

ساخت اضافه در زبان فارسی: بررسی پیکره‌بنیاد

مینو نساجیان

دانش آموخته کارشناسی ارشد زبان‌شناسی رایانشی، مرکز زبان‌ها و زبان‌شناسی، دانشگاه صنعتی شریف

راضیه شجاعی

دانش آموخته دکتری زبان‌شناسی، دانشگاه تهران

محمد بحرانی^۱

استادیار گروه علوم رایانه، دانشگاه علامه طباطبائی

تاریخ دریافت مقاله: ۹۷/۱۰/۱۵؛ تاریخ پذیرش مقاله: ۹۸/۳/۱

چکیده

ساخت اضافه همواره در نظریه‌های مختلف زبان‌شناسی نظیر آوایی، ساختواژی و نحوی حائز اهمیت بوده است و زبان‌شناسان ایرانی تاکنون تحلیل‌های متفاوتی از این ساخت به دست داده‌اند. عدم تظاهر کسره اضافه در نوشتار، ابهامات بسیاری را در تحلیل و درک متون فارسی موجب شده است و برنامه‌های مختلف پردازش زبان اعم از برچسب‌زن اجزای کلام، تشخیص موجودیت‌های نام‌مند، تشخیص کلمات هم‌مرجع، تبدیل متن به گفتار، ترجمه ماشینی، تجزیه نحوی جملات و غیره را با چالش‌های بسیاری روبرو ساخته است. به همین روی، شناسایی جایگاه این عنصر از مهم‌ترین چالش‌های پردازش متون زبان فارسی به‌شمار می‌رود. پژوهش حاضر می‌کوشد تا به شیوه‌ای تحلیلی و پیکره‌بنیاد و از منظر دستور وابستگی به بررسی ساخت اضافه بپردازد. از آنجا که دستور وابستگی به لحاظ سادگی، استفاده کم از فضای حافظه رایانه و تسریع در امر پردازش در مطالعات پردازش متن در حوزه زبان‌شناسی رایانشی از اهمیت چشمگیری برخوردار است، بهترین پایگاه نظری را برای این دست مطالعات فراهم می‌آورد. به همین سبب، پژوهش حاضر در تلاش است تا با استفاده از این دستور روشی قاعده‌مند جهت تشخیص کلمات حاوی نشانه کسره اضافه در متون فارسی ارائه دهد. بدین منظور، با ارائه کلیه ساخت‌های نمونه‌ای که حاوی نشانه اضافه هستند و از پیکره وابستگی زبان فارسی دانشگاه اویسالا استخراج شده‌اند، در چارچوب دستور وابستگی به تجزیه و تحلیل آن‌ها خواهیم پرداخت. از رهگذر این بررسی، تنها هفت قاعده منطقی برای درج کسره اضافه در گروه‌های غیرفعلی اعم از گروه‌های اسمی، صفتی، حرف اضافه‌ای، قیدی، گروه‌هایی با بیش از یک وابسته پسین، گروه‌هایی با بیش از یک وابسته از نوع گروه و ساخت‌های همپایه استخراج می‌شود که با استناد به آن‌ها می‌توان در انواع پیکره‌های وابستگی و سامانه‌های رایانه‌ای مبتنی بر تجزیه وابستگی به شناسایی جایگاه نشانه اضافه پرداخت. افزون بر این، در این پژوهش به جایگاه‌های نشانه اضافه‌ای نیز اشاره خواهد شد که تاکنون در پژوهش‌های نظری و رایانه‌ای پیشین به آن‌ها پرداخته نشده است.

واژه‌های کلیدی: نشانه اضافه، ساخت اضافه، دستور وابستگی، قواعد درج اضافه، پردازش متن فارسی.

۱- مقدمه

واکه‌های کوتاه فارسی معمولاً در نوشتار تظاهر نمی‌یابند. همین امر ابهامات فراوانی را در درک متون موجب می‌شود. از پیامدهای عدم استفاده از واکه‌های کوتاه در زبان فارسی، مشکل در تشخیص مرز سازه‌ها و کسره اضافه است. بی‌جن‌خان (۱۳۸۲) تکواژ اضافه را یک صورت‌ساز^۱ تعریف می‌کند که کلمات زبان را به یکدیگر پیوند می‌دهد تا از این رهگذر ساخت‌های بزرگ‌تر از کلمه ساخته شود. به گفته وی از نظر صرفی، کسره اضافه پی‌چسبی که جایگاه ساختاری آن در انتهای کلمه در جایگاه فاقد تکیه است (بی‌جن‌خان: ۱۳۸۷: ۳۲-۳۳). تشخیص جایگاه اضافه در متون نوشتاری و درج آن در جایگاه درست، علاوه بر این که درک متون را آسان‌تر می‌کند، موجب بالا رفتن دقت در پردازش‌های زبان طبیعی چون تحلیل نحوی جملات، ترجمه ماشینی، برچسب‌زنی اجزای کلام، شناسایی موجودیت‌های نام‌مند^۲، استخراج معنی، خلاصه‌سازی متون، تشخیص مرز گروه‌ها و جملات و غیره می‌شود.

کسره اضافه، به‌عنوان چالشی مهم در پردازش زبان فارسی از دیرباز مورد توجه پژوهشگران مختلف (چون قیومی و ممتازی (۲۰۰۹)، شمس‌فرد (۲۰۱۱)، ساگت^۳ و والتر^۴ (۲۰۱۰) در میان دیگر پژوهشگران) بوده است. پیچیدگی‌های این عنصر محققان را به مطالعه ساخت اضافه^۵ از جنبه‌های گوناگون سوق داده است؛ ویژگی‌های آوایی، ساختواژی و نحوی این ساخت همواره مدنظر بوده است و زبان‌شناسان از منظر رویکردهای مختلف بدان پرداخته‌اند. نظر به این که هدف مقاله حاضر بررسی و شناسایی نشانه اضافه و ارائه راهکاری در این زمینه است که در حوزه پردازش متن به کار آید، چارچوب نظری انتخاب‌شده دستور وابستگی^۶ است. دلیل این انتخاب، همسویی بیشتر این دستور با اصل ساده‌سازی نظام دستوری نسبت به رویکردهای رقیب است. در مقایسه با دیگر دستورها، دستور وابستگی در تجزیه نحوی جملات از تعداد گره‌های کمتری استفاده می‌کند؛ تنها گره‌های ساخته‌شده در نمودارهای درختی آن محدود به کلمات حاضر در جمله است. به همین سبب، در حوزه پردازش متن، این دستور در میزان سرعت و حافظه مصرفی رایانه نقش بسزایی دارد (کووینگتون^۷، ۲۰۰۱)؛ بنابراین، می‌توان گفت بهترین پایگاه نظری را برای مطالعات رایانه‌ای فراهم می‌آورد. از این رو، بر پایه این دستور مقاله حاضر بر آن است تا به بررسی ساخت اضافه و جایگاه کسره اضافه

1. Formative. 2. Named Entity Recognition (NER). 3. Sagot. 4. Walter. 5. Ezafe construction. 6. dependency grammar. 7. Covington

پردازد. پیکره وابستگی فارسی دانشگاه اوپسالا^۱، به کوشش سراجی (۲۰۱۵) داده‌های این پژوهش را فراهم می‌آورد. این پیکره حاوی ۶۰۰۰ جمله، ۳۱ برچسب اجزای کلام^۲ و ۴۸ برچسب رابطه وابستگی^۳ است^(۱). از رهگذر بررسی جملات این پیکره و روابط وابستگی میان کلمات، به استخراج قواعد منطقی ساده و روشن برای درج کسره اضافه در گروه‌های اسمی، صفتی، حرف اضافه‌ای، قیدی و همپایه می‌پردازیم. انتظار می‌رود با استناد به این قواعد بتوان در سامانه‌های رایانه‌ای به صورت کارا به شناسایی نشانه اضافه پرداخت. پس از مقدمه حاضر، در بخش دوم که در دو بخش تنظیم شده است، نخست پیشینه نظری پژوهش و سپس مطالعات رایانشی را بررسی می‌کنیم. در بخش سوم، به اجمال به مبانی دستور وابستگی می‌پردازیم و مفاهیمی که در این پژوهش به کار برده‌ایم، معرفی می‌کنیم. در بخش چهارم، قواعدی منطقی برای درج نشانه اضافه پیشنهاد می‌دهیم که در سامانه‌های رایانه‌ای، جایگاه نشانه اضافه را شناسایی می‌کند. سرانجام در بخش پنجم، نتیجه‌گیری کلی از تحلیل حاضر ارائه می‌دهیم.

۲- پیشینه پژوهش

ساخت اضافه و جایگاه کسره اضافه، افزون بر اینکه توجه زبان‌شناسان ایرانی و غیر ایرانی را به خود جلب کرده، در میان متخصصان علوم رایانه نیز همواره از اهمیت بالایی برخوردار بوده است. در ادامه این بخش که در دو زیر بخش تنظیم شده است، نخست به اختصار به پیشینه پژوهش‌های زبان‌شناسی در این حوزه اشاره می‌شود. از برخی از دستاوردهای این مطالعات در پیشبرد پژوهش حاضر استفاده می‌گردد. سپس در زیر بخش دوم به پژوهش‌های رایانه‌ای حول ساخت اضافه پرداخته می‌شود.

۲-۱- مطالعات زبان‌شناسی

ساخت اضافه فارسی دست‌مایه مطالعات زبان‌شناسان مختلف قرار گرفته است. از آن میان، راستارگویوا (۱۳۴۷)، ابوالقاسمی (۱۳۷۵) و خانلری (۱۳۷۷) کسره اضافه را به لحاظ تاریخی بررسی کرده و آن را حاصل تحول /i/ فارسی میانه می‌دانند که آن نیز به نوبه خود حاصل تحول ضمیر موصولی hya- در ایرانی باستان است. تباعیان (۱۹۷۴) اولین زبان‌شناسی است که در چارچوب دستور زایشی، تحلیلی نحوی از ساخت اضافه به دست داده است. ماهوتیان (۱۹۹۳) نیز در قالب دستور زایشی، ساخت اضافه را به‌عنوان فرافکن پیشینه هسته-پایانی معرفی می‌کند که هسته آن کسره اضافه است.

1. Uppsala Persian Corpus (UPC). 2. Part-Of-Speech tag (POS tag). 3. dependency relation. 4. dependency relation

معین‌زاده (۲۰۰۱) بر روی این مطلب که اضافه فرافکن بیشینه است، با ماهوتیان اتفاق نظر دارد، اما سازه‌ها را برخلاف نظر ماهوتیان هسته-ابتدا تلقی کرده، گروه اضافه^۱ را متمم هسته در نظر گرفته و به تحلیل آن در چارچوب برنامه کمینه‌گرا^۲ می‌پردازد. کریمی و بریم^۳ (۱۹۸۶) و سمولین (۲۰۰۵، ۲۰۰۸) از طرفداران جدی رویکرد ساختارآژی به ساختار اضافه هستند و به توضیح جایگاه‌های اضافه پرداخته‌اند. سمیعان (۱۹۸۳، ۱۹۹۴) و هاشمی‌پور (۱۹۸۹) اضافه را نشانه حالت^۴ و درج آن را در سطح نمود آوایی می‌دانند و ساختار درونی مقولات را در چارچوب نظریه ایکس-تیره^۵ و نظریه حالت بررسی می‌کنند. قمشی (۱۹۹۷) نیز اضافه را با استفاده از تحلیل‌های آوایی بررسی می‌کند و کسره اضافه را واکه‌ای می‌داند که در سطح صورت واجی^۶ بر روی عناصری که ویژگی اسمی دارند می‌تواند اضافه شود. لارسن و یاماکی‌دو^۷ (۲۰۰۵) در پرتو بررسی جایگاه زیرساختی توصیف‌گرهای اسمی به کسره اضافه در زبان‌های ایرانی پرداخته و معتقدند اضافه در این زبان‌ها کنترل‌کننده حالت است که از طریق حرف تعریف موجود در گروه اسمی به موضوع‌هایش اعطا می‌شود.

کهنمویی‌پور (۲۰۰۶، ۲۰۱۴، ۲۰۱۶) در چارچوب برنامه کمینه‌گرا به ساخت اضافه در زبان پرداخته است. او جایگاه اولیه صفات، توصیف‌گرها و متمم‌های اسم را پیش از اسم و جایگاه روساختی آن‌ها را بعد از اسم دانسته و ظهور کسره اضافه در میان آن‌ها را نتیجه حرکت نحوی اسم می‌داند. لطفی (۲۰۱۴) نیز اضافه را در چارچوب زبان‌شناسی شناختی بررسی می‌کند و آن را به عنوان پیکر^۸ تلقی کرده است. شمار دیگری از پژوهش‌ها نیز در سایر زبان‌های ایرانی به حوزه ساخت اضافه پرداخته‌اند: هلمبرگ^۹ و آدن^{۱۰} (۲۰۰۵) در گویش هورامی، محمدی بلبان آبادی (۱۳۸۷) در گویش سورانی، لارسون^{۱۱} و یاماکی‌دو^{۱۲} (۲۰۰۶) و توسروندی و ون‌یورک^{۱۳} (۲۰۱۴) در زازاکی و یاراحمدزهی (۱۳۸۹) و کریمی (۲۰۰۷) به ترتیب در بلوچی و در گویش سورانی.

۲-۲- مطالعات رایانه‌ای

چنانکه پیشتر گفته شد، عدم استفاده از نشانه اضافه ابهاماتی را در نوشتار موجب شده که تجزیه نحوی و پردازش‌های معنایی جملات را با مشکل مواجه می‌سازد. به این سبب

1. Ezafe phrase. 2. Minimalist Program. 3. Brame. 4. case marker. 5. X-bar theory. 6. Phonological Form (PF). 7. Larson and Yamakido. 8. figure. 9. Holmberg. 10. Odden. 11. Larson. 12. Yamakido. 13. Van Urk

بررسی این پدیده در مطالعات رایانه‌ای نیز راه یافته است. رویکرد اتخاذ شده در این گونه مطالعات عمدتاً مبتنی بر روش‌های قاعده-محور^۱ یا داده-محور^۲ است که در ادامه به پژوهش‌های صورت گرفته با هر یک از روش‌های نامبرده اشاره می‌شود.

۲-۲-۱- مطالعات قاعده-محور

محققانی که از روش‌های قاعده-محور جهت تشخیص جایگاه درج نشانه اضافه استفاده کرده‌اند، از یافته‌های نظری در حوزه ساخت اضافه بسیار بهره گرفته و بر مبنای آن‌ها به استخراج قواعد تشخیص کلمات دارای نشانه اضافه در متون فارسی همت گمارده‌اند. مگردومیان^۳ (۲۰۰۰) در پژوهش خود، بر پایه روش‌های قاعده-محور، به طراحی یک تحلیل‌گر ساختوازی^۴ پرداخته است. وی برای کلمات فارسی ویژگی‌های مختلفی اعم از اضافه تعریف کرده که با این ویژگی و تعریف قواعد نحوی، وجود یا عدم وجود کسره اضافه برای کلمات موجود در جمله مشخص می‌گردد. به‌عنوان پژوهش دیگری که از روش‌های قاعده‌مند بهره گرفته است، می‌توان به قیومی و مولر^۵ (۲۰۱۰) اشاره کرد که با استفاده از دستور ساختار گروهی هسته-بنیاد^۶، به یک سری ویژگی‌ها و پدیده‌های زبان فارسی از جمله نشانه اضافه و نحوه درج آن بر پایه قواعد نظری برای مدل‌سازی دستور زبان فارسی اشاره کرده‌اند.

۲-۲-۲- پژوهش‌های داده-محور

اسکویی‌پور (۲۰۱۱) سامانه‌ای را بر پایه مدل مخفی مارکوف^۷ برای تشخیص جایگاه نشانه اضافه طراحی کرده است. نجومیان (۲۰۱۱) سامانه‌ای را بر پایه دو سطح ساختوازی برای تحلیل خودکار ساختمان واژه و تشخیص واژه‌های کوتاه فارسی با استفاده از ماشین مبدل حالت قطعی^۸ طراحی نموده است که پس از استخراج تک‌به‌تک کلمات از جمله ورودی و نرمال‌سازی آن‌ها، در سطح اول، کلمات به لحاظ تکواژهای سازنده آن و برچسب جزء کلام تجزیه و تحلیل می‌شوند و همزمان، واژه‌های کوتاه برای کلمات درج می‌شوند. در نهایت، اضافه نیز با توجه به توالی برچسب‌های اجزاء کلام به کلمات اضافه می‌گردد. نوفرستی و شمس‌فرد (۲۰۱۴) با استفاده از الگوریتم ژنتیک و

1. rule-based methods. 2. statistical methods. 3. Megerdoomian. 4. morphological analyzer. 5. Müller. 6. Head-driven Phrase Structure Grammar (HPSG). 7. Hidden Markov Model (HMM). 8. Finite-State Transducer (FST).

