

پارس مورف: تحلیلگر صرفی زبان فارسی

وحید مواجی^۱، محرم اسلامی^۲ و بهرام وزیرنژاد^۳
دانشگاه صنعتی شریف^۱ و دانشگاه زنجان^۲

چکیده:

در مقاله حاضر می‌کوشیم مبنای نظری، نحوه طراحی و عملکرد سامانه تحلیلگر صرفی زبان فارسی با عنوان اختصاری "پارس مورف" را معرفی کنیم. پارس مورف سامانه‌ای مبتنی بر قواعد صرفی زبان فارسی است که ساخت درونی کلمات فارسی را با توجه به نظام تصrif و نظام واژه‌سازی زبان، تجزیه و تحلیل و مقوله دستوری و نقش هر کدام از اجزای سازنده کلمه را مشخص می‌کند. پارس مورف با استفاده از یک واژگان در حدود ۴۵۰۰۰ واژه‌ای و نیز در چارچوب قواعد صرفی زبان فارسی که بر یک تحقیق جامع زبان‌شناسی استوار است می‌تواند واژه‌های پیچیده و نیز صورت‌های ممکن تصrifی و حتی واژه‌های خارج از واژگان را تحلیل کند.

دقت نسخه اول پارس مورف حدود ۹۵٪ است که افزودن اطلاعات نحوی و مسائل مربوط به همویشه‌ها و نیز لحاظ‌کردن ویژگی‌های خط فارسی می‌تواند این دقต را به ۱۰۰٪ نزدیک کند. از پارس مورف می‌توان در مطالعات محض زبانی و نیز پردازش ماشینی زبان فارسی استفاده کرد.

واژگان کلیدی: پارس مورف، تحلیلگر صرفی، زبان فارسی، واژه‌سازی، تصrif، اشتراق و ترکیب.

می‌گیرد. هر کدام از واژه‌ها^۱ با توجه به جایگاهشان در جمله می‌توانند صورت-کلمه‌های متفاوت داشته باشند که در اصطلاح غیردقیق به آنها کلمه^۲ نیز می‌گویند. صورت-کلمه‌ها واژه‌های جدید نیستند، بلکه صورت تصrifی واژه به حساب می‌آیند. در نظام واژه‌سازی، واژه‌ها بسیط و یا پیچیده‌اند. واژه‌های بسیط می‌توانند با هم ترکیب شوند و واژه مرکب بسازند و از طرف دیگر واژه‌های بسیط و یا مرکب می‌توانند با عناصر واژه‌ساز (تکواژه‌های اشتراقی) زبان ترکیب شوند و به ترتیب واژه‌های پیچیده مشتق و یا مرکب-مشتق سازند. به نمونه‌های زیر توجه کنید:

(الف) واژه بسیط: اکار

(ب) صورت کلمه: کار، کارها، کارش، کارشے

(ج) واژه مشتق: اکارگر، کارمند، کارایی، کاری(کارکن)/

(د) واژه مرکب: اکارخانه‌ساز، کاربلد، کارآگاه، کارنامه/

۱- مقدمه

صرف یا ساختواژه^۱ به عنوان بخشی از زبان‌شناسی ناظر بر مطالعه ساخت درونی کلمات و روابط نظام‌مند صورت و معنا در کلمات است (Booij, Geert., 2007). وقتی صحبت از ساخت (کلمه) می‌شود، تلویحاً می‌پذیریم که کلمه دارای اجزایی است و درنتیجه هر کدام از اجزا نقش و جایگاه مشخصی در ساخت کلمه دارد. در این تحقیق پس از صورت‌بندی ساخت درونی کلمه، سامانه‌ای را طراحی کرده‌ایم که قادر است به طور خودکار ساخت درونی کلمات فارسی را تحلیل و اجزای سازنده کلمه را مشخص کند. پیش از ورود به بحث اصلی، در ادامه به اختصار به مبانی صرف اشاره می‌کنیم تا در جریان بحث درخصوص عملکرد پارس مورف، ابهامی وجود نداشته باشد.

ساختواژه از دو نظام مستقل تصrif^۲ و واژه‌سازی^۳

تشکیل می‌شود و در تصrif صورت-کلمه‌ها^۴ مورد بحث قرار

¹ morphology

² inflection

³ word-formation

⁴ word-forms

⁵ lexemes

⁶ word

۵) واژه مركب-مشتق: اکاردرمانی، کارخانه‌چی، کارمندمحور، کاردرمانگری / با توجه به نمونه‌های ارائه شده در مثال ۱، دستگاه صرف دو نقش اساسی بر عهده دارد: تصريف و واژه‌سازی (Bauer, 1983، طباطبایی ۱۳۱۲ ص ۶). واژه بسيط (الف) واژه‌ای است که از یك ريشه تشکيل شده است. صورت-كلمه (ب) صورت‌های تصريف شده يك واژه می‌گويند. از ۱.ج تا ۱.ج واژه‌های پیچیده را می‌بينيم که در چارچوب امكانات واژه‌سازی زبان درست شده‌اند. واژه مشتق (ج) واژه‌ای است که در ساختمان آن يك واژه بسيط يا واحد واژگانی بسيط و يك يا چند تکواز اشتقاقي به کار رفته است. واژه مركب واژه‌ای است که از دو يا چند واژه بسيط يا واحد واژگانی بسيط درست شده باشد. واژه مركب -مشتق (م) واژه‌ای است که در ساختمان آن دو يا چند واژه بسيط يا واحد واژگانی بسيط و يك يا چند تکواز اشتقاقي به کار رفته باشد. با توجه به مثال‌های بالا می‌بينيم که بخشی از تکوازهای مقيد، واژه جديد نمي‌سازند (مانند ۱.ب) که به آنها تکوازهای غيراشتقاقي و بهتسامح كل آنها را تکوازهای تصريفی می‌گويم. بخشی دیگر (در ۱.ج و ۱.م) واژه جديد می‌سازند که به آنها تکوازهای اشتقاقي می‌گويم. به عبارت دیگر در تصريف با واژه‌پردازي سروکارداريم، اما در اشتقاقي با واژه‌سازی (طباطبایي ۱۳۷۶ ص ۱۵). ستاك (stem) بهصورت-كلمه‌ای اطلاق می‌شود که وندهای تصريفی از آن حذف شده باشند. ستاك می‌تواند بسيط يا پیچیده باشد و ستاك بسيط همان ريشه است (Booij, Geert., 2007).

در طراحي پارس مورف، ساخت تصريفی کلمه در زبان فارسي (اسلامي و علیزاده ۱۳۸۱) بهصورت ساختمند در نظر گرفته شده است و جايگاه طبقات مختلف وندهای تصريفی در ساختمان انواع کلمات مشخص شده است. پارس مورف در تحليل تصريفی کلمه ابتدا ستاك را شناسايي می‌کند و متناسب با ساخت تصريفی پيش‌بياني شده برای آن نوع ستاك به دنبال انواع وندهای تصريفی در جايگاه‌های خاص، با در نظر گرفتن صورت‌های مختلف نوشتاري هرکدام از طبقات، می‌گردد. يافتن وندی به عنوان مثال در جايگاه دوم پس از ستاك اسمی، نشان می‌دهد که جايگاه اوّل خالي مانده است. در بخش ۲ در مقاله حاضر بهطور كامل به موضوع ساخت تصريفی کلمه در زبان فارسي پرداخته‌aim.

اگر ستاك پیچیده باشد، پارس مورف با اعمال قواعد واژه‌سازی زبان فارسي، ستاك مورد نظر را از حيت مشتق يا مركب‌بودن تجزيه و تحليل و اجزاي سازنده آن را با ذكر مقوله دستوري و نقش آنها مشخص می‌کند. به دليل معتبر

نبودن فاصله به عنوان مرز کلمه در متون فارسي، پارس مورف در تجزيه ستاك‌های پیچیده ترکيب‌های بالقوه را نيز به عنوان گزينه‌های بعدی در اختيار ما می‌گذارد. به عنوان مثال ستاك پیچیده "کارگر" در بخش اشتقاقي به عنوان کلمه مشتق شناسايي و يا در ترکيب گفته می‌شود که در فارسي ممکن است آن ترکيب "کار" (N1) + "گر" (Adv) يعني اسم + قيد باشد که اين دو کلمه بيفاصله در کنار هم آمداند. در بخش ۳ به عملکرد پارس مورف در ريشه‌يابي و تعين اجزاي واژه‌های پیچیده خواهيم پرداخت.

