان رشته و همان دانشگاه موفق به اخذ مدرک دکترا شد. از ایشان بیش از سی مقاله در نشریات و همایش‌های داخلی و خارجی به چاپ رسیده است و همچنین صاحب تألیفاتی هستند. وی هم‌اکنون عضو هیئت علمی دانشگاه زنجان است. زمینه‌های تخصصی ایشان زبان‌شناسی همگانی، واج‌شناسی، آهنگ و آواشناسی و پردازش زبان طبیعی است.

نشانی رایانامهٔ ایشان عبارت است از:

meslami@znu.ac.ir

پیوست

(جدول- ۱ پیوست): روابط وابستگی موجود در پیکرهٔ وابستگی زبان فارسی (رسولی و دیگران، ۱۳۹۳)

توضیح	اختصار	توضیح	اختصار
متمم بندی صفت	ACL	صفت پیشین اسم	NPREMOD
قید	ADV	جزء اسمی	NPRT
متمم قیدی فعل	ADVC	فعل یار	NVE
متمم حرف اضافه‌ای صفت	AJPP	مفعول	OBJ
بند افزودهٔ فعل	AJUCL	مفعول دوم	OBJ2
وابستهٔ پسین صفت	APOSTMOD	بند وصفی	PARCL
بدل	APP	افزوده پرسشی فعل	PART
وابستهٔ پیشین صفت	APREMOD	همپایهٔ حرف اضافه	PCONJ
همپایهٔ قید	AVCONJ	وابستهٔ پسین	POSDEP
حرف اضافهٔ تفضیلی	COMPPP	صفت پیشین اسم	NPREMOD
فعل یار پی بستی	ENC	جزء اسمی	NPRT
جزء همکرد	LVP	گزاره	PRD
ممیز	MESU	وابستهٔ پیشین	PREDEP
مسند	MOS	مستمرساز	PROG
مضافٌ الیه	MOZ	علامت نگارشی	PUNC
قید اسم	NADV	ریشهٔ جمله	ROOT
بند اسم	NCL	فاعل	SBJ
همپایهٔ اسم	NCONJ	تمیز	TAM
اسم یار	NE	بند متممی فعل	VCL
متمم نشانهٔ اضافه‌ای صفت	NEZ	همپایهٔ اسم	NCONJ
صفت پسین اسم	NPOSTMOD	مفعول حرف اضافه‌ای	VPP
حرف اضافهٔ اسم	NPP	حرف اضافهٔ فعلی	VPRT