مدل‌های قاعده‌مند برچسب‌زن به تشخیص جایگاه اضافه پرداخته‌اند و اصغری، مالکی و فیلی (۲۰۱۴) نیز با استفاده از الگوریتم‌های برچسب‌دهی اجزای کلام، آنتروپی بیشینه^۱ و میدان‌های تصادفی شرطی^۲ مدلی احتمالاتی را برای تشخیص جایگاه کسره اضافه طراحی کردند. بر اساس پیشینه حاضر، این پژوهش می‌کوشد تا با بهره‌گیری از دستاوردهای پژوهش‌های پیشین، در چارچوب دستور وابستگی، به شناسایی جایگاه‌های درج کسره اضافه بپردازد. در بخش آتی به اجمال اصول و مبانی این دستور را معرفی می‌کنیم.

۳- مبانی دستور وابستگی

نظریه دستور وابستگی در سال‌های میانی قرن بیستم به کوشش لوسین تسنی^۳ (۱۹۵۳) در فرانسه بالیدن گرفت. به موجب عینی بودن مبانی این دستور و احتراز آن از بحث‌های ذهنی و انتزاعی و نیز به علت عرضه اطلاعات در قالب فرمول‌های ساده، دستاوردهای این دستور را به‌سادگی می‌توان صورت‌بندی و الگوریتمی کرد و از آن در زمینه پردازش زبان استفاده نمود (طیب‌زاده، ۱۳۹۳: ۲۷-۲۸). مزیت مهم دستور وابستگی، تجزیه جملات زبان‌های با آرایش (نسبتاً) آزاد واژگانی است. این در حالی است که در رویکردهای مبتنی بر سازه برای هر جایگاه احتمالی در درخت تجزیه قانونی مجزا و ثابت نیاز است. برای مثال در دستور زایشی کلمات جمله دارای جایگاه‌های ساختاری ثابت و مشخصی هستند و چنانچه گشتار جا به جایی هم بر جمله اعمال شده باشد، همچنان جایگاه اولیه کلمات در درخت نحوی مدنظر قرار می‌گیرد.

نقطه قوت دیگر دستور یادشده این است که گره‌های تشکیل‌دهنده درخت وابستگی جمله به تعداد کلمات موجود در جمله است و تجزیه‌گر با در نظر گرفتن قوانین وابستگی، تنها به اتصال گره‌های موجود می‌پردازد (کووینگتون، ۲۰۰۱)؛ بنابراین، می‌توان گفت دستور وابستگی نسبت به دیگر رویکردهای رقیب، همسویی بیشتری با اصل ساده‌سازی نظام دستوری دارد؛ بی‌تردید می‌توان گفت این دستور پایگاه نظری مناسبی را برای پردازش متنی زبان‌های ضمیرانداز با ترتیب خطی (نسبتاً) آزاد چون زبان فارسی^(۲) فراهم می‌آورد. در ادامه، شیوه تحلیل در این دستور را به اختصار توضیح خواهیم داد.

1. Maximum Entropy (ME). 2. Conditional Random Fields (CRFs). 3. Lucien Tesnière.

تحلیل نحوی در دستور وابستگی از فعل آغاز می‌شود و سپس متمم‌های بلافصل آن (فاعل، متمم مستقیم، متمم حرف‌اضافه‌ای و ...) به دنبال می‌آیند. در این دستور، هر گروه نحوی، چه جمله باشد و چه گروه کوچک‌تری همچون گروه اسمی یا صفتی و غیره، از بیش از پنج نوع عنصر تشکیل‌دهنده فراتر نمی‌رود. عناصر سازنده گروه‌ها تنها دارای یک هسته هستند که مقوله گروه با آن مشخص می‌شود و گروه‌های اسمی، صفتی، حرف‌اضافه‌ای، قیدی و فعلی را تشکیل می‌دهد. می‌توان به این هسته‌ها چهار نوع وابسته (به ترتیب میزان وابستگی‌شان به هسته) اضافه نمود. این وابسته‌ها موارد زیر را شامل می‌شوند:

الف- متمم اجباری هسته که در ساختار موضوعی هسته قرار دارند و حذف آن‌ها گروه را نادستوری می‌کند؛ به‌عنوان نمونه در جمله «استناد به آیات شایسته است.»، کلمه «استناد» هسته گروه اسمی و عبارت «به آیات» متمم اجباری آن است. ب- متمم اختیاری هسته است که در ساختار موضوعی هسته قرار دارند اما حذف این نوع متمم‌ها گروه را نادستوری نمی‌کند؛ برای مثال در جمله «بحث با او بی‌فایده است.»، کلمه «بحث» هسته گروه اسمی و عبارت «با او» متمم اختیاری آن است. ج- افزوده خاص هسته که از منظر مناسبت‌های معنایی تنها با برخی از هسته‌ها به کار می‌روند اما در ساختار موضوعی هسته شرکت نمی‌کنند؛ قیدی نظیر «باعلاقه» تنها با برخی افعال نظیر «گوش دادن، درس خواندن و ...» به کار می‌رود و استفاده از آن در ترکیب با افعالی همانند «آتش گرفتن، باریدن و ...» موجب بدساخت شدن جمله می‌شود. د- افزوده‌های عام هسته که وابسته کل گروه محسوب می‌شوند با هر هسته و هر گروهی به کار می‌روند؛ به‌عنوان مثال قید «متأسفانه» را می‌توان با هر فعلی استفاده کرد.

در اینجا شایسته است به این نکته اشاره کنیم که در پیکره وابستگی فارسی دانشگاه اویسالا که مورد استفاده پژوهش حاضر است، برچسب‌های متفاوتی برای انواع وابسته‌ها (مانند وابسته حرف‌اضافه‌ای/ اسمی) به کار رفته است؛ اما در این میان، اجباری یا اختیاری بودن و خاص یا عام بودن وابسته‌ها لحاظ نشده است. این دسته‌بندی کلی و عدم ورود به جزئیات موجب ساده‌سازی تحلیل شده و با استناد به آن قواعد موجزتری برای درج اضافه می‌توان ارائه داد. به همین سبب، در ادامه این پژوهش نیز برای تجزیه و تحلیل جملات پیکره و استخراج قواعد تنها به سه عنوان هسته، وابسته پیشین و

وابسته پسین اکتفا می‌کنیم. در سطرهای آتی، ساخت اضافه را مبنا قرار می‌دهیم و به معرفی هر یک از عناصر سازنده گروه‌های نحوی خواهیم پرداخت.

۴- روابط وابستگی در ساخت اضافه و تشخیص جایگاه نشانه اضافه

همان‌طور که در بخش‌های پیشین اشاره شد، به‌منظور تحلیل جملات فارسی در قالب دستور وابستگی و استخراج قواعد تشخیص جایگاه نشانه اضافه از پیکره وابستگی دانشگاه اویسالا استفاده شده است؛ اما تلاش ما بر این است تا به ارائه قواعد جامعی پردازیم تا تنها با جایگزینی برچسب‌های خاص هر پیکره در این قواعد، بتوانند مورد استفاده پیکره‌های وابستگی فارسی مختلف با برچسب‌های نحوی متفاوت نیز قرار گیرند. از این‌رو، در تعریف قواعد تنها به دو طبقه وابسته‌های پیشین و پسین هسته استناد خواهیم کرد. از آنجا که از کسره اضافه برای افعال و حروف ربط استفاده نمی‌شود، در بخش حاضر، با نگرشی ریاضی‌بنیاد و در چارچوب دستور وابستگی به ارائه قواعد درج کسره اضافه در گروه‌های اسمی، صفتی، حرف‌افزای، قیدی و همپایه می‌پردازیم. از رهگذر این قواعد می‌توان در سامانه‌های رایانه‌ای به‌صورت کارا به تشخیص نشانه اضافه در پیکره‌های متفاوت پرداخت. پیش از آنکه به ارائه قواعد پردازیم، این نکته را یادآور می‌شویم که دستور وابستگی تجزیه نحوی جملات را به‌صورت خطی انجام می‌دهد. از این ویژگی می‌توان برای تشخیص نوع وابسته هسته (به لحاظ پیشین یا پسین بودن) استفاده کرد و جایگاه هر کلمه را با نمایه‌ای در جمله نشان داد. اکنون، با توجه به این مطلب، قاعده‌های کلی زیر را برای تشخیص نوع وابسته به کار می‌بریم:

$$\text{Premodifier}(x_n) = \{x_i \in S \mid i < n \text{ and } \exists r \ x_n \xrightarrow{r} x_i\}$$

$$\text{Postmodifier}(x_n) = \{x_i \in S \mid i > n \text{ and } \exists r \ x_n \xrightarrow{r} x_i\}$$

در قاعده‌های بالا، منظور از x کلمات موجود در جمله S است. x_n هسته و x_i وابسته آن است و رابطه وابستگی بین هسته و وابسته با نماد \xrightarrow{r} نشان داده شده است که r در این نماد، نشانه رابطه وابستگی (Relation) است. چنانچه اندیس کلمه وابسته در جمله کوچک‌تر از n (اندیس هسته) باشد، x_i وابسته پیشین هسته خواهد بود؛ وابسته‌های پیشین هسته با مجموعه Premodifier معین شده است. حال چنانچه این اندیس بزرگ‌تر از n باشد، x_i وابسته پسین هسته محسوب می‌شود که این نوع وابسته‌ها با مجموعه Postmodifier مشخص شده‌اند. در ادامه، برای تعریف قواعد درج

نشانه اضافه از زبان منطق مرتبه اول استفاده شده و چهار تابع $Ezafe$ ، CAT_d ، POS و $Indef$ تعریف شده که به ترتیب شرایط درج نشانه اضافه برای کلمات جمله، برچسب مقوله‌ای نحوی کلمه (به همراه رابطه وابستگی آن)، نقش نحوی کلمه در جمله و نکره بودن کلمه را نشان می‌دهند. بعلاوه، از آنجاکه تنها برخی از پیش‌توصیف‌گرها^(۳) و سورها^(۴) در ساخت اضافه شرکت دارند، تابعی با عنوان $EzMod$ تعریف شده که این کلمات را شامل می‌شوند.

۴-۱- قواعد درج نشانه اضافه در گروه اسمی

اسم‌ها گاه وابسته یک هسته در جمله هستند و گاه به صورت هسته در گروه اسمی ظاهر می‌شوند و وابسته می‌پذیرند. در گروه‌های اسمی (و صفتی) فارسی آرایش عناصر عمدتاً ثابت و مشخص است. به همین دلیل باید وابسته‌های هر هسته را در این زبان به دو دسته پسین و پیشین تقسیم کرد (طیب‌زاده، ۱۳۹۳: ۳۷)، باطنی (۱۳۸۹: ۱۳۹-۱۴۳)، مشکوٰۃ‌الدینی (۱۳۷۰: ۱۶۲) و طیب‌زاده (۱۳۹۳: ۲۱۱) وابسته‌های پیشین اسم را به ترتیب سورهای بسیط و مرکب، صفات (اشاره، پرسشی، تعجبی، مبهم، تأکیدی)، اعداد (ترتیبی، اصلی و مبهم) و صفات عالی، ممیزها و شاخص‌ها معرفی کرده‌اند که ترتیب این وابسته‌ها قبل از اسم ثابت است. آن‌ها وابسته‌های پسین اسم را نیز به ترتیب شامل شاخص، صفات (بیانی، شمارشی، نسبی، تفضیلی و مطلق)، وابسته اضافه‌ای غیر مالک، بدل، گروه حرف‌اضافه‌ای، وابسته اضافه‌ای مالک و بند دانستند.

در جایگاه پیشین اسم، تنها پیش‌توصیف‌گرها و سورهای بسیط نظیر «اکثر، همه، بیش‌تر و تمام» و شاخص‌هایی نظیر «خانم، آقا، حضرت و جناب» با کسره اضافه به هسته اسمی وابسته می‌شوند. به علاوه، صفات پیشین عددی (عدد+م) نیز با کسره به هسته اسمی متصل می‌شوند (مثال ۱- الف). در خصوص صفات عالی و صفات پیشین عددی (عدد+م) ذکر این نکته الزامی است که اگر قبل از ضمیر به کار روند، درج کسره برای آن‌ها الزامی است (مثال ۱- ب). سایر وابسته‌های پیشین بدون استفاده از کسره با هسته اسمی ترکیب می‌شوند.

(۱) الف. نهم دی

ب. بهترین/دومین آن‌ها

در جایگاه بعد از اسم، به جز وابسته شاخص پسین، بدل و بند، حضور سایر وابسته‌ها در جمله باعث درج کسره اضافه در انتهای هسته اسمی می‌شوند. بعلاوه، بعد از ضمائر

نیز می‌توان وابسته صفتی استفاده کرد. در مثال‌های زیر روابط وابستگی هسته اسمی و ضمیری و وابسته‌های آن‌ها در ساخت اضافه مشاهده می‌شود:

(۲) الف. استاد بزرگوار آمدند. (وابسته صفتی)

ب. راه تهران را یافتیم. (وابسته اضافه‌ای غیر مالک)

پ. کتاب علی (وابسته اضافه‌ای مالک)

ت. من دیوانه (وابسته صفتی)

خاطر نشان می‌شود چنانچه هسته اسمی دارای تکواژ نکره‌ساز «ی» باشد، در آن صورت هسته اسمی فاقد نشانه اضافه خواهد بود.

(۳) او راهی مناسب به ما نشان داد.

در خصوص وابسته اضافه‌ای اسم، لارسون (۲۰۰۹) و لارسون و یاماکیو (۲۰۰۸) در پژوهش خود به این نکته اشاره دارند که بند موصولی مخفف^۱ (مثال ۴- الف) برخلاف بندهای خود ایستا^۲ (مثال ۴- ب) در جایگاه وابسته اسمی کسره اضافه می‌گیرند. به عبارتی، کسره اضافه با بند موصولی محدودکننده^۳ ظاهر نمی‌شود:

(۴) الف. این جوان از سوئیس برگشته

ب. این جوانی که از سوئیس برگشته

کهنمویی‌پور (۲۰۱۶) نیز عدم مشاهده کسره اضافه در بند موصولی محدودکننده را این‌گونه تبیین می‌کند که نشانه موصولی^۴ واج‌گونه‌ای از نشانگر اضافه به حساب می‌آید. به عقیده ایشان در چنین ساخت‌هایی به جای استفاده از کسره از شکل واجگونه آن یعنی «ی» استفاده می‌شود. بدین ترتیب، قواعد درج نشانه اضافه برای هسته اسمی و ضمیری را می‌توان به دو قاعده کلی (۱) و (۲) تقسیم‌بندی کرد. این قواعد به ترتیب مربوط به وابسته‌های پیشین و پسین هسته‌های ذکر شده هستند.

قاعده اول: درج نشانه اضافه در گروه‌های اسمی با وابسته‌های پیشین

$\forall x_i \in S \text{ Ezafe}(x_i) = \text{True} \text{ IF } x_i \in \text{Premodifier}(x_n) \text{ and } (\text{CAT}_d(x_i) \in \text{Y or } \text{CAT}_d(x_i) \in \text{EzMod}) \text{ and } \text{POS}(x_n) = \text{Noun and } \text{Indef}(x_n) = \text{False}$ الف

$Y = \{ \text{صفت شمارشی ترتیبی (عدد + م)} \}$

$\text{EzMod} = \{ \text{برخی پیش توصیفگرها و سورهای بسیط} \}$

1. reduced relative clauses. 2. finite relatives. 3. restrictive relative clause. 4. relative marker.

۱- $\forall x_i \in S \text{ Ezafe}(x_i) = \text{True}$ IF $x_i \in \text{Premodifier}(x_n)$ and $\text{CAT}_d(x_i) \in$
 Y or $\text{CAT}_d(x_i) \in \text{EzMod}$ and $\text{POS}(x_n) = \text{Pronoun}$ ب

Y = {صفت شمارشی ترتیبی (عدد + - مین)، صفت‌های عالی}

طبق قاعده (۱-الف)، چنانچه قبل از هسته اسمی (x_n) وابسته‌های پیشین از نوع پیش توصیف‌گرها و سوره‌های بسیط EzMod و صفت شمارشی ترتیبی (عدد + - م) به کار رود، این وابسته‌ها پذیرای کسره اضافه خواهند بود. طبق قاعده (۱-ب) چنانچه هسته اسمی از نوع ضمیر باشد، وابسته‌های پیشین از نوع سوره‌های بسیط EzMod، صفت شمارشی ترتیبی (عدد + - مین) و صفت‌های عالی واجد نشانه کسره هستند.

قاعده دوم: درج نشانه اضافه در گروه‌های اسمی با وابسته‌های پسین

۲ $\text{Ezafe}(x_n) = \text{True}$ IF $\exists x_i x_i \in \text{Postmodifier}(x_n)$ and $\text{CAT}_d(x_i) \in$
 Y and $\text{POS}(x_n) = \text{Noun or Pronoun}$ and $\text{Indef}(x_n) = \text{False}$ -

Y = {صفت شمارشی، کلیه صفت‌های بیانی، صفت برتر، وابسته اضافه‌ای (غیرمالک)، وابسته حرف {اضافه‌ای (از نوع بند موصولی مخفف)، وابسته اضافه‌ای (مالک)}

طبق قاعده بالا، چنانچه بعد از هسته اسمی وابسته‌های پسین از نوع صفت شمارشی، کلیه صفت‌های بیانی، صفت برتر، وابسته اضافه‌ای (غیر مالک)، وابسته حرف اضافه‌ای (از نوع بند موصولی مخفف)، وابسته اضافه‌ای (مالک) استفاده شود، هسته اسمی دارای کسره است.

۴-۲- قواعد درج نشانه اضافه در گروه صفتی

متمم‌های صفت از نوع وابسته‌های پسین هستند که شامل متمم حرف اضافه‌ای، متمم اضافه‌ای و متمم بندی می‌شوند. تنها متمم اضافه‌ای صفت است که باعث درج کسره در انتهای هسته صفتی می‌شود. در زیر مثال‌هایی از متمم‌های صفت مشاهده می‌شود:

(۵) الف. این اتاق مشرف بر باغ است.

ب. او مغل آسایش شما است.

پ. او شاهد بود که چگونه مردم را کشتند.

افزوده‌های صفت به هر دو دسته وابسته پیشین و پسین تعلق دارند. افزوده پیشین صفت، قید صفت است که بدون کسره قبل از هسته صفتی به کار می‌رود (مثال ۶- الف) و افزوده پسین صفت، «صفت صفت» است که افزوده خاص به‌شمار می‌رود. این

وابسته در صورت وجود، باعث درج کسره اضافه در انتهای هسته صفتی می‌شود (مثال ۶ - ب):

(۶) الف. این خانه اصلاً خیلی قشنگ نیست.

ب. رنگ آن ماشین، سبز متمایل به آبی بود.

با توجه به مثال‌های بالا، در ادامه به قاعده درج نشانه اضافه برای وابسته‌های پسین هسته صفتی اشاره می‌گردد.

قاعده سوم: درج نشانه اضافه در گروه‌های صفتی با وابسته‌های پسین

$Ezaf(x_n) = \text{True}$ IF $\exists x_i x_i \in \text{Postmodifier}(x_n)$ and $CAT_d(x_i) \in$ ۳
 Y and $POS(x_n) = \text{Adjective}$ -

$Y = \{\text{صفت صفت، اسم}\}$

طبق قاعده بالا، چنانچه بعد از هسته صفتی (x_n) ، وابسته‌هایی از نوع صفت صفت و اسم استفاده شود، هسته صفتی پذیرای نشانه اضافه خواهد بود.

۳-۴- قواعد درج نشانه اضافه در گروه حرف‌اضافه‌ای

هر گروه حرف‌اضافه‌ای از هسته (حرف‌اضافه) و گروه اسمی (وابسته) تشکیل شده است که تنها متمم اجباری آن، گروه اسمی به حساب می‌آید. البته گروه‌های حرف‌اضافه‌ای گاهی قیودی را به عنوان افزوده عام می‌پذیرند که محل آن‌ها در ابتدای گروه است. در مثال زیر «تنها» افزوده عام حرف اضافه «از» است که بدون کسره به هسته حرف‌اضافه‌ای وابسته شده است:

(۷) تنها از این راه می‌توان مشکل را حل کرد.

در اینجا نکته شایان ذکر آن است که در پیکره سراجی برای تمامی حروف اضافه برچسب واحد P در نظر گرفته شده و تمایزی میان حروف اضافه واقعی و سایر حروف اضافه وجود ندارد؛ حال آنکه برای استخراج قاعده درج نشانه اضافه در گروه‌های حرف‌اضافه‌ای قائل شدن به این تمایز امری اجتناب‌ناپذیر است؛ بنابراین، در پژوهش حاضر به پیروی از سمیعان (۱۹۹۴) قائل به سه طبقه مستقل برای حروف اضافه می‌شویم.

الف- حروف اضافه‌ای که با کسره اضافه به کار نمی‌روند. تعداد این حروف اضافه زیاد

نیست. به این دسته از حروف اضافه، حروف اضافه واقعی^۱ گویند: در، با، به، از، بر، پس از، نسبت به، بعد از، پیش از، الی، تا، همچون.

ب- حروف اضافه‌ای که به طبقه باز^۲ تعلق دارند و لزوماً قبل از متمم آن‌ها کسره اضافه استفاده می‌شود: بدون، بین، زیر، پشت، پایین، علیه، توسط، روی، حدود، نزد، درباره، پس، مورد، ضمن، بعد، طی، پیش، علی‌رغم، جهت، برابر، ضد، قابل، طبق، بابت، قریب، خلاف، سر، مجاور، دنبال، کنار، پی، مقابل، حسب.

ج- حروف اضافه‌ای که به طور اختیاری می‌توانند با کسره اضافه به کار روند: تو(ی)، رو(ی)، پهلو(ی)، جلو(ی)، بالا(ی).

بر اساس دسته‌بندی بالا، قاعده زیر را برای درج نشانه اضافه در هسته حرف اضافه ارائه می‌دهیم:

قاعده چهارم: درج نشانه اضافه در گروه‌های حرف اضافه‌ای با وابسته‌های پسین

$$\text{Ezafe}(x_n) = \text{True} \text{ IF } \exists x_i x_i \in \text{Postmodifier}(x_n) \text{ and } \text{CAT}_d(x_i) \in \text{Y and POS}(x_n) \neq \text{True_Preposition} \quad ۴$$

$$Y = \{\text{اسم، ضمیر}\}$$

طبق قاعده بالا، چنانچه بعد از حروف اضافه (به غیر از حروف اضافه واقعی) وابسته‌های پسین از نوع اسم و ضمیر استفاده شود، هسته حرف اضافه‌ای دارای نشانه اضافه است.

۴-۴- قواعد درج نشانه اضافه در گروه قیدی

قیدها نیز دارای دو وابسته پیشین و پسین هستند. وابسته پیشین قیدها افزوده‌هایی نظیر «اندکی، بسیار، خیلی، شدیداً، کمی و...» هستند که هسته گروه قیدی را از حیث مقدار توصیف می‌کنند. وابسته پسین قیدها، گروه حرف اضافه‌ای «زی» و گروه اسمی است. تنها گروه اسمی با کسره اضافه به هسته قیدی متصل می‌شود.

(۸) الف. او آرام‌تر از من غذا می‌خورد. (متمم حرف اضافه‌ای قید)

ب. نظیر این کتاب / او (متمم اسمی قید)

قاعده پنجم: درج نشانه اضافه در گروه‌های قیدی با وابسته‌های پسین
 $Ezafe(x_n) = True \text{ IF } \exists x_i \ x_i \in Postmodifier(x_n) \text{ and } CAT_d(x_i) \in Y \text{ and } POS(x_n) = Adverb$ ۵

$Y = \{\text{اسم، ضمیر}\}$

طبق قاعده بالا، چنانچه بعد از هسته قیدی، وابسته‌های پسین از نوع اسم و ضمیر استفاده شود، هسته قیدی دارای نشانه اضافه است.

۴-۵- قواعد درج نشانه اضافه در گروه‌های با بیش از یک وابسته

در گروه‌های اسمی یا صفتی ممکن است چندین وابسته پسین به صورت پی‌درپی بعد از هسته به کار روند و هیچ کدام از این وابسته‌ها، به‌عنوان هسته، وابسته نداشته باشند؛ اما به دلیل حضور وابسته‌های اسمی یا صفتی بعد از آن‌ها، حامل نشانه اضافه شوند. (۹) کتاب قرمزِ بزرگِ مریم که گم شد.

در مثال بالا «قرمز» به‌عنوان وابسته پسین هسته به دلیل وجود وابسته اسمی دیگری پس از خود، دارای نشانه اضافه است؛ بدون آن که وابسته‌ای داشته باشد. در این حالت، باید تمامی وابسته‌های پسین از نوع اسم و صفت به غیر از آخرین وابسته دارای نشانه اضافه باشند؛ بنابراین، در مثال (۹) وابسته‌های «قرمز» و «بزرگ» دارای نشانه اضافه هستند و «مریم» به‌عنوان آخرین وابسته اسمی «کتاب» نشانه اضافه نمی‌گیرد. در زیر ساختار مربوطه قابل مشاهده است:

$X_n + Ezafe + X_{n+1} + Ezafe + X_{n+2} + Ezafe + \dots + X_{n+k-1} + Ezafe + X_{n+k}$

بنابراین، اگر X_n هسته اسمی، ضمیری یا صفتی باشد و k تعداد وابسته پسین از مجموعه Y داشته باشد، آنگاه قاعده درج کسره اضافه برای وابسته‌های پسین هسته به شکل زیر خواهد بود که تنها وابسته آخر هسته فاقد نشانه اضافه خواهد بود:

قاعده ششم: درج نشانه اضافه در گروه‌های اسمی/صفتی با چند وابسته پسین

$\forall x_i \in S \ Ezafe(x_i) = True \text{ IF } x_i \in Postmodifier(x_n) \text{ and } CAT_d(x_i) \in Y \text{ and } POS(x_n) = Noun \text{ or } Pronoun \text{ and } Indef(x_n) = False \text{ and } i \neq n + k$ ۶ الف

$Y = \{\text{صفت شمارشی، کلیه صفت‌های بیانی، صفت برتر، وابسته اضافه‌ای (غیرمالک)، وابسته حرف (اضافه‌ای (از نوع بند موصولی مخفف)، وابسته اضافه‌ای (مالک)}$

۶
 $\forall x_i \in S \text{ Ezafe}(x_i) = \text{True} \text{ IF } x_i \in \text{Postmodifier}(x_n) \text{ and } \text{CAT}_d(x_i) \in$
 $Y \text{ and } \text{POS}(x_n) = \text{Adjective} \text{ and } i \neq n + k$

Y = {صفت، اسم}

قواعد بالا متناظر قواعد هسته اسمی و ضمیری و صفتی هستند که اندکی تغییر یافته‌اند، به این صورت که تعداد بیش از یک وابسته پسین نیز برای آن‌ها در نظر گرفته شده است.

۴-۶- قواعد تشخیص نشانه اضافه در گروه‌های هم‌پایه

طبق نظر طبیب‌زاده (۱۳۹۳) حروف عطف به‌عنوان پیونددهنده عناصر هم‌پایه تعیین شده‌اند که گروه‌های مختلف نحوی را به هم متصل می‌کند (مثال ۱۰).

(۱۰) [[پسرِ شجاعِ مریم] و [دخترِ باهوشِ مینا]]

اگر x_n که هسته اسمی، ضمیری یا صفتی است، دارای k وابسته پسین باشد و حرف

عطف (x_m)، m امین وابسته پسین x_n و ($m < k$) باشد، آنگاه داریم:

$X_n + \text{Ezafe} + X_{n+1} + \text{Ezafe} + \dots + X_{n+m-1} + \text{Conjunction}_{n+m} + X_{n+m+1} + \text{Ezafe} +$
 $\dots + X_{n+k-1} + \text{Ezafe} + X_{n+k}$

در مثال (۱۰)، علاوه بر حالت همپایگی، گروه‌های اسمی با بیش از یک وابسته هم مشاهده می‌شود؛ بنابراین، برای لحاظ کردن حالت همپایگی نیز از قاعده ششم که در زیر بخش پیشین ارائه شد، استفاده می‌کنیم و تنها اندیس کلمه قبل از حروف عطف به کار رفته در جمله را در این قاعده اضافه می‌کنیم؛ بدین ترتیب، نشانه اضافه برای این کلمه لحاظ نمی‌گردد.^(۶)

قاعده هفتم: درج نشانه اضافه در گروه‌های هم‌پایه با بیش از یک وابسته

۷-
 $\forall x_i \in S \text{ Ezafe}(x_i) = \text{True} \text{ IF } x_i \in \text{Postmodifier}(x_n) \text{ and } \text{CAT}_d(x_i) \in$
 $Y \text{ and } \text{POS}(x_n) = \text{Noun or Pronoun} \text{ and } \text{Indef}(x_n) = \text{False} \text{ and } i \neq n +$
 $k \text{ and } i \neq n + m - 1$

Y = {صفت شمارشی، کلیه صفت‌های بیانی، صفت برتر، وابسته اضافه‌ای (غیرمالک)، وابسته حرف
 {اضافه‌ای (از نوع بند موصولی مخفف)، وابسته اضافه‌ای (مالک)}

۷
 $\forall x_i \in S \text{ Ezafe}(x_i) = \text{True} \text{ IF } x_i \in \text{Postmodifier}(x_n) \text{ and } \text{CAT}_d(x_i) \in Y \text{ and } \text{POS}(x_n) = \text{Adjective} \text{ and } i \neq n + k \text{ and } i \neq n + m - 1$
 -
 ب
 -

$Y = \{\text{صفت صفت، اسم}\}$

قاعده بالا برای حرف عطف نیز کسره اضافه در نظر نمی‌گیرد زیرا در مجموعه Y وجود ندارد. بدین ترتیب، برای اسامی، ضمائر و صفات با هر تعداد وابسته پسین و شرایط همپایگی می‌توان تنها دو قاعده بالا را به عنوان قاعده جامع در نظر گرفت.

۴-۷- قواعد تشخیص نشانه اضافه در گروه‌های اسمی با بیش از یک وابسته از نوع گروه

در گروه‌های اسمی ممکن است چندین وابسته پسین به صورت پی‌درپی بعد از هسته به کار روند و یکی از این وابسته‌ها، خود هسته گروه دیگری محسوب شود:

(۱۱) الف. [کپسول [ویتامین [دی]] من]

ب. [چشمان [آبی [کمرنگ]] مینا]

در مثال (۱۱-الف)، کلمه «کپسول» دارای دو وابسته «ویتامین» و «من» است که خود کلمه «ویتامین» دارای وابسته دیگری است. با در نظر داشتن قاعده (۶-الف)، کلمه «ویتامین» به سبب وجود کلمه «من» نشانه اضافه دریافت می‌کند. به منظور لحاظ کردن کسره اضافه برای کلمه «دی»، باید وضعیت کسره‌دار بودن هسته آن بررسی شود. چنانچه هسته (یعنی کلمه «ویتامین») از قبل دارای کسره اضافه باشد، تمامی وابسته‌های آن نیز دارای نشانه اضافه هستند.

۸-
 $\forall x_i \in S \text{ Ezafe}(x_i) = \text{True} \text{ IF } x_i \in \text{Postmodifier}(x_n) \text{ and } \text{Ezafe}(x_n) = \text{True} \text{ and } \text{CAT}_d(x_i) \in Y \text{ and } \text{POS}(x_n) = \text{Noun} \text{ and } \text{Indef}(x_n) = \text{False}$
 الف
 -

صفت شمارشی، کلیه صفت‌های بیانی، صفت برتر، وابسته اضافه‌ای (غیرمالک)، وابسته {حرف} اضافه‌ای (از نوع بند موصولی مخفف)، وابسته اضافه‌ای (مالک)

به عبارتی قاعده شش را برای این نوع ساخت لحاظ کرده و تنها شرایط کسره‌دار بودن هسته را به آن اضافه می‌کنیم. بعلاوه، در این قاعده برای تمامی وابسته‌های هسته کسره لحاظ می‌شود. بدین ترتیب محدودیت عدم اعمال کسره بر روی آخرین وابسته را

نیز لحاظ نمی‌کنیم. این نوع ساخت برای مثال (۱۱-ب) نیز صادق است که یک گروه صفتی، وابسته میانی گروه اسمی در نظر گرفته می‌شود:

$$\forall x_i \in S \text{ Ezafe}(x_i) = \text{True} \text{ IF } x_i \in \text{Postmodifier}(x_n) \text{ and Ezade}(x_n) = \text{True and CAT}_d(x_i) \in Y \text{ and POS}(x_n) = \text{Adjective}$$

$$Y = \{\text{صفت صفت، اسم}\}$$

۵- نتیجه

از آنجایی که نشانه اضافه و عدم درج آن در نوشتار معضل اصلی پژوهش‌های مرتبط با پردازش متن به حساب می‌آید، پژوهش حاضر به بررسی جایگاه‌های درج نشانه اضافه در چارچوب دستور وابستگی پرداخت. پیرو این امر، با در نظر داشتن روابط هسته و وابسته در گروه‌های نحوی دارای نشانه اضافه، هفت قاعده منطقی برای درج کسره اضافه در گروه‌های اسمی، صفتی، حرف اضافه‌ای، قیدی و همپایه استخراج شد. در قواعد تنها به دو طبقه وابسته‌های پیشین و پسین هسته استناد شده است که این مسئله موجب سادگی قواعد شده و آن‌ها را از وابستگی به پیکره‌ای خاص می‌رهاند. بدین ترتیب، فارغ از برجسب‌های به کار رفته در هر پیکره وابستگی، قواعد ارائه شده بر روی انواع این پیکره‌ها قابل اعمال هستند و با استناد به این قواعد می‌توان در سامانه‌های رایانه‌ای به شناسایی جایگاه نشانه اضافه پرداخت.

پی‌نوشت‌ها

۱. در اینجا باید به این نکته اشاره کرد که پیکره سراجی (۲۰۱۵) کاستی‌هایی چون عدم نرمال‌سازی کلمات، جداسازی اشتباه کلمات مرکب، عدم برجسب‌دهی صحیح اجزای کلام، عدم تعیین صحیح روابط وابستگی گروه‌های هم‌پایه و غیره است که پرداختن به آن‌ها از حوصله پژوهش حاضر بیرون است.
۲. همان‌طور که می‌دانیم در زبان‌های ضمیرانداز، گروه اسمی فاعل گاه تظاهر آوایی ندارد. البته شایان ذکر است که تظاهر فاعل در زبان فارسی (همچون بسیاری از زبان‌های ضمیرانداز) در پاره‌ای از ساخت‌ها الزامی است؛ در صورت حذف گروه فعلی کوچک (VP) و موضوع‌های درونی فعل در برخی از جملات، حذف فاعل به ساختی نادرستی می‌انجامد. برای اطلاعات بیشتر در این زمینه نگاه کنید به انوشه (۱۳۸۷: ۱۶۸-۱۶۹).
۳. پیش‌توصیف‌گرها کلماتی هستند که پیش از توصیف‌گرها به کار برده می‌شوند. در عبارتی نظیر «همه این کتاب‌ها»، کلمه «همه» پیش‌توصیف‌گر تلقی شده که پیش از توصیف‌گر «این» ظاهر شده است.
۴. کلماتی نظیر «همه، کلی، تمام، اکثر» سورها و پیش‌توصیف‌گرهایی هستند که در گروه اسمی نیاز به کسره اضافه دارند. کلماتی همانند «برخی، بعضی، این، آن» سورها و توصیف‌گرهایی هستند که پیش از هسته اسمی

بدون کسره اضافه استفاده می‌شوند. از آنجایی که در پیکره وابستگی اوپسالا بین این دو دسته از کلمات تمایزی وجود ندارد و برچسب یکسانی برای هر دو گروه در نظر گرفته شده است، در سامانه حاضر، تابعی با عنوان EzMod جهت تشخیص درج نشانه اضافه در نظر گرفته شده است که تنها شامل سورها و پیش توصیف‌گرهایی می‌شود که با کسره اضافه پیش از هسته اسمی به کار می‌روند.

۵. در پیکره وابستگی فارسی دانشگاه اوپسالا، برچسب‌های مختلفی برای اجزای کلام در ساخت اضافه مورد استفاده قرار می‌گیرند. این برچسب‌ها مورد استفاده پژوهش حاضر نیستند اما به آن‌ها اشاره می‌کنیم تا خواننده علاقه‌مند را در درک بهتر پیکره یاری رساند. برچسب‌های N_SING و N_PL برای کلیه اسامی مفرد و جمع، PRO برای ضمائر، ADJ برای کلیه صفات عددی ترتیبی و کلیه صفات بیانی به غیر از صفات مفعولی، ADJ_CMPR برای صفات برتر، ADJ_SUP برای صفات عالی، ADJ_INO برای صفات مفعولی، ADV برای قیود، CON برای حروف عطف، DET برای پیش‌توصیف‌گر، NUM برای صفات عددی شمارشی، P برای کلیه حروف اضافه. از ایرادات برچسب‌زنی این پیکره می‌توان به این مسئله اشاره داشت که برچسب واحد ADJ برای کلیه صفات عددی ترتیبی (ترکیب با م و -مین) استفاده شده است و انواع طبقات حرف اضافه برچسب واحدی دارند.

۶. در پیکره وابستگی حاضر، تجزیه وابستگی همپایگی به اشتباه صورت گرفته است. در ساخت‌های هم‌پایه، حروف عطف باید به‌عنوان هسته لحاظ شوند و دو گروه هم‌پایه وابسته آن‌ها در نظر گرفته شوند. در مثال مذکور چنانچه حرف عطف «و» و کلمه «دختر» وابسته کلمه «پسر» در نظر گرفته شوند، با حذف این وابسته‌ها جمله حاصل بدساخت خواهد شد و فعل جمع جمله با فاعل مفرد به لحاظ شمار مطابقت نخواهد داشت. به‌منظور پوشش این نوع ساخت‌ها در این پیکره، قاعده جدیدی برای تشخیص جایگاه اضافه تعریف می‌کنیم که حروف عطف به‌عنوان وابسته پسین هسته لحاظ شده‌اند.

منابع

- ابوالقاسمی، محسن (۱۳۷۵). *دستور تاریخی زبان فارسی*، (چاپ اول)، تهران، سمت.
- باطنی، محمدرضا (۱۳۸۹). *توصیف ساختمان دستوری زبان فارسی بر بنیاد یک نظریه عمومی زبان*. (چاپ بیست و دوم). تهران: امیرکبیر.
- بی‌جن‌خان، محمود (۱۳۸۲). کسره اضافه از دیدگاه معنایی، *جشن‌نامه دکتر علی‌اشرف صادقی*، صص ۲۲۹-۲۵۳.
- بی‌جن‌خان، محمود (۱۳۸۷). *امکان‌سنجی برای تجزیه و تحلیل کسره اضافه زبان فارسی*، با روش تطبیق الگو، تهران، پژوهشگاه فرهنگ، هنر و ارتباطات.
- خانلری، پرویز (۱۳۷۷). *تاریخ زبان فارسی*، تهران، فردوس.
- راستارگویا، وراسرگی یونا (۱۳۴۷). *دستور زبان فارسی میانه*، (مترجم: ولی‌الله شادان)، تهران، بنیاد فرهنگ ایران.
- طباطبائی، علاءالدین (۱۳۸۲). *اسم و صفت مرکب در زبان فارسی*، تهران، مرکز نشر دانشگاهی.
- طیب‌زاده، امید (۱۳۹۳). *دستور زبان فارسی بر اساس نظریه گروه‌های خودگردان در دستور وابستگی*، (چاپ دوم)، تهران، نشر مرکز.
- فرشیدورد، خسرو (۱۳۸۸). *دستور مختصر امروز بر پایه زبانشناسی جدید*، تهران، سخن.

- لازار، ژیلبر (۱۳۸۹). دستور زبان فارسی معاصر، ترجمه مهستی بحرینی، تهران، انتشارات هرمس.
- محمدی بلبان آبادی، صادق (۱۳۸۷). رویکردی کمینه‌گرا به ساخت اضافه در کردی (گوش سوران)، پایان‌نامه کارشناسی ارشد، دانشگاه کردستان.
- محمودی بختیاری، بهروز (۱۳۸۹). ترکیب‌های نحوی در زبان فارسی، رشد آموزش زبان و ادب فارسی. شماره ۹۳: ۳۲-۳۷.
- مشکوة‌الدینی، مهدی (۱۳۷۰). دستور زبان فارسی بر پایه نظریه گشتاری، (چاپ دوم)، خراسان رضوی، انتشارات دانشگاه فردوسی مشهد.
- معین، محمد (۱۳۹۴). اضافه، تهران، صدای معاصر.
- یاراحمدزهی، ناهید. (۱۳۸۱). ساخت درونی گروه اسمی گویش بلوچی سرحدی بر اساس نظریه ایکس-تیره، رساله کارشناسی ارشد، دانشگاه تهران، ایران.
- Abolghassemi, M. 1996. *A historical grammar of the Persian language*. (first edition). Tehran: Samt. [In Persian].
- Asgari, H., J., Maleki and H., Faili. 2014. A Probabilistic Approach to Persian Ezafe Recognition. *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics*.
- Batani, M. R. 2010. *The description of grammatical structure of Persian language based on a general theory of language*. (Twenty-second edition). Tehran: Amirkabir. [In Persian].
- Bauer, L. 1983. *English Word-Formation*. Cambridge: Cambridge University Press.
- Bijankhan, M. 2003. Ezafe from semantic approach. *Festschrift for prof Dr. A. A. Sadeghi*: pp. 229-253. [In Persian].
- Bijankhan, M. 2008. *A Feasibility Study for Analysis of Ezafe in Persian Using Pattern Matching*. Tehran: Research Center for Culture, Art and Communication. [In Persian].
- Brown, K. and J., Miller. 2013. *The Cambridge Dictionary of Linguistics*. Cambridge University Press.
- Chomsky, N. 1993. A minimalist program for linguistic theory. *The View from Building 20*. Ed. By K. Hale and S. Keyser. Cambridge: MIT Press: 1-52.
- Covington, M. A. 2001. "A Fundamental Algorithm for Dependency Parsing". *Proceedings of the 39th Annual ACM Southeast Conference*: 1-8.
- Farshidvard, K. 2008. *The concise grammar based on modern linguistics*. Tehran: Sokhan. [In Persian].
- Ghayoomi, M. and S., Müller. 2010. PerGram: A TRALE implementation of an HPSG fragment of Persian. In *Computer Science and Information Technology (IMCSIT), Proceedings of the 2010 International Multiconference on*: 461-467.

- Ghomeshi, J. 1996. Projection and Inflection: A study of Persian phrase structure. PhD Dissertation. Department of Linguistics, University of Toronto: Canada.
- Ghomeshi, J. 1997. Non-projecting Nouns and the Ezafe Construction in Persian. *Natural Language and Linguistic Theory* 15: 729-788.
- Hashemipour, M. 1989. Pronominalization and Control in Modern Persian. PhD dissertation, University of California at San Diego.
- Holmberg, A. and D. Odden. 2005. The Noun Phrase in Hawramani. *Paper presented at the First International Conference on Iranian Linguistics: Leipzig*.
- Kahnemuyipour, A. 2000. Persian Ezafe Construction Revisited: Evidence for Modifier Phrase. In *John T. Jensen, J. T., van Herk, G. (Eds.), Cahiers Linguistique d'Ottawa, proceedings of the 2000 annual conference of the Canadian Linguistic Association: pp.173-185*.
- Kahnemuyipour, A. 2006. Persian Ezafe construction revisited: Evidence for modifier phrase. In *Proceedings of the annual Conference of the Canadian Linguistic Association: pp. 17-185*.
- Kahnemuyipour, A. 2014. Revising the Persian Ezafe construction: A roll-up movement analysis. *Lingua: pp. 1-24*.
- Kahnemuyipour, A. 2016. The Ezafe Construction: Persian and Beyond. *ConCall-2, Indiana University..*
- Karimi, S., and M., Brame. 1986. A Generalization Concerning the Ezafe Construction in Persian. *Paper presented at Western Conference on Linguistics: Canada*.
- Karimi, Y. 2007. Kurdish Ezafe Construction: Implications for DP Structure, *Lingua* 177: 2159-2177.
- Khanlari, P. 1998. *The history of Persian language*. Tehran: Ferdows. [In Persian].
- Larson, R. 2009. "Chinese as a Reverse Ezafe Language". *Yuyanxue Lancong (Journal of Linguistics): 39*. Peking University, Beijing, pp. 30-85.
- Larson, R. and H., Yamakido. 2005. Ezafe and the Deep Position of Nominal Modifiers. *Paper presented at the Barcelona Workshop on Adjectives and Adverbs: Universitat Pompeu Fabra*.
- Larson, R. and H., Yamakido. 2006. Zazaki "Double Ezafe" as double case-marking. *Paper presented at the Linguistics Society of America Annual Meeting, Aluquerque, NM*.
- Larson, R., Yamakido, H. 2008. Ezafe and the Deep Position of Nominal Modifiers. In: McNally, L., Kennedy, C. (Eds.), *Adjectives and Adverbs: Syntax, Semantics and Discourse. Oxford: Oxford University Press*, 43-70.
- Lazard, G. 2010. *Grammaire du persian contemporain*. (Translator: M. Bahreyni). Tehran: Hermes. [In Persian].

- Lotfi, A. R. 2014. Persian Ezafe as a “figure” marker: A unified analysis. *Colifornia Linguistic Notes*, 39:1.
- Mahmoodi Bakhtari, B. 2010. Word combination in Persian. *The Development of Persian Teaching and Literature Journal*. No. 93: pp. 32-37. [In Persian].
- Mahootian, S. 1993. A Null Theory of Code-switching. PhD dissertation. Northwestern University, Illinois.
- Megerdooimian, K. 2000. Persian Computational Morphology: A Unification-Based Approach. *Memoranda in Computer and Cognitive Science MCCS-00-320*, Computing Research Lab, New Mexico State University.
- Meshkatod Dini, M. 1991. *Introduction to Persian Transformational Syntax*. (second edition). Razavi Khorasan: Ferdowsi University Press. [In Persian].
- Moein, M. 2015. *Ezafe*. Tehran: Sedaye Moaser. [In Persian].
- Mohammadi Bolbolan Abadi, S. 2009. A minimalist approach to Ezafe constructions in Kurdish language (Sorani dialect). MA thesis. Kurdistan University. [In Persian].
- Moinzadeh, A. 2001. An Antisymmetric, Minomalist Approach to Persian Phrase Structure. PhD dissertation. University of Ottawa.
- Moinzadeh, A. 2006. The Ezafe Phrase in Persian: How Complements are added to N°s and A°s. *Journal of Social Sciences & Humanities of Shiraz University* Vol. 23, No. 1.
- Nojoomian, P. 2011. *Towards the Development of an Automatic Diacritizer for the Persian Orthography based on the Xerox Finite State Transducer*. PhD dissertation. Ottawa: Canada.
- Oskouipour, N. 2011. *Converting text to phoneme stream with the ability to recognizing Ezafe marker and homographs applied to Persian speech synthesis*. Msc. Thesis, Sharif University of Technology: Iran.
- Palmer, A. 1971. *The EZAFE Construction in Modern Standard Persian*. PhD Dissertation, University of Michigan: USA.
- Rastorqueva, V. S. 1968. *Dastūr-i zabān-i Fārsī-i miyānah*. (Translator: W. Shadan). Tehran: Iranian Culture Foundation. [In Persian].
- Samiian, V. 1983. *Origins of Phrasal Categories in Persian, an X-bar Analysis*. PhD Dissertation: UCLA: USA.
- Samiian, V. 1994. *The Ezafe Construction: Some Implications for the Theory of X-bar Syntax*. In: Marashi, M. (Ed.). *Persian Studies in North America, Bethesda, MD: Iranbooks*, 17-41.
- Samvelian, P. 2005. *When morphology does better than Syntax: The Ezafe construction in Persian*.
- Samvelian, P. 2008. *The Ezafe as a Head-marking Inflectional Affix: Evidence from Persian and Kurmanji Kurdish*. In: Karimi, S., Samiian, V., and Stilo, D. (Eds.). *Newcastle upon Tyne: Cambridge Scholars Publishing*: 339-361.

- Seraji, M. 2015. *Morphosyntactic Corpora and Tools for Persian*. Uppsala Universitet: Sweden.
- Shamsfard, M. 2011. Challenges and Open Problems in Persian Text processing. *Proceedings of LTC*.
- Shamsfard, M. and S., Noferesti. 2014. A Hybrid Algorithm for Recognizing the Position of Ezafe Constructions in Persian Texts. *International Journal of Interactive Multimedia and Artificial Intelligence* Vol. 2, N. 6:17-25.
- Tabaian, H. 1974. *Conjunction, Relativization and Complement in Persian*. PhD dissertation. University of Colorado.
- Tabatabaei, A. 2003. *Compound nouns and adjectives in Persian language*. Tehran: Iran University Press. [In Persian].
- Tabibzadeh, O. 2014. *Persian Grammar*. Tehran: Markaz. [In Persian].
- Tesnière, L. 1953. *Esquisse d'une Syntaxe structural*. Paris: Klincksieck.
- Toosarvandi, M. and C. van Urk. 2014. The syntax of nominal concord: what ezafe in Zazaki shows us. *In proceedings of NELS 43*, eds. Hsin-Lun Hung, Ethan Poole, and Amanda rsyling, Volume 2, 209-220. Amherst: GLSA.
- Yarahmadzahi, A. 2002. *The internal structure of noun phrases in Sarhad Balochi dialect based on X-bar theory*. MA thesis. University of Tehran. [In Persian].