در تمامي تحقيقات مربوط به پردازش‌های خودکار زيانی در زبان فارسي، به خصوص در پردازش متن فارسي برای مقاصد مختلف اعم از ترجمه ماشيني، تبديل متن به گفتار و غيره از نوعی تحليل گر صرفی استفاده می‌شود؛ اگر چه در اغلب مواقع تحليل گرهای صرفی در تحقيقات پيشين محدود، هدف محور و فاقد پشتوانه جامع زبان‌شناختي است. به عنوان مور پيشينه پژوهش در ادامه تنها به مواردي اشاره می‌كنيم که در تحليل ساخت درونی کلمه فارسي نگاه ساخت‌مند داشته‌اند و يا يك رویکرد زبان‌شناختي-مهندسي به تحليل صرفی کلمه فارسي پرداخته‌اند. در اين خصوص ابتدا می‌توان به مطالعات دقیقی اشاره کرد که در پروژه شيراز (Megerdormian, et al., 2000) در طراحي ترجمه سامانه ماشيني فارسي-انگلسي اشاره کرد که در ميانه راه متوقف شد. دومين مورد از طراحي تحليل خودکار فارسي مربوط به تحليل گر تصريفي زبان فارسي است که در سامانه تبديل متن به گفتار فارسي "گويا" به کار گرفته شد (اسلامي و همکاران ۱۳۱۳) که تنها به تحليل تصريفی کلمه محدود می‌شد. آماده‌سازی متن معيار برای زبان فارسي ۱ با عنوان اختصاری STeP1 سامانه دیگري است که قادر است کلمات فارسي را از نظر صرفی تجزيه و تحليل کند (شمس‌فرد ۱۳۱۳). STeP1 با استفاده از واژگان زياری زبان فارسي (اسلامي و همکاران ۱۳۱۳) و قواعد صرفی که طراحان آن در نظر گرفته‌اند کار می‌کند.

اگرچه STeP1 در تجزيه کلمه به اجزاي سازنده آن تا حد زياری موفق عمل می‌کند، ولی مبناي علمي زيان‌شناختي آن نياز به بازنگري دارد تا از اشكالات فعلی پرهيز شود. به عنوان مثال برای کلمه "دانشگاه‌ها" سه ريشه در نظر می‌گيرد، به ترتيب "دانشگاه، دانش، دان" که نشان می‌دهد حداقل منظور پديدآورندگان آن از "ريشه" در مفهوم علمي آن اصطلاح نسيت. يا در کلمه "دانشگاه‌هايمان" علاوه بر اختصاص سه ريشه فوق همزمان دو تحليل زير برای آن ارائه می‌شود:

همدیگر قرار دارند و در ساخت کلمه، اجزای هر کدام عنوان و نقش منحصر به فردی دارند. پارس مورف صورت واژگانی کلمه را از واژگان زیای زبان فارسی (اسلامی و همکاران ۱۳۱۳) می‌گیرد و در چارچوب نظام تصريف کلمه در زبان فارسی (اسلامی و علیزاده ۱۳۱۱) به تحلیل تصريفی کلمه می‌پردازد. در واژگان زایا مقوله دستوری هر کدام از واژه‌ها مشخص شده است و بعد از شناسایی واژه با توجه به مقوله دستوری آن ساخت تصريفی را برای آن در نظر می‌گیرد. اگر در ساختمان کلمه از وابسته‌های تصريفی پيش‌بینی شده در جايگاه‌های خاص مورد يا مواردي وجود داشته باشد، در آن صورت عنوان هر کدام را مشخص می‌کند. واژه ممکن است بدون هیچ‌گونه وند تصريفی در جمله به کار رود که در چنین شرایطی واژه و کلمه صورت يكسانی خواهد داشت. وابسته‌های تصريفی عناصر اختياري در ساختمان کلمه هستند و پارس مورف قادر است حتی با استفاده از قواعد تصريف که در اختيار دارد کلمات خارج از واژگان را شناسایي و عناصر آن را مشخص کند و اگر ستابک ناشناخته پيچيده باشد، اجزای آن را هم با توجه به قواعد واژه‌سازی روشن می‌کند. درخصوص کلمه در کلمه‌ها (words within word) نيز در پارس مورف تمھيداتي اندیشيده شده است که در گزینه‌های بعدی که در اختيار کاربر قرار می‌دهد به آنها اشاره می‌کند. کلمه در کلمه‌ها واژه‌هایي هستند که در درون آنها می‌توان کلمات ديگري نيز پيدا کرد، مانند "مادر" که در عين حال واژه بسيط و دو کلمه مجزا ("ما + در") است. البته در شناسایي کلمه در کلمه‌ها در پارس مورف از قواعد صرفی استفاده کرده‌است.

در ادامه به ساخت تصريفی اسم و فعل اشاره می‌کنيم. شایان ذكر است که در زبان فارسي، صفت‌ها بالقوه اسم هستند و با استتفاق صفر به اسم تبديل می‌شوند و صورت‌های تصريفی اسم را می‌پذيرند. اين موارد در طراحي پارس مورف لحظ شده است. در ساخت تصريفی اسم، هسته صورت واژگانی اسم و عنصر اجباري است که می‌تواند وابسته‌های تصريفی اختياري داشته باشد که به همين دليل در ساخت تصريفی اسم در درون () قرار دارند.

(نحو: نکره) (نحو: بهم مجموعی) [اسم] (فتحه: حسن)، [فرزه: پنهان]، [ضمیر: ضمیر مصل] (گرداً اتفاق)

ساختار تصريفی اسم (اسلامی و علیزاده ۱۳۸۸)

چنانچه در ساخت تصريفی اسم می‌بینیم، اسم بالقوه می‌تواند در سه جايگاه، وابسته‌هایي پذيرد که اين جايگاهها

سال ۱۳۹۰ شماره ۱ پيادي ۱۵

الف) اسم + علامت جمع + i + ضمير ملكي اول شخص جمع که در آن "دانشگاه" ريشه است، به صورت "دانشگاه + ها + i + مان".

ب) اسم + گاه + علامت جمع + i + ضمير ملكي اول شخص جمع که در آن "دانش" ريشه است، به صورت "دانش + گاه + ها + i + مان".

چنانچه می‌بینيم هیچ مبنای علمی برای ريشه‌بودن "دانشگاه" يا "دانش" در مثال فوق وجود ندارد؛ از طرف ديگر اجزای تجزيه شده لزوماً واحدهای نظام واژه‌سازی و نظام تصريف نیستند. به عنوان مثال "i" در مثال بالا چه عنوان و نقش زبانی دارد؟ جزاينکه به خاطر شرایط واژه‌اجی در زنجیره واجی کلمه ظاهر شده است. همچنین کلماتي مانند "کفشهایمه" يا "کفشهایمند" را STEPI نمی‌تواند تجزيه کند و هیچ ريشه‌ای برای اين قبيل تصريفها و صورت-کلمه‌ها و همچنین برخی ستاك‌های پيچيده و بسيط خارج از واژگان پيدا نمی‌کند. اين در حاليست که پارس مورف با دقت كامل موارد فوق را تجزيه و تحليل می‌کند.

پارس مورف، تحليل گر صرفی زبان فارسي، که با زبان برنامه‌نويسی پايتون^۱ نوشته شده است بر يك مبنای بهطور كامل علمی زبانی استوار است و در چارچوب ساختمان صرفی کلمه فارسي به تجزيه و تحليل و تعين نقش هر کدام از اجزا در درون کلمه می‌پردازد. پس از صورت‌بندی دقیق اطلاعات نظام تصريف و واژه‌سازی در زبان فارسي سعی کرديم در مرحله اجرا و طراحی سامانه پارس مورف تمامی آن اطلاعات را به شكل دقیق به کار بگیریم. در حال حاضر پارس مورف با استفاده از آخرین ويرايش واژگان زایای زبان فارسي (اسلامی، شريفي آتشگاه، علیزاده لمجيري و زندي، ۱۳۱۳) که حدود ۴۵۰۰ واژه در آن قرار دارد و در چارچوب قواعد صرفی زبان فارسي که در اختيار دارد، با دقت بالاي ۹۵٪ می‌تواند ساخت درونی کلمات فارسي را تحليل کند. نيز می‌تواند با استفاده از امكانات و اطلاعات صرفی که در اختيار دارد کلمات خارج از واژگان را نيز از حيث تصريف و واژه‌سازی تجزيه و تحليل کند.

۲- نظام تصريف زبان فارسي

چنانچه در مقدمه گفتيم، در طراحي پارس مورف رویکرد ساخت-بنiard اتخاذ شده است که به موجب آن عناصر صرفی تشکيل‌دهنده کلمه در يك ارتباط ساختاري و نظاممند با

¹ Python

مقاله تهییه شد و پارس مورف با در اختیار داشتن این فهرست و الگوهای استتفاق به تجزیه کلمات مشتق فارسی می‌پردازد. در الگوهای استتفاق مشخص شده است که هر کدام از تکوازهای استتفاقی به چه نوع ستابکی از نظر مقوله افزوده می‌شود و واژه حاصل از رهگذر استتفاق چه مقوله دستوری خواهد داشت. به عنوان مثال کلمه "دانشگاهی" به صورت "دان(N) + ش(N) اسم‌ساز" + گاه(N=اسم‌ساز) + "اسم‌ساز" تجزیه می‌شود. پارس مورف با در اختیار داشتن امکانات ذکر شده می‌تواند کلمات خارج از واژگان را نیز تجزیه و تحلیل کند و نیز ترکیب‌های بالقوه را برای کلمه مورد نظر در گرینه‌های بعدی پیشنهاد کند. نظر به عدم قطعیت در استفاده از فاصله در متون فارسی به عنوان مرز کلمه، پارس مورف این قابلیت را دارد که علاوه‌بر تجزیه و تحلیل کلمه در سطح اشتتفاق برای کلمه مورد نظر صورت‌های ترکیبی نیز در نظر بگیرد. با افزودن اطلاعات نحوی به پارس مورف در موارد مشکوک به اشتتفاق یا ترکیب و یا دو کلمه مستقل بدون فاصله می‌توان از گرینه‌هایی دیگری استفاده کرد که پارس مورف ارائه می‌کند.

مسئله مهم دیگر در تجزیه کلمات مشتق، موضوع همنویسه‌ها^۱ است که کار تجزیه خودکار کلمه را دچار مشکل می‌کند. به عنوان مثال "ای" با ساخت واجی "(i)" می‌تواند در نقش‌های زیر به کار رود: (الف) تکواز اشتتفاقی (یا نسبت) که به طور معمول اسم را به صفت و بر عکس صفت را به اسم تبدیل کند. به عنوان مثال در کلمات "دانشگاهی، زیبایی"، ب) در نقش "ای" نکره به عنوان مثال در "كتابي" در جمله "كتابي خريدم" که اشاره به کتاب نامعین می‌کند، ج) در نقش "ای" بند موصولی ساز، به عنوان مثال در "پسری" در جمله "پسری که دیدی" و د) در نقش واژه‌بست ربطی، به عنوان مثال در "تو پسری" به معنی "تو پسر هستی". علاوه بر صورت‌های یکسان تکوازهای اشتتفاقی و غیراشتفاقی نوع دیگری از همنویسه هست که در آنها بین صورت واژگانی و صورت تصریفی ابهام وجود دارد. به عنوان مثال "مردم" که به صورت‌های زیر تجزیه می‌شود:

الف) مردم(N1)

ب) مرد(N1) + am (واژه بست ربطی)

ج) مرد (N1) + am (واژه بست شخصی یعنی مردِ من)

د) مُرد (V2) + بست (شخصی) (واژه am)

البته در ساختهای مبهم از اطلاعات واژه‌سازی و تصریفی برای رفع ابهام استفاده می‌شود. به عنوان مثال در

4 homographs

معتبر هستند و هرگونه جایه‌جاوی در نظم وابسته‌ها منجر به کلمه بدساخت می‌شود. در هر کدام از جایگاه‌ها نیز صورت‌های مختلف تکواز (به عنوان مثال تکوازهای مختلف جمع در جایگاه اول) یا طبقات گوناگونی از وابسته‌ها به کار می‌رود که حضور یکی از آنها مانع حضور بقیه گونه‌های یک تکواز یا دیگر طبقات مربوط به آن جایگاه می‌شود. بنابراین پارس‌مورف هیچ‌گاه دو تکواز جمع یا دو تکواز مربوط به جایگاه دوم یا سوم در ساختمان اسم شناسایی نمی‌کند. به این محدودیت در اصطلاح توزیع تکمیلی^۱ می‌گویند که پارس‌مورف در تحلیل ساخت تصیریفی کلمه محدودیت بالا را از نظر تصیریف و واژه‌سازی لحظ می‌کند. در واژه‌سازی نیز پارس‌مورف تکرار تکواز مثلاً "گر" را به طور هم‌زمان در ساختمان کلمه مجاز نمی‌داند.

تحلیل ساخت تصیریفی فعل به دلیل پذیرفتن پیشوند و پسوند تصیریفی دشوارتر است. در واژگان زایه، بن مضارع و بن ماضی صورت‌های واژگانی فعل در نظر گرفته شده است و پارس‌مورف با توجه به این موضوع ساخت تصیریفی فعل را تحلیل می‌کند. در ادامه ساخت تصیریفی فعل فارسی را می‌بینیم که محدودیت‌های ذکر شده در خصوص اسم در فعل نیز مشاهده می‌شود.

$$[(\text{النسبة المئوية} \cdot \text{الناتج المحلي الإجمالي}) + ((\text{الناتج المحلي الإجمالي}) - (\text{الناتج المحلي الإجمالي})) + \text{الناتج المحلي الإجمالي}] + [\frac{\text{الناتج المحلي الإجمالