

مدلسازی دانش تکواژشناختی زبان فارسی

به کمک گرامرهای پیوندی^۱

آرمین سجادی و محمد مهدی همایون پور
دانشکده‌ی کامپیوتر، دانشگاه صنعتی امیرکبیر
نویسنده‌ی عهده‌دار مکاتبات: آرمین سجادی

چکیده

ساختار وندی زبان فارسی کلمات بسیار پیچیده با ویژگی‌های متنوعی تولید می‌کند که طراحی یک واژگان کامل برای آن بسیار مشکل می‌باشد. اما با توجه به منظم بودن نسبی این ساختار، این مشکل به کمک تحلیل تکواژشناختی قابل رفع می‌باشد. در زبان فارسی، واج‌های همسایه عموماً تأثیرات زیادی بر یکدیگر ایجاد نمی‌کنند و همچنین از طرف دیگر، واج‌های متعدد می‌توانند بارها بر یک تکواژ اثر گذاشته و طبقه‌ی نحوی آن را عوض کنند. به همین دلیل گرامر پیوندی می‌تواند به عنوان یک صورت‌بندی^۲ قوی برای بازنمایی این خواص به‌کار رود و انتقال دانش را از روش‌های سنتی به محاسباتی با هزینه‌ی کمی امکان‌پذیر نماید.

ما با معرفی سیستمی که قادر به تجزیه‌ی کلمه به همه‌ی واج‌های تشکیل‌دهنده‌ی آن در زبان فارسی و انعکاس ویژگی‌های کلمه‌ی مورد نظر می‌باشد، تکواژشناسی زبان فارسی را به‌طور مفصل و از دیدگاه محاسباتی بررسی می‌کنیم. در این سیستم بر خلاف سیستم‌های موجود، همه واج‌های صرفی و اشتقاقی مورد پوشش قرار داده می‌شود که به کمک روش‌های معمولی، چون انتقال دهنده‌های با حالت محدود (FST) امکان‌پذیر نمی‌باشد. با توجه به اینکه گرامرهای پیوندی فاقد ویژگی هستند و ویژگی، خاصیت لازم هر تحلیل‌گر تکواژشناختی است، در این تحقیق روشی برای بازنمایی و استخراج ویژگی‌ها در این صورت‌بندی ارائه می‌شود. در این مطالعه بر خلاف تحقیق‌های مشابه، همه‌ی دانش سیستم ارائه شده است تا بتواند در تحقیقات مشابه دیگری نیز مورد استفاده قرار گیرد.

کلمات کلیدی: گرامر پیوندی، تحلیل تکواژشناختی، پردازش زبان طبیعی.

۱- مقدمه

با توجه به تعداد زیاد وندهایی که می‌تواند به یک کلمه در زبان فارسی متصل شود، واژگان به سرعت رشد کرده و ناکارا می‌شود. هر چند این اتصالات از قواعد زیرین ساختار زبان پیروی می‌کنند، اما محدودیتی در این زمینه وجود ندارد و انواع میانوند، پسوند و پیشوندهای اشتقاقی می‌توانند بر یک کلمه ظاهر شده و طبقه‌ی نحوی آن را تغییر دهند. بسیاری از ابهام‌های ساختاری نیز حاصل افزودن متوالی این وندها می‌باشد. برای مثال، می‌توان بیشتر از ده وضعیت مثال زد که واژگ -ی- می‌تواند به یک کلمه متصل شود و معنی و POS کلمه را عوض کند. تحلیل تکواژشناختی^۳ زبان فارسی

سابقه‌ی کوتاهی دارد، یکی از تحقیقات ارزشمند اخیر [1] می‌باشد که سیستمی برای ساختارهای تصریفی مبتنی بر انتقال دهنده‌های با حالت محدود (FST) و ویژگی‌های نوع‌دار ارائه داده است که تنها وندهای تصریفی^۴ را تحت پوشش قرار می‌دهد.

در این تحقیق خواهیم کوشید که ساختار تکواژشناختی زبان فارسی را از دیدگاه گرامرهای پیوندی به صورت مفصل بررسی کنیم. تحلیل‌گر معرفی شده قادر به تحلیل و تشخیص انواع ساختارهایی می‌باشد که از افزودن وندهای

² Formalism

³ Morphological Analysis

⁴ Inflectional

¹ Link Grammer

- **مسطح بودن**^۷. پیوندها هنگامی که از بالای کلمات وصل می‌شوند نباید هم‌دیگر را قطع کنند.
- **هم بندی**^۸. پیوندها همه‌ی کلمات را به هم وصل کنند (گراف حاصل هم بند باشد).
- **ترتیب**^۹: هنگامی که اتصال‌گرهای یک کلمه از چپ به راست پیمایش می‌شوند، کلماتی که به آن‌ها متصل شده‌اند از نزدیک به دور تغییر می‌یابند.
- **انحصار**^{۱۰}. هیچ دو پیوندی یک جفت کلمه را به هم متصل نمی‌کند.
- دیگر مفاهیم گرامر پیوندی که به آن‌ها رجوع می‌شود به‌طور خلاصه عبارتند از:
- **پیوند**^{۱۱}. پیوندها جفت کلمات را به هم مرتبط می‌کنند که به هریک اسمی اختصاص داده می‌شود.
- **اتصال‌گر**^{۱۲}. پیوندها از طریق اتصال‌گرهای کلمات آن‌ها متصل می‌شوند که از آن به عنوان ارضا کردن اتصال‌گر ذکر می‌شود. هر اتصال‌گر شامل یک نام پیوند به همراه + یا - می‌باشد، دو اتصال‌گر هنگامی به هم متصل می‌شوند که اسامی آن‌ها با هم مطابق بوده و اتصال‌گر + (-) در سمت چپ (راست) پیوند قرار گیرد.
- **ملزومات پیوندی**^{۱۳}. تعریف پیوندها برای یک اتصال‌گر به شکل فرمول‌های خوش فرمی است که توسط عمل‌گرهای گرامر پیوندی ساخته می‌شود که فهرست آن‌ها در (جدول ۱) آمده است.
- **تطابق پیوندی**^{۱۴}. در حالت کلی یک اتصال‌گر با یک یا بیشتر حرف بزرگ شروع می‌شود که نام آن را مشخص می‌کند و با دنباله‌ای از حروف کوچک ادامه می‌یابد که زیرنویس نامیده می‌شود. قاعده‌ی کلی تطابق بین دو اتصال‌گر را می‌توان به صورت زیر خلاصه کرد:
 دو اتصال‌گر $Ax_1...x_m$ و $By_1...y_n$ قابلیت اتصال‌گر دارند هرگاه (الف) A با B مطابقت کند و (ب) $x_1...x_n$ با $y_1...y_n$ مطابقت کند و $m \geq n$ یا $x_1...x_m$ با $y_1...y_m$ مطابقت کند و $m < n$.
- **تعریف**^{۱۵}. به مجموعه قواعد یک کلمه، تعریف آن گفته

تصریفی و هم‌چنین اشتقاقی^۱ حاصل می‌شود. این روش علاوه بر مزایای محاسباتی با توجه به شباهت بین گرامرهای پیوندی و شیوه‌ای که برای توصیف گرامر به‌صورت سنتی به کار می‌رود، توصیف‌کنندگی بیشتری دارد. هم‌چنین با توجه به بی‌قاعده بودن^۲ سیستم اشتقاقی استفاده از یک صورت‌بندی مستقل از متن ضروری می‌باشد.

در این تحقیق بر خلاف کارهای مشابه، همه‌ی قواعد و دانش‌های سیستم ارائه و توضیح داده شده است و مرور کاملی بر تکواژشناسی زبان فارسی انجام شده است. ارائه‌ی یک روش محاسباتی برای ساخت زبان فارسی، می‌تواند راه را برای استفاده از دیگر روش‌ها و ابزارها هموارتر کند. در [۲] روشی برای مدل‌سازی الگوهای دستوری زبان فارسی ارائه شده است. اما در این تحقیق بر روی بخش دیگری از زبان فارسی تأکید می‌باشد و آن ساختار خود کلمه است که از قواعدی متفاوتی پیروی می‌کند. مهم‌ترین تفاوت این نوع تحلیل در این است که خروجی آن طبیعتاً نه یک پیوندبندی، بلکه ویژگی‌های^۳ کلمه است. با توجه به این‌که گرامر پیوندی دارای زیرساختی برای پشتیبانی از ویژگی نیست، ما روشی را برای کد کردن این ویژگی‌ها در پیوند بندی و استخراج آن‌ها ارائه می‌کنیم. پارادایم‌های زبان‌شناسی این تحقیق مبتنی بر [۳] بوده و نمادگذاری به کار رفته نیز مطابق [1] صورت گرفته است. هم‌چنین تبدیل نوشتار فارسی به لاتین به کمک یک نگاشت یک‌به‌یک بین حروف صورت گرفته که تا حد زیادی شبیه به [1] می‌باشد.

۲- معرفی گرامر پیوندی

گرامر پیوندی به عنوان یکی از گرامرهای وابستگی توسط دانیل اسلیتور^۴ معرفی شد [4] و [5]. هر گرامر پیوندی شامل مجموعه‌ای از کلمات (نمادهای پایانی)^۵ است که هر یک دارای تعدادی ملزوم پیوندی هستند. هر جمله که توسط این گرامر مورد قبول قرار می‌گیرد، اگر بتوان بین کلمات یال‌هایی کشید به طوری که این‌ها ملزومات پیوندی کلمات را ارضا کنند و قواعد زیر (که فراقواعد^۶ نامیده می‌شوند) را رعایت کنند.

¹ Derivational

^۲ منظور از بی‌قاعده بودن (Irregularity)، عدم تعلق به دسته‌ی اول زبان‌ها در سلسله مراتب چامسکی است.

³ Feature

⁴ Daniel Sleator

⁵ Terminal

⁶ Meta Rules

⁷ Planarity

⁸ Connectivity

⁹ Ordering

¹⁰ Exclusion

¹¹ Link

¹² Connector

¹³ Linking Requirements

¹⁴ Link Matching

¹⁵ Definition

در این تحقیق معرفی می‌شود، در مجموع دارای حدود ۳۰۰ قاعده‌ی پیوندی می‌باشد که در مقایسه به کل واژگان زبان فارسی بسیار اندک است. از طرف دیگر این روش به دلیل مبتنی بودن بر یک صورت‌بندی مستقل از متن، توانایی بیشتری در پوشش ساخت‌های پیچیده‌ی زبان فارسی دارد و به صورت کارآیی قادر به استخراج POS کلمات است. اما مهم‌ترین دلیل استفاده از این روش، ارائه‌ی یک چارچوب کاملاً سازگار با مجموعه‌ی گرامر پیوندی و استفاده در سیستم‌هایی چون [۲] می‌باشد. در بخش قبل ذکر شد که مبانی نظری و زبان‌شناسانه‌ی گرامر پیوندی با دیگر صورت‌بندی‌های متداول، متفاوت است و استفاده از پارادایم‌های دیگر صورت‌بندی‌ها در یک سیستم مبتنی بر وابستگی، نشانه‌ی ضعف نظری رهیافت مزبور است. هدف ما رفع این مشکل و ارائه‌ی یک مجموعه‌ی یک‌پارچه‌ی مبتنی بر نظریه‌ی وابستگی می‌باشد.

تحلیل در این روش، شامل دو مرحله می‌باشد، در مرحله‌ی اول کلمه به تعدادی تکواژ تجزیه می‌شود و به تقطیع گر گرامر پیوندی داده می‌شود. در این حالت هر یک از تجزیه‌ها به صورت یک بازنمایی واژگی معتبر در نظر گرفته می‌شود. در فاز دوم از نتیجه‌ی تقطیع، دانش زبانی استخراج می‌شود که نیازمند یک پس‌پردازش می‌باشد. منظور از دانش زبانی ویژگی‌های کلمه می‌باشد. با توجه به عدم وجود ویژگی در گرامرهای پیوندی، در بخش بعد ادغام و استخراج ویژگی از یک پیوندبندی ذکر می‌شود.

۳-۱- ساختار ویژگی

ویژگی‌ها را می‌توان با توجه به پیوندهایی که واژک‌ها با سمت چپ‌ترین کلمه، یعنی کلمه‌ی پیش فرض LEFT-WALL تشکیل می‌دهند، شناسایی کرد. هر پیوند با دیوار چپ می‌تواند نشان دهنده‌ی یک یا بیشتر ویژگی باشد. برای مثال، زبان فارسی، شناسه‌ی فاعلی در سوم شخص مفرد افعال ندارد و در نتیجه پیوندی که چنین فعلی ممکن است تشکیل دهد یک پیوند Vspts است که بیان‌کننده‌ی

{POS=V, FORM=simple-past, PERSON=third, NUMBER=singular}

می‌باشد. خصوصیات واژگانی، نیازی به کد شدن ندارند چرا که به کمک وندها تغییر نمی‌کنند. در شکل ۱ می‌توان مفهوم توضیح داده شده را دید. در این شکل منظور از stem ریشه‌ی کلمه و aff، هر وندی می‌باشد.

می‌شود.

- **واژه‌نامه^۱**: به مجموعه‌ای که حاوی کلمات و تعاریف آن‌ها می‌باشد، واژه‌نامه گفته می‌شود که با تلقی متداول از آن تفاوت دارد.
- **زیرنویس کلمات^۲**: کلماتی که دارای دو تعریف مختلف باشند، باید با زیرنویس‌های مخالف جدا شوند. زیرنویس حروف کوچکی است که بعد از نقطه قرار می‌گیرد، مانند: *weekly.e*.

جدول ۱. عمل‌گرهای گرامر پیوندی

| عمل‌گر | کارکرد |
|---------------------|---|
| & | هر دو اتصال‌گرگر باید ارضا شوند. |
| or | یکی از اتصال‌گرگرها بایستی ارضا شود. |
| {exp} | عبارت اختیاری است. |
| @ (unary) | یک یا تعداد بیشتری از این اتصال‌گرگر می‌تواند ارضا شود. |
| <MAC_NAME> : exp | عبارت کوتاه <MAC_NAME> می‌تواند به جای عبارت طولانی‌تر exp به کار رود که به آن ماکرو گفته می‌شود. |

۳- استفاده از گرامر پیوندی جهت تحلیل تکواژشناسی

مهم‌ترین روش مورد استفاده برای تحلیل تکواژشناختی استفاده از انتقال دهنده‌های با حالات محدود^۳ (FST) می‌باشد [6]. این روش دارای محدودیت‌های زیادی است. یکی از مهم‌ترین آن‌ها این است که کل واژگان باید به صورت FST، کد شود [7] که هم سرباره زیادی در طراحی دارد و هم سرعت پردازش آن کند می‌شود. از طرف دیگر FST به صورت مستقیم نمی‌تواند POS کلمه را استخراج کند، که این روش با افزودن قواعد همسان‌سازی^۴ باید حل شود [8]. ساختار آوایی زبان فارسی عموماً از تأثیر متقابل واژک‌های کنار هم، بر یکدیگر جلوگیری می‌کند و در نتیجه تغییر و تبدیل واژک‌ها کمتر رخ می‌دهد، به همین دلیل استفاده از یک روش مبتنی بر قاعده، مانند گرامر پیوندی از کد کردن کلّ واژگان در یک FST، هزینه‌ی کمتری دارد و می‌تواند مبتنی بر قواعد انجام گیرد. برای مثال سیستمی که

¹ Dictionary

² Subscripts

³ Finite State Transducers

⁴ Unification

دوم (۴-۲) تمرکز را بر روی ساختارهای غیرفعلی قرار می‌دهیم.

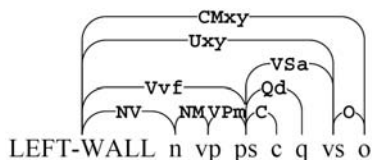
۴- تکواژشناسی فارسی

ما در حالت کلی دو نوع ساختار را در نظر می‌گیریم، در بخش ۴-۱ ساختارهای فعلی را بررسی می‌کنیم و در بخش

جدول ۲. چند مثال از ساختار کلی کلمه فعل

| n | vp | ps | c | q | vs | o | Phonetics |
|---|----------------|-----------------|---------------|-----------------|--------------|---|--|
| n | my my my | pr xnd gf | an an - | ayd ayd t | m yd - | S | [nemiaranidam] [mixandabidaš] [migofi] |

استراتژی پیوندبندی در حالت کلی در (شکل ۲) قابل مشاهده است.



شکل ۲. ساختار کلی یک پیوند بندی برای کلمه فعل

در بخش‌های بعد هر یک از اجزای فوق بررسی مختصری می‌شوند و تعاریف پیوندی هر یک ارائه می‌گردد.

عنصر منفی کننده

این واژگ، به اولین واژگ فعل از طریق یک پیوند NM متصل شده و فعل را منفی می‌کند. ویژگی Negative با ایجاد یک پیوند NV با LEFT-WALL شناسایی می‌شود.

n : NM+ & NV-;

پیشوند

این طبقه، شامل دو عضو می‌باشد -my و -b که یک پیوند VP آن‌ها را به فعل متصل می‌کند. توجه شود که در تعاریف زیر، -b و عنصر منفی با هم روی یک فعل ظاهر نمی‌شوند.

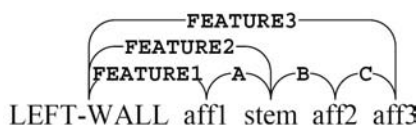
my : VPm+ & {NM-};
b.vp:VPb;

واژه‌بست‌های فاعلی و مفعولی

با توجه به شخص و عدد، شش واژه‌بست مختلف وجود دارد که در (جدول ۳) آمده است.

جدول ۳. شناسه‌های فعلی

| person, number | singular | plural |
|----------------|----------|--------|
| first | m | im |
| Second | y/Φ | yd |
| Third | d/Φ | nd |



شکل ۱. روش بازنمایی ویژگی‌ها در گرامر پیوندی

۴-۱. ساختارهای فعلی

ساختارهای فعلی، یکی از معمول‌ترین ساختارهای واژه‌ای است که بدون وجود یک تحلیل‌گر تکواژشناسی، مدخل‌های زیادی در واژگان به وجود می‌آورد. ما در این فصلی خواهیم کوشید، مجموعه قواعدی ارائه دهیم که بتوان به کمک آن‌ها، ساخت‌واژه‌ی افعال را تشخیص داد. باید اشاره کنیم که ما، بین کلمه‌ی فعل و عبارت فعلی تفاوت قائل می‌شویم که ممکن است شامل یک یا چند کلمه‌ی فعلی باشد. در نتیجه چیزی که در این بخش تحلیل می‌شود و از این به بعد برای اختصار فعل نامیده می‌شود، هسته‌ی این ساختار است.

۴-۱-۱. ساختار فعل

یک فعل شامل هفت عنصر ساختاری، عنصر منفی کننده (n)، پیشوند (vp)، ستاک حال (ps)، عنصر سببی ساز (c)، پسوند گذشته (q)، پسوندهای فعلی (vs) و واژه‌بست^۱ (o) می‌باشد که ۵ تای آن‌ها اختیاری است. (واژه‌بست طبق تعریف، جزء کلمه به‌شمار نمی‌آید؛ اما برای سادگی تحلیل در اینجا این‌گونه در نظر گرفته شده است). ساختار یک فعل را می‌توان به کمک {n}{vp}{ps}{c}{q}{vs}{o} نشان داد که در آن واژگ‌های بین { } به معنی اختیاری بودن آن‌ها می‌باشد. در (۰) تعدادی مثال مشاهده می‌شود.

¹ Clitic

a و b زیرنویس‌هایی هستند که این ویژگی را نشان می‌دهند. از آنجا که این خاصیت در ساختارهای مشابه دیگری نیز وجود دارد؛ در ادامه، منظور از اسم نوع B یا فعل نوع B آن‌هایی است که مختوم به a یا v و منظور از اسم نوع A یا فعل نوع A، دیگر انواع می‌باشد. گوناگونی‌هایی از این شناسه‌ها وجود دارد که برای ساخت افعال نقلی به‌کار می‌روند (رفته‌ام) که در اینجا بررسی نمی‌شوند؛ چراکه می‌توان به آن‌ها به‌عنوان واژک‌های غیرمقید نگاه کرد.

| | | |
|-----|------------------|---------------------|
| m | : Oafs- & CMfs-; | |
| t | : Oass- & CMss-; | پسوند گذشته |
| S | : Oats- & CMts-; | |
| man | : Oafp- & CMfp-; | چهار گوناگونی از |
| tan | : Oasp- & CMsp-; | این پسوند وجود دارد |
| San | : Oatp- & CMtp-; | (-t, -d, -yd, -ad) |

که با توجه به نظام آوایی، هر یک برای گروهی از افعال به‌کار می‌روند و همه‌ی آن‌ها از طریق Q به کلمه‌ی فعل وصل می‌شوند.

| | |
|-----|---------|
| t | : Qt- |
| d | : Qd- |
| yd | : Qyd |
| ad | : Qad- |
| yyd | : Qyyd- |

این پیوندها، فقط با افعال باقاعده به‌کار می‌روند. در کلمات زیر برای هر یک از آن‌ها مثالی ارائه شده است و صورت گذشته‌ی افعال بی‌قاعده باید مستقیماً در واژه‌نامه ذکر شود.

| | |
|--------|------------------|
| kš+t | [/košt/, killed] |
| xr+d | [/xord/, ate] |
| dv+yd | [/david/, ran] |
| aft+ad | [/?oftad/, fell] |
| sa+iid | [/sa?id/,] |

عنصر سببی‌ساز

تنها عضو این طبقه -an می‌باشد که از طریق پیوند C به فعل متصل می‌شود.

an.s : C;-

۴-۱-۲. فرم‌های مختلف فعلی

در زبان فارسی، ده صورت مختلف یک فعل با توجه به وجود یا عدم وجود واژک‌های (جدول ۵) وجود دارد که همه‌ی آن‌ها بررسی می‌شوند. باید اشاره کرد که برخی افعال از این قواعد پیروی نمی‌کنند؛ مانند افعال بودن و

این واژک‌ها دارای ویژگی شخص و عدد هستند که هر دو با یک پیوند CMxy مشخص می‌شوند؛ که x, y به ترتیب بیان‌گر آن‌ها می‌باشند و یک پیوند O آن‌ها را به واژک بعدی متصل می‌کند. در مورد افعال منتهی به a یا v استثنایی وجود دارد که در آن‌ها یک y مابین ریشه و پسوند قرار می‌گیرد؛ یا به عبارتی به صورت -ym, -yt, و تغییر می‌یابند. تعریف هر دو مشابه است، فقط با این تفاوت که زیرنویس‌های a و b، آن‌ها را از هم متمایز می‌کند. تعاریف گروه اول در زیر آمده است:

و به همین ترتیب برای گروه دوم داریم.

im : im.cm: Obfs- & CMfs-;
...

پسوندهای فعلی

این گروه، شامل سه زیرطبقه است که دوتای آن‌ها دارای یک عضو و دیگری دارای شش عضو می‌باشد و همه‌ی آن‌ها از طریق پیوند VS به فعل متصل می‌شوند.

(۱) -n: واژک مصدری (مانند رفتن)

(۲) -h: واژک صفت مفعولی (رفته)

(۳) شناسه‌های فعل (جدول ۴)

جدول ۴. شناسه‌های فعلی

| person, number | singular | plural |
|----------------|----------|--------|
| first | m | im |
| Second | y/Φ | yd |
| Third | d/Φ | nd |

شخص و عدد به کمک پیوند Uxy تنظیم می‌شود که x و y مانند قبل به ترتیب نشان‌دهنده‌ی آن‌ها می‌باشند. در تعاریف زیر می‌توان توجه کرد که پیوند O قواعد اتصال‌گر را مشخص می‌کند. برای مثال از تعریف اول، این مسأله استنباط می‌شود که یک شناسه‌ی اول شخص مفرد، فقط به یک واژه‌بست دوم یا سوم شخص متصل می‌شود.

| | |
|----|--------------------------------|
| m | : VSa & Ufs- & {Oas+ or Oat+}; |
| ym | : VSa & Ufp- & {Oas+ or Oat+}; |
| y | : VSa & Uss- & {Oaf+ or Oat+}; |
| yd | : VSa & Usp- & {Oaf+ or Oat+}; |
| d | : VStsa- & Uts- & {Oa+}; |
| nd | : VSa & Utp- & {Oa+}; |

در اینجا استثنایی وجود دارد و آن این‌که در افعال مختوم به a یا v، یک y بین ریشه و پسوند قرار می‌گیرد. مانند واژه‌بست، تعاریف مشابه زیر را داریم:

ym : VSb- & Ufs- & {Oas+ or Oat+};
...

دوم شخص مفرد می‌باشد که در آن پیشوند، می‌تواند حذف و به فعل امری تبدیل شود.

<IP-A> : (VPb- or NM-) & Vip- & {C+} & {Oaf+ or Oat+};

گذشته‌ی ساده و استمراری (CP و SP)

برای پوشش این صورت، افعال باقاعده و بی‌قاعده به ترتیب زیر تعریف می‌شوند:

افعال با قاعده

<SCPPR> : (VPm- & Vcp-) or ({NM-} & Vsp-);

<SCPPR>، ماکروبی برای توصیف پیش-واژک‌های عمومی است. وجود پیوند VPm و Vcp با LEFT-WALL استمراری بودن را مشخص می‌کند. تنها استثنا در مورد سوم شخص مفرد است که هیچ پسوندی ندارد و پیوند V باید هم ساخت و هم شخص و تعداد را نشان دهد. (مثلاً POS=verb, FORM=Continuous Past, } بیان‌گر Vcpts {PERSON=third, NUMBER=singular می‌باشد). سوم شخص مفرد به کمک <SCPPRk> تعریف می‌شود.

داشتن در زمان حال که به صورت متفاوتی صرف می‌شود. هم‌چنین، صورت‌های مختلف فعل، نام‌گذاری شده‌اند و می‌شد به جای اسامی حال و غیره از نمادها استفاده کرد. در نتیجه دلالت معنایی کلمات مورد توجه نیست. در ساختارهای زیر، صورت فعلی در پیوند Vvf با دیوار سمت چپ قرار داده می‌شوند. یک نکته‌ی بسیار مهم این‌که، این قواعد برای کلمه‌ی فعل نوشته می‌شود.

حال

با توجه به (جدول ۵)، صورت حال فعل می‌تواند به صورت زیر تعریف شود.

<PT-A> : {(VPm- or NM-)} & Vpt- & {C+} & (VSa+ or VStsa+);

تعاریف مناسبی برای افعال B نیز وجود دارد.

امری و التزامی

<SB-A> : (VPb- or NM-) & Vsb- & {C+} & (VSa+ or VStsa+);

مانند زمان حال، این ماکرو برای پوشش دادن دیگر اشخاص و تعداد، باید بسط داده شود. تنها استثنا در مورد

جدول ۵. صورت‌های مختلف فعلی

| Verb Form | Abbr. | Particles | | | | | Example |
|----------------------------|-------|------------|----|---|---|----|----------------------|
| | | vp | Ps | c | q | vs | |
| Present | PT | my | + | ± | - | + | myxrm/mixranm |
| Present | PT | - | + | - | - | + | darm |
| Simple Past | SP | - | + | ± | + | + | xvrmd/xvrandm |
| Continuous Past | CP | my | + | ± | + | + | myxvrmd/myxvrandm |
| Subjunctive | SB | be/ bi/ bo | + | ± | - | + | bxvrm/bxvranm |
| Imperative | IP | be/ bi/ bo | + | ± | - | + | bxvr, bxvran |
| Infinitive | IF | - | + | ± | + | an | xvrnd/ xvrandn |
| Short Infinitive | SI | - | + | ± | + | - | xvrd/xvrand |
| Past Participle | PP | - | + | ± | + | h | xvrndh/ xvrandh |
| Continuous Past Participle | CP | my | + | ± | + | e | myixvrndh/ myxvrandh |

سببی

در نهایت، تعریف زیر را بر افعال گذشته‌ی ساده/ استمراری داریم؛ (یعنی SCPR)، RT بیان‌گر باقاعده بودن و داشتن واژک t- به‌عنوان نشانه‌ی ستاک گذشته می‌باشد و طبیعتاً برای دیگر فعل‌ها، تعاریف مشابهی وجود دارد.

<SCPPRk> : (VPm- & Vcpts-) or ({NM-} & Vsppts-);

این ماکرو، برای پوشش حالت سببی به‌کار می‌رود، باید توجه شود که در هنگام حضور این عنصر، همه‌ی افعال باقاعده از واژک d- برای نشان‌گر گذشته استفاده می‌کنند.

$$\{pr\}^+ s\{su\}^+ \{c\}\{pl\}\{io \cup ez\}$$

که در آن { } نشانه‌ی انتخابی بودن و + به معنی تکرار است. s ریشه و دیگر نشانه‌ها به ترتیب برابر معانی زیر به کار رفته‌اند:

pr: پیشوند

su: پیشوندها

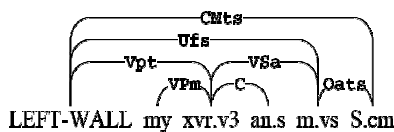
c: واژگ مقایسه‌ای

pl: واژگ‌های جمع‌ساز

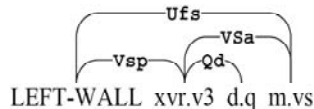
io: واژه‌بست‌های نکره‌ساز یا مفعولی

ez: واژه‌بست اضافه

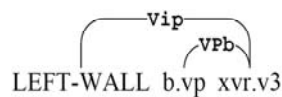
باید اشاره شود که رابطه‌ی فوق یک طرح اولیه است؛ چون همان‌طور که قبلاً گفته شد، اضافه شدن واژگ‌ها تابع قواعدی است که از ساختار زبان و POS تکواژها حاصل می‌شود؛ مثلاً یک واژگ مقایسه‌ای نمی‌تواند به یک حرف اضافه متصل شود. از آن‌جایی که صفات می‌توانند جای



(الف)



(ب)



(پ)



(ت)

شکل ۳. چهار مثال برای تکواژشناسی فعل

(الف) می‌خورم (ب) خوردم (پ) بخور (ت) خوردن

<SCPRT> : (<SCPPR> & ((Qt+ or <SCPC>) & VSa+)) or (<SCPPRk> & ((Qt+ or <SCPC>) & (Oas+ or Oat+)));

افعال بی قاعده

همان‌گونه که قبلاً ذکر شد، ریشه‌ی افعال بی قاعده باید با تعریف زیر به واژه‌نامه افزوده شود.

<SCPI> : (<SCPPR> & VSa+) or (({VPm-} & Vmmts-) & {(Oas+ or Oat+)});

مصدر و اسم مفعول

در تعاریف زیر، <YPS>، ماکروبی برای تعریف همه‌ی واژگ‌های انتهایی می‌باشد. <IF> ماکروبی است که رفتار مصدر را توصیف می‌کند؛ زیرا مصدر می‌تواند وندهای مختلف دیگری نیز اختیار کند. CP و PP پیوندهایی هستند با دیوار تشکیل شده و به استمراری یا مفعولی بودن اشاره دارند و <SI> ماکروبی است که مصدر کوتاه را تعریف می‌کند. این ماکروها، در بخش ۴-۲-۶ و (جدول ۶) بررسی شده است. ماکروی اصلی برای افعال باقاعده، ماکروی <YRT> می‌باشد.

<YSU> : (VS_n+ & <IF>) or (VSe+ & PP-) or (VPm- & CP- & VSe+) or <SI>;

<YRT> : {NM-} & (Qt+ or <SCPC>) & <YSU>;

و برای افعال بی‌قاعده

<YI> : {NM-} & <YPS>;

۴-۱-۳. نتایج و مثال‌ها

ما در اینجا چهار مثال برای نشان دادن مفهوم پیوندبندی‌ها نشان می‌دهیم. در مثال اول [شکل ۳ (الف)] واژه‌ی می‌خورم یک فعل حال (Vpt) سوم شخص مفرد (Ufs) با یک واژه‌بست مفعولی سوم شخص (Cmts) آمده است. در مثال دوم [شکل ۳ (ب)]، خوردم یک فعل گذشته‌ی ساده‌ی اول شخص مفرد (Vsp) و مثال سوم، [شکل ۳ (پ)]، بخور، یک فعل امری (Vip) و مثال چهار [شکل ۳ (ت)]، خوردن، یک مصدر (IF) است.

۴-۲. ساختارهای غیرفعلی

از یک دیدگاه عمومی، ساختار یک کلمه می‌تواند به صورت زیر نمایش داده شود:

نشانه‌ی نکره و y برای واژه‌بست‌های مفعولی xy (شخص و عدد) می‌باشد.

y.f: La- & Fy-;
m.f: La- & Ffs-;
...

۴-۲-۲. واژه‌بست اضافه

اضافه، واژگی است که به ریشه متصل می‌شود؛ اما در نوشتار ظاهر نمی‌شود؛ جز در مورد اسامی نوع B. در این حالت اضافه به‌صورت یک y نوشته می‌شود که به کمک پیوند K به کلمه متصل می‌شود و ویژگی ez را تنظیم می‌کند.

y.ez: K- & EZ-;

۴-۲-۳. نشانه‌ی جمع

انواع گوناگونی از نشانه‌ی جمع وجود دارد که به اسم یا هر طبقه‌ای که بتواند جایگاه اسم را اختیار کند، متصل می‌شود. همه‌ی آن‌ها تعریف مشابهی دارند.

x.p : Wx- & P- & {K+ or Ly+};

که x در آن نشان دهنده‌ی خود واژک و y نوع تکواژ می‌باشد (a یا b).

تعریف نشان می‌دهد که بعد از آن می‌تواند اضافه یا واژه‌بست مفعولی ظاهر شود. (پیوند K برای کلمه‌هایی که به a یا v ختم نمی‌شوند، حذف می‌شود). همه‌ی آن‌ها به کمک یک پیوند Wx (نشان دهنده‌ی واژک است) به کلمه متصل شده و از طریق پیوند P که نشان دهنده‌ی جمع بودن است به LEFT-WALL متصل می‌شوند. در مورد هر یک از این واژک‌ها در زیر توضیح مختصری می‌آید.

ها (ha)

این واژک، عمومی‌ترین نشانه‌ی جمع است که به‌صورت زیر تعریف می‌شود:

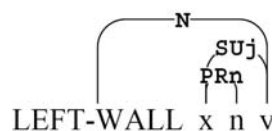
ha.p: Wha- & P- & {K+ or Lb+};

ان (an)

این واژک نیز، واژک عمومی به شمار می‌آید. اگر کلمه به v یا y ختم شود، به $-yan$ تبدیل می‌شود (مثال: زیبايان یا مهربويان)، هر چند برای همه‌ی آن‌ها صادق نیست (گیسوان). کاربرد آن‌ها بیشتر به محیط آوایی کلمه بستگی دارد.

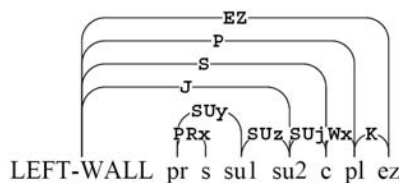
an.p: Wan- & P- & {La+};

اسامی را بگیرند، دارای ملزومات پیوندی مشابهی هستند، تنها با این تفاوت که از یک پیوند J برای صفات و یک پیوند N برای اسامی استفاده می‌شود. قواعدی که افزایش وندها باید از آن تبعیت کند سازگاری با POS و دیگر مشخصات آوایی کلمه می‌باشد. سازو کار که در این تحقیق برای پوشش این خاصیت ارائه شده است به این ترتیب است که POS کلمه‌ی جدید به کمک وند متصل شده، بازنمایی می‌شود. برای مثال فرض کنید x پیشوندی باشد که روی اسامی ظاهر شده و آن‌ها را تبدیل به صفت می‌کند، y پسوندی است که روی صفات ظاهر شده و آن‌ها را به اسم تبدیل می‌کند و n یک اسم باشد، در نتیجه کلمه‌ی xny به‌صورت شکل ۴ تقطیع می‌شود.



شکل ۴. پیوندها و پسوندهایی که روی کلمه ظاهر شده و POS آن را عوض می‌کنند.

طبیعتاً در این گونه موارد، ابهام‌های زیادی رخ می‌دهد که عموماً به‌دلیل کارکردهای مختلف وندها است. برای مثال بیشتر از ده حالت وجود دارد که $-y$ می‌تواند روی کلمه ظاهر شده و POS و معنی آن‌ها را عوض می‌کند که برخی از آن‌ها می‌تواند با تحلیل نحوی حل شود. روشی که تکواژهای معرفی شده به هم پیوند زده می‌شوند، و هم‌چنین نحوه‌ی بازنمایی ویژگی‌ها، در (شکل ۵) آمده است.



شکل ۵. طرح پیوند بعدی برای ساختار کلی یک کلمه

۴-۲-۱. نشانه‌ی نکره و واژه‌بست مفعولی

زبان فارسی دارای یک نشانه‌ی نکره $-l$ است که هنگام اتصال گر به اسامی نوع B به $-yy$ تغییر می‌یابد. این واژک از طریق یک پیوند F به دیوار چپ متصل می‌شود. واژه‌بست‌های مفعولی در بخش ۴ بررسی شدند. هر دو با یک پیوند La (یا Lb برای اسامی نوع B) به کلمه متصل می‌شوند و یک پیوند Fv با دیوار می‌سازد که v برای

برای صفات نیز تقسیم‌بندی مشابهی وجود دارد. از این به بعد زیرنویس‌ها (a, b, \dots, e) بیان‌کننده‌ی همین ویژگی می‌باشند.

۴-۲-۴. واژک‌های مقایسه

این واژک ($-tr$ برای تفصیلی و $-tryn$ برای برترین) به کمک پیوند R به صفت متصل شده و ویژگی comparison را با ایجاد پیوند S با دیوار، تنظیم می‌کند.

trjn.jsup yn.jsup tr.jsup Rj- & S- & ({Wha+ or Wan+ or La+});

۴-۲-۵. ریشه

ریشه می‌تواند به هر یک از طبقات نحوی متعلق باشد که هر یک با ماکروبی خاص تعریف می‌شود؛ برای مثال اسامی به صورت زیر تعریف می‌شوند:

<N> : SUN+ or IN+ or IN- or (N- & ({Wha+} or {Wjat+}));

این ماکرو به این معنی است که یک اسم می‌تواند به پیشوندها، میانوندها و نشانه‌ی جمع عمومی $-ha$ متصل یا با یک N- به دیوار متصل شود که نشان دهنده‌ی POS آن می‌باشد. آنچه ذکر شد، یک تعریف عمومی است که همه‌ی اسامی به نحوی آن را علاوه بر تعریف خاص خود دارند. ماکروهای عمومی برای POS های مختلف در (۰) آمده است. صفات دارای تعریف مشابهی می‌باشند (جز یک پیوند R با نشانه‌ی تفصیلی). توجه کنید که آخرین پیوند، همیشه پیوندی است که با دیوار تشکیل می‌شود. دیگر ماکروها در (جدول ۶) ذکر شده‌اند.

۴-۲-۶. وندها

همه‌ی پیشوندها یک پیوند PRxy دارند که x، POS و y نوع کلمه (a, b, \dots, e) را مشخص می‌کند. دلیل این که این گروه‌بندی به پیشوندها هم اعمال می‌شود، این است که همان‌گونه قبلاً نیز ذکر شد، POS و دیگر ویژگی‌های تأثیرگذار کلمه، بعد از اتصال گروند، به آن متصل می‌شوند.

ات (at)

از لحاظ دستوری این واژک فقط باید با کلمه‌های عربی به کار می‌رود (مثال: اجتماعات) هر چند با برخی کلمات فارسی نیز به کار می‌رود (مثال: گزارشات). برخی کلمات عربی نیز، به اشتباه با این واژک جمع بسته می‌شوند (مثال: خطرات) و حتی بعضی از جمع‌های مکسر نیز با این واژک جمع بسته می‌شوند (مثال: جواهرات). به علت عمومیت این موارد، نوع تعاریف، همه‌ی آن‌ها را تحت پوشش قرار می‌دهد.

at.p : Wat- & (PRna- or P-) & {La+};

ین (yn)

این واژک، برای برخی کلمات عربی به کار می‌رود (مثال: مسلمین).

yn.p : Wyn- & (PRna- or P-) & {La+};

ون (vn)

این واژک خاص کلمات عربی است که دارای $-y$ صرفی در انتها می‌باشد (مثال: اجتماعیون). SUNPpu پیوندی است که yn- را به $-y$ متصل می‌کند.

vn.p : SUNPpu- & P- & {La+};

جات (Jat)

این واژک معمولاً دلالت بر نوع دارد؛ مانند شیرینی‌جات یا طلاجات که از طریق پیوند Wjat به اسم متصل می‌شود.

Jat.p : Wjat- & P- & {K+ or Lb+};

با توجه به آنچه گفته شد، صفات و اسامی را به پنج گروه متفاوت تقسیم می‌کنیم:

<N-A> : اسامی نوع A

<N-B> : اسامی نوع B

<N-C> : اسامی نوع B که $-an$ نشانه‌ی جمع آن‌ها است.

<N-D> : کلمات عربی که نشانه‌ی جمع $-at$ اختیار می‌کنند.

<N-E> : کلمات عربی که با $-in$ جمع بسته می‌شوند.

جدول ۶. ماکروهای عمومی برای POS های مختلف

| Category | Macro |
|--------------------------|--|
| Nouns* | <N-x> : <N> or PRnx- or Ynx+ or (N- & (Wan+ or Lx+)); |
| Adjectives* | <J-x> <J> or PRjx- or Yjx+ or (J- & (Rj+ or Wan+ or Lx+)); |
| Short Infinitive | <SI> : PRsi- or SUsi+ or SI-; |
| Infinitive | <IF> : PRif- or SUIf+ or IF-; |
| Present Stem | <PS-A> : PRpsa- or SUpS+ + or PS-; |
| Adverb | <A> : PRa- or SUa+ or A-; |
| Infinitive Result** | <IR> : PRh- or SUh+ or PRna- or SUn+ or IR-; |
| Demonstrative Adjectives | <D> : PRd- or Pod+ or D-; |
| Cardinal Numbers | <CN> : PRen- or SUCn+ or CN-; |
| Ordinal Numbers | <ON> : PRtn- or SUn+ or ON-; |
| Prepositions | <PREP> : PRpre- or SUpre+ or PREP-; |
| Augment | <AP> : PRap- or SUap+ or AP-; |

* نوع پنج‌گانه‌ی اسم یا صفت می‌باشد.

** مانند خوبی، که جایگاه اسم را اشغال می‌کنند، در نتیجه در تعریف آن‌ها پیوندهای اسم نیز ذکر شده‌است.

که در آن x، POS کلمه است. حرف A برای پیشوندها، Pre (+) و SUFF (-) برای پسوندها می‌باشد. Y نیز POS کلمه‌ی حاصل از افزودن وندها را مشخص می‌کند که توسط ماکروی <Y>، (+) تعریف شده‌است. بیشتر وندهای زبان فارسی در (جداول ۷ تا ۹) خلاصه شده‌اند. برای مثال در این جدول منظور از **Adjective** → **Noun**، وندهایی است که بر روی یک اسم ظاهر شده و آن‌ها را به صفت تبدیل می‌کنند. بیشتر اطلاعات مربوط به وندها از [۹] استخراج شده است.

پسوندها به کمک پیوند SUx (جز در مورد آن‌هایی که مختوم به y هستند) به کلمه متصل می‌شوند. دلیل آن این است که در این حالت با توجه به کلمه‌ی ماقبل، تغییر شکل می‌دهند و به yy تبدیل می‌شوند در نتیجه در این مورد، از یک پیوند Yxy استفاده می‌کنیم که x و y مانند قبل می‌باشند. برای میان‌وندها از پیوند IN برای اتصال‌گر استفاده می‌کنیم، تعریف عمومی وندها به صورت زیر است:

<A-XY> Ax ± & <Y>

جدول ۷. پیشوندهای فارسی و ملزومات پیوندی آن‌ها

| Noun → Noun | | | | | |
|---|------------------|-----|-------------------|-----|-----------------|
| <Pre-NN>* : PRna+ & <N-A> or (PRnb+ & <N-B>) or PRnc+ & <N-C>) or (PRnd+ & <N-D>) or (PRne+ & <N-E>); | | | | | |
| Pre | Example | Pre | Example | Pre | Example |
| b | bdst /bedast/ | hm | Hmkar /hamkâr/ | va | vaknS /vâkoneš/ |
| vr | vrdst /vardast/ | | | | |
| Noun → Adjective | | | | | |
| <Pre-NJ> : (PRna+ & <J-A>) or (PRnb+ & <J-B>) or (PRnc+ & <J-C>) or (PRnd+ & <J-D>) or (PRne+ & <J-E>); | | | | | |
| Pre | Example | Pre | Example | Pre | Example |
| b | bxd /bexrad/ | by | bygnah /bigonâh/ | ba | bahnR /bâhonar/ |
| frv | frvtn /forutan/ | br | brqrar /barğarâr/ | n | nstvh /nastuh/ |
| na | nakam /nâkâm/ | - | - | - | - |
| Noun → adverb | | | | | |
| <Pre-NA> : (PRna+ or PRnb+ or PRnc+ or PRnd+ or PRne+) & <A>; | | | | | |
| Pre | Example | Pre | Example | Pre | Example |
| b | bsrat /besor?at/ | - | - | - | - |
| Adjective → Adjective | | | | | |

| <Pre-JJ> : (PRja+ & <J-A>) or (PRjb+ & <J-B>) or (PRjc+ & <J-C>) or (PRjd+ & <J-D>) or (PRje+ & <J-E>); | | | | | |
|---|---------------------|-----|-------------------|-----|---------------------|
| Pre | Example | Pre | Example | Pre | Example |
| n | nsnjydh /nasanjidh/ | na | nadrst /nâdorost/ | vr | vrpnydh /varparidh/ |
| Indefinite Result → Adverb | | | | | |
| <Pre-IRA> : PRh+ & <A>; | | | | | |
| Pre | Example | Pre | Example | Pre | Example |
| b | bxvby /bexubi/ | | | | |
| Short Indefinite → Noun | | | | | |
| <Pre-SIJ> : PRsi+ & <J-A>; | | | | | |
| Pre | Example | Pre | Example | Pre | Example |
| baz | bazdaSt /bâzdâst/ | frv | frvrft /foruft/ | br | brxrd /barxord/ |
| va | vaknS /vâkoneš/ | dr | dramd /darâmad/ | vr | vrSkst /varšekast/ |
| Present Stem → Noun | | | | | |
| <Pre-PSN> : (PRpsa+ & <N-A>) or (PRpsb+ & <N-B>); | | | | | |
| Pre | Example | Pre | Example | Pre | Example |
| baz | bazprs /bâzprs/ | frv | frvkS /forukeš/ | br | brCsp /barčasp/ |
| va | vadar /vâdâr/ | dr | drgyr /dargir/ | vr | vrandaz /varandaz/ |
| fra | fraxor /faraxor/ | | | | |
| Short Indefinite → Adjective | | | | | |
| <Pre-SIJ > : PRmk+ & <J-A>; | | | | | |
| Pre | Example | Pre | Example | Pre | Example |
| by | byryxt /birixt/ | n | ndyd /nadid/ | na | naSayst /našâyest/ |
| hm | hmzad /hamzâd/ | | | | |
| Present Stem → Adjective | | | | | |
| <Pre-PSJ> : (PRpsa+ & <J-A>) or (PRpsb+ & <J-B>); | | | | | |
| Pre | Example | Pre | Example | Pre | Example |
| by | bytab /bytâb/ | na | naSnas /nâsenâs/ | n | nfhm /nafahm/ |
| va | vagyr /vâgir/ | by | bytab /bitâb/ | hm | hmgra /hamgara/ |

زیر نویس های (a,...,e) نوع اسم یا صفت را مشخص می کنند.

| Demonstrative Adjective → Demonstrative Adjective | | | | | |
|--|------------------|-----|---------|-----|---------|
| <Pre-DD > : PRd+ & <D>; | | | | | |
| Pre | Example | Pre | Example | Pre | Example |
| hm | hman /hamân/ | | | | |
| Ambiguous Pronoun → Ambiguous Pronoun | | | | | |
| <Pre-APAP > : PRap+ & <AP>; | | | | | |
| Pre | Example | Pre | Example | Pre | Example |
| hm | hmdygr /hmdigar/ | | | | |
| Ambiguous Pronoun → Adjective | | | | | |
| <Pre-APJ > : PRAP+ & <J-A>; | | | | | |
| Pre | Example | Pre | Example | Pre | Example |
| dr | drhm /darham/ | | | | |
| Suffixe → Noun | | | | | |
| <Pre-SUN > : PRsu+ & <N-A>; | | | | | |
| Pre | Example | Pre | Example | Pre | Example |
| hm | hmsan /hamsân/ | | | | |
| Preposition → Demonstrative Adjective | | | | | |
| <Pre-PRED > : PRpre+ & <D>; | | | | | |
| Pre | Example | Pre | Example | Pre | Example |
| hm | hmCvn /hamčon/ | | | | |

جدول ۸. پسوندهای زبان فارسی و ملزومات پیوندی آنها

| Nouns → Nouns | | | | | |
|------------------------------|---------------------|-----|----------------------|------|-----------------------|
| <Suff-NN-A>* : POn- & <N-A>; | | | | | |
| Suf | Example | Suf | Example | Suf | Example |
| an | xauran /xavarân/ | gan | mhrġan /mehregân/ | anh | šbhanh /sobhâne/ |
| ganh | jdaganh /jodâġâne/ | C | ktabCh /ketâbče/ | gany | xdaygany /xodayegâni/ |
| dan | nmkdan /namakdân/ | gah | tvqfġah /tavaġofġâh/ | dys | tndys /tandis/ |
| gr | msggr /mesgr/ | zar | Cmzgar /čamanzâr/ | lax | snglax /sanglâx/ |
| sar | kvhsar /kuhsâr/ | var | bhStvar /beheštvar/ | tan | bvstan /bustân/ |
| varh | mahvarh /mâhvare/ | k | mrđk /mardak/ | h | dndanh /dandane/ |
| Noun → Adjective | | | | | |
| <Suff-NJ-A> : SUn- & <J-A>; | | | | | |
| Suf | Example | Suf | Example | Suf | Example |
| asa | nhġnasa /nahangâsâ/ | nak | xSmnak /xašmnâk/ | agyn | zhragyn /zerangtarin/ |
| v | rySv /rišû/ | anh | dyvanh /divane/ | var | omydvar /omidvâr/ |
| any | rohany /rohâni/ | vr | sxnvr /soxanvar/ | Cy | tvpcy /tupči/ |
| vr | rnjvr /ranjvr/ | sar | Srmsar /šarmsâr/ | vS | SyrvS /širvaš/ |
| san | lahsan /lalesân/ | vnd | tnvmnd /tanvmand/ | syr | grmsyr /garmsir/ |
| y | kvhy /kuhil/ | gr | hylhgr /hilegar/ | yn | zryn /zarin/ |
| gyn | xSmgyn /xašmgyn/ | ynh | symynh /simine/ | mnd | danSmnd /danešmand/ |
| Noun → Adverb | | | | | |
| <Suff-NA > : SUn- & <A>; | | | | | |
| Suf | Example | Suf | Example | Suf | Example |
| an | bamdadan /bâmdâdân/ | y | ašry /?asri/ | | |

* باید توجه شود که فقط تعاریف مربوط به کلمات نوع A در این جدول ذکر شده است، تعاریف مشابهی نیز برای کلمات نوع B وجود دارد، مانند <Suff-NN-B> و غیره.

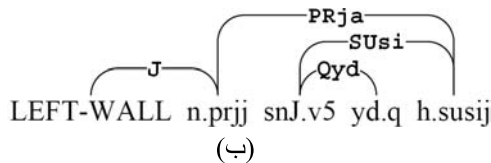
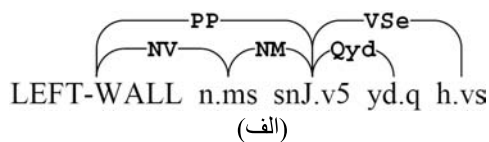
| Adjective → Adjective | | | | | |
|--|-------------------|-----|-------------------|-----|---------------------|
| <Suff-JJ-A> : SUj- & <J-A>; | | | | | |
| Suf | Example | Suf | Example | Suf | Example |
| a | phna /pahnâ/ | y | pnhany /penhâni/ | ar | pdydar /padidâr/ |
| yn | rastyn /râstin/ | man | Sadman /šadmân/ | ynh | symynh /simine/ |
| na | tngna /tangnâ/ | h | zrdh /zarde/ | h | myxarh /maixâre/ |
| Short Infinitive → Adjective | | | | | |
| <Suff-SIN > : SUsi- & <N-A>; | | | | | |
| Suf | Example | Suf | Example | Suf | Example |
| ar | krđar /kerdâr/ | gr | rftgar /roftegar/ | man | saxtman /saxtemân/ |
| h | dydh /dide/ | | | | |
| Present Stem → Noun | | | | | |
| <Suff-PSN > : SUpS- & <N-A>; | | | | | |
| Suf | Example | Suf | Example | Suf | Example |
| ar | prstar /parastâr/ | k | Qlġk /ġaltak/ | k | xrak /xarak/ |
| gr | rvygr /ruyggr/ | S | xorS /xoreš/ | man | zayman /zayemân/ |
| ndh | dvndh /davande/ | h | nalh /nale/ | | |
| Short Infinitive → Adjective | | | | | |
| <Suff-SIJ > : SUsi- & <J-A>; | | | | | |
| Suf | Example | Suf | Example | Suf | Example |
| ar | grftar /gereftâr/ | h | mrđh /morde/ | | |
| Present Stem → Adjective | | | | | |
| <Suff-PSJ > : SUpS- & <J-A>; | | | | | |
| Suf | Example | Suf | Example | Suf | Example |
| a | gyra /girâ/ | an | lQzan /laġzân/ | gra | bSrggra /bašargarâ/ |
| Adjective → Comparative Adjective | | | | | |
| <Suff-JSUP> : SUj- & <J-SUP>; | | | | | |

| Suf | Example | Suf | Example | Suf | Example |
|---|------------------|------|---------------------|-----|--------------|
| tr | zybatr /zibâtar/ | tryn | xvStryn /xuštarin/ | yn | bhyn /behin/ |
| Cardinal Number → Ordinal Number | | | | | |
| <Suff-CNON > : SUCn- & <ON> | | | | | |
| Suf | Example | Suf | Example | Suf | Example |
| m | Charm /čahâr/ | myn | Charmyn /čahâromin/ | | |
| Cardinal Number → Adjective | | | | | |
| <Suff-CNJ > : SUCn- & <J-A>; | | | | | |
| Suf | Example | Suf | Example | Suf | Example |
| y | dvmy /dovomil/ | | | | |
| Adjective → Infinitive Result | | | | | |
| <Suff-JIRa> : Ija- & <IR>; | | | | | |
| Suf | Example | Suf | Example | Suf | Example |
| y | xvby /xubi/ | | | | |
| Infinitive → Adjective | | | | | |
| <Suff-IFJ > : SUif- & <J-A>; | | | | | |
| Suf | Example | Suf | Example | Suf | Example |
| y | xvrdny /xordani/ | | | | |

* افعالی مانند گرا، نگر، انگار، باور، خواه

جدول ۹. میان‌وندهای زبان فارسی و ملزومات پیوندی آن‌ها

| Noun (Infix) Noun → Noun | | | | | |
|-----------------------------------|-------------------|-----|---------|-----|---------|
| <M-NNN > : N+ & N- & <N>; | | | | | |
| Inf | Example | Inf | Example | Inf | Example |
| bh | srbhsr /sarbesar/ | | | | |
| Noun (Infix) Noun → Adverb | | | | | |
| <M-NNA> : N+ & N- & <A> | | | | | |
| Pre | Example | Inf | Example | Inf | Example |
| bh | dmbhdm /dambdam/ | | | | |

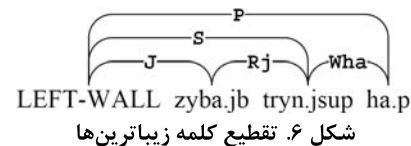


شکل ۷. دو تقطیع مختلف برای کلمه‌ی نسنجیده

اما در شکل ۸، ابهامی برای کلمه‌ی زیبایی وجود دارد در حالت اول، این کلمه یک حاصل مصدر و در حالت دوم یک صفت معین (Fi) می‌باشد که طبیعتاً به کمک تحلیل نحوی، قابل رفع می‌باشند.

۴-۲-۷. مثال‌ها

اولین مثال این بخش، یک صفت تفضیلی جمع (زیباترین‌ها) می‌باشد؛ واضح است که همه‌ی ویژگی‌های لازم از پیوندها استخراج می‌شود.



شکل ۶. تقطیع کلمه زیباترین‌ها

شکل ۷) مثالی نشان می‌دهد که در آن دو تقطیع مختلف برای کلمه وجود دارد. چنین موردی ابهام به‌نظر نمی‌رسد، چون صفت مفعولی، خود نوعی صفت به‌شمار می‌آید.

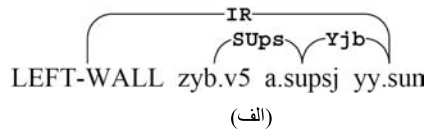
مشکل برای استفاده در محیط‌های واقعی، عدم رعایت قواعد نگارشی یکسان است که نیاز به بررسی مفصل‌تری دارد [12]. هم‌چنین واژگان سیستم دارای نقص زیادی می‌باشد که تکمیل آن در جهت معنا دار بودن ارزیابی این رهیافت بر روی پیکره‌های طبیعی بسیار لازم است. هر چند عدم وجود این واژگان، مانعی بر سر راه ارزیابی کیفی سیستم، یعنی صحت و کمال قواعد و مهم‌تر از آن نگرش به تکواژشناسی از طریق گرامر پیوندی، قرار نمی‌دهد.

در بخش ۴-۲-۷ به ابهام‌های سیستم اشاره شد. ساده‌ترین راه برای کاهش ابهام‌ها استفاده از روش هزینه‌گذاری بر روی اتصالات گره‌ها است [4]. در این روش به هر اتصال گرگر یک هزینه داده می‌شود تا در نهایت بتوان بین دو پیوندبندی، انتخاب کرد. در سیستم کنونی در برخی موارد از این روش استفاده شده است، اما علی‌رغم ابهامات فراوان، استفاده از تحلیل نحوی در بیشتر موارد به رفع آن‌ها کمک می‌کند.

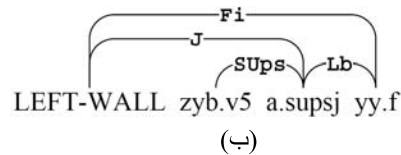
۷- نتیجه‌گیری و کارهای آینده

ما در این بخش تحلیل مفصلی از تکواژ شناختی زبان فارسی ارائه دادیم، توسعه‌ی این سیستم برای پذیرش موارد جدید، به‌سادگی صورت می‌گیرد و دیدیم که برخلاف پیچیدگی فراوان سیستم، وندی، به‌خصوص وندهای اشتقاقی، می‌توان به کمک قواعد زبانی، آن‌ها را مدیریت کرد. پیچیده‌ترین بخش تکواژشناسی زبان فارسی، مربوط به وندهای اشتقاقی می‌باشد که استفاده از یک صورت‌بندی مستقل از متن را ضروری می‌کند. بزرگ‌ترین مشکلی که پوشش همه‌ی این قواعد ایجاد می‌کند، ابهام‌های بسیار زیاد است که برخی از آن‌ها به کمک تحلیل‌های نحوی و معنایی قابل رفع هستند و رخداد تعدادی از آن‌ها نیز احتمال خیلی کمی دارد که در بیشتر موارد به اعمال هزینه بر روی پیوندها قابل رفع است. استفاده از مدل‌های احتمالاتی می‌تواند یک رهیافت دیگر برای رفع این ابهام‌ها باشد؛ به‌خصوص که گرامرهای پیوندی بستر بسیار مناسبی را برای تقطیع احتمالاتی ایجاد می‌کنند [۸]. از طرف دیگر تکمیل جداول وندها از اهداف دیگر این تحقیق می‌باشد.

در این مقاله ساختارهای فعلی، غیرفعلی، صرفی و اشتقاقی به صورت کامل بررسی شد و واژه‌نامه کامل برای پرهیز از هر گونه ابهام در پیش‌فرض‌های زبان‌شناسانه‌ی ما، ذکر شد که می‌تواند در کارهای مشابهی که در مورد



(الف)



(ب)

شکل ۸. نمونه‌ای از ابهام برای تقطیع کلمه‌ی زیبایی (الف) یک حاصل مصدر (ب) یک صفت معین

۵- محدودیت‌ها

هرچند در بخش‌های قبل، ذکر شد که تأثیرات آوایی در زبان فارسی کمتر به چشم می‌خورد؛ اما حالت‌هایی وجود دارد که در آن اضافه‌شدن یک پسوند، کلمه‌ی ماقبل را تغییر می‌دهد. گرامر پیوندی و هر گرامر مستقل از متن دیگری به شکل صوری و به‌سادگی قادر به رفع این مشکل نمی‌باشد. اما برای زبانی مثل فارسی با توجه به باقاعده بودن این تغییرات، می‌توان بر آن غلبه کرد. برای مثال بیشترین حالت مربوط به کلمات مختوم به "های" غیرملفوظ می‌باشد که هنگام گرفتن پسوندی چون -y یا -an، به g تبدیل می‌شوند. (نامه + ان = نامگان) در نتیجه، با ذکر این موارد تغییر فرم یافته‌ی (نامگ) در واژه‌نامه به‌سادگی آن را حل کرد.

۶- ارزیابی سیستم

همه‌ی قواعد این سیستم بر روی یک بستر آزمایش مصنوعی ارزیابی شده است. این بستر شامل مجموعه ساختارهایی است که کمال^۱ و صحت سیستم را مورد بررسی قرار داده است. این مجموعه شامل واژه‌هایی که این سیستم تحت پوشش قرار می‌دهد و هم‌چنین کلمات غیر معتبری که باید توسط گرامر رد شوند، می‌باشد [۱۰]. این روش ارزیابی مهم‌ترین روش آزمایش سیستم‌های NLP می‌باشد [11]. مصنوعی بودن این بستر آزمایش، این امکان را فراهم می‌کند که بتوان ساختارهای بسیار متنوعی را توسط سیستم آزمود. همان‌گونه که اشاره شد، این سیستم برخلاف مورد مشابه [1]، قادر به تشخیص وندهای اشتقاقی نیز می‌باشد که از مزایای آن به حساب می‌آید. مهم‌ترین

¹ Completeness

[۱۰] سجادی، ا.، همایون پور، م.م. ارائه‌ی یک سیستم تحلیل‌گر تکواژشناختی مبتنی بر گرامر پیوندی، گزارش فنی، آزمایشگاه پردازش گفتار، دانشکده‌ی مهندسی کامپیوتر و فناوری اطلاعات دانشگاه صنعتی امیرکبیر، ۱۳۸۴.

[11] Arnold, D. J., "Machine Translation, an Introductory Guide", Cochester, 1993.

[12] Megerdoomian, K., "Extending a Persian Morphological Analyzer to Blogs", The 2nd Workshop on Persian Language and Computer, Tehran University, 2006

[13] Lafferty, J., Sleator, D. and Temperley, D., "Grammatical Trigrams: A Probabilistic Model of Link Grammar", Proceedings of the AAAI conference on Probabilistic Approaches to Natural Language, October, 1992.

تکواژشناسی زبان فارسی ارائه می‌شود، مورد استفاده قرار گیرد. هم‌چنین روشی برای بازنمایی و استخراج ویژگی‌ها در صورت‌بندی پیوندی ارائه شد. محدوده‌ی پوشش داده شده توسط این صورت‌بندی، بخش اعظم زبان فارسی را تحت پوشش قرار می‌دهد. هرچند محدودیت‌هایی نیز وجود دارد که عموماً با پس‌پردازش‌هایی قابل رفع هستند. همه‌ی قواعد ارائه شده که تقریباً مشتمل بر ۲۵۰ قاعده‌ی لغوی می‌باشد، بر روی یک بستر آزمایش، شامل ساختارهای معرفی شده ارزیابی شده است.

۸- مراجع

[1] Megerdoomian, K., "Unification-based Persian morphology", In proceedings of CICLing, Alexander Gelbukh, Center of Investigation on Computation-IPN, Mexico, 2000.

[۲] سجادی، ا.، عبدالله زاده بارفروش، ا.، بازنمایی دانش دستوری زبان فارسی به کمک گرامرهای پیوندی، دومین کارگاه پژوهشی زبان فارسی و رایانه، دانشگاه تهران، ۱۳۸۵

[۳] باطنی. م.، "توصیف ساختمان دستوری زبان فارسی"، ۱۳۸۳، انتشارات امیرکبیر.

[4] Sleator, D. and Temperley, D., "Parsing English with a Link Grammar", Carnegie Mellon University Computer Science technical report CMU-CS-91-196, October 1991

[5] Sleator, D. and Temperley, D., "Parsing English with a Link Grammar", Third International Workshop on Parsing Technologies, 1991

[6] Koskenniemi, K., "Two level morphology: a general computational model for word-form recognition and production", Publication No. 11. Helsinki: University of Helsinki Department of General Linguistics, 1983.

[7] Antworth, E. L., "Morphological Parsing with Unification-based Word Grammar.", A paper presented at North Texax Natural Language Processing Workshop, 1994.

[8] Allen, J., "Natural language understanding", The Benjamin / Cumming Publishing Company, 1994.

[۹] انوری، ح.، گیوی، "حسن، دستور زبان فارسی ۲"، ویرایش ۲، چاپ بیست و سوم، ۱۳۸۳، انتشارات فاطمی



آرمین سجادی در سال ۱۳۶۰ در شهر سنندج متولد شد. تحصیلات تا مقطع دیپلم را در شهر سنندج سپری و دیپلم متوسطه‌ی خود را در سال ۱۳۶۰ دریافت کرد. وی

تحصیلات خود در مقطع کارشناسی را در رشته‌ی مهندسی کامپیوتر در دانشگاه علم و صنعت ایران (سال ۱۳۸۳) و کارشناسی ارشد را در رشته‌ی هوش مصنوعی از دانشگاه صنعتی امیر کبیر (پلی تکنیک تهران) (سال ۱۳۸۵) به پایان رسانید. زمینه‌ی تخصصی مورد علاقه‌ی ایشان پردازش نمادی و زبان‌شناسی محاسباتی می‌باشد.

نشانی (رایانامک) پست الکترونیکی ایشان عبارت است از:

لاف.ع.م.مع.اف.ع.ف.ع.ل



محمد مهدی همایون پور در سال ۱۳۳۹ در شهر شیراز متولد شد. تحصیلات تا مقطع دیپلم را در شهر شیراز سپری و دیپلم متوسطه‌ی خود را در سال ۱۳۵۸ دریافت کرد. وی

تحصیلات خود در مقطع کارشناسی را در رشته‌ی مهندسی برق (الکترونیک) در دانشگاه صنعتی امیرکبیر (سال ۱۳۶۶)، کارشناسی ارشد را در رشته‌ی برق (مخابرات)، از دانشگاه خواجه نصیرالدین طوسی (سال ۱۳۶۹)، کارشناسی ارشد دوم خود را در زمینه‌ی فونتیک (۱۳۷۴) در دانشگاه سوربون جدید در فرانسه و هم‌زمان دوری دکترای خود را در دانشگاه پاریس ۱۱ در زمینه‌ی مهندسی برق (۱۳۷۴) به پایان رسانید. نامبرده از سال ۱۳۷۴ در سمت عضو هیأت علمی دانشکده‌ی مهندسی کامپیوتر و فناوری اطلاعات دانشگاه صنعتی امیرکبیر به تدریس و تحقیق مشغول می‌باشد. ایشان علاوه بر تدریس، راهنمایی پروژه‌های کارشناسی، کارشناسی ارشد و دکتری در زمینه‌های مهندسی کامپیوتر و فناوری اطلاعات و نیز هدایت تعداد زیادی پروژه‌های صنعتی و ملی را برعهده داشته است. نامبرده عضو انجمن‌های علمی کامپیوتر، ارتباطات و فناوری اطلاعات و رمز می‌باشد و مسؤولیت‌های اجرایی متعدد از جمله ریاست و معاونت‌های آموزشی و پژوهشی دانشکده‌ی مهندسی کامپیوتر و فناوری اطلاعات دانشگاه صنعتی امیرکبیر و شرکت در برگزاری چندین کنفرانس و مسابقه علمی را بر عهده داشته و موفق به انتشار بیش از ۸۰ مقاله علمی - پژوهشی در مجلات و کنفرانس‌های علمی داخل و خارج از کشور گردیده است.

آدرس پست الکترونیکی ایشان عبارت است از:

homayoun@aut.ac.ir

homayoun@ce.aut.ac.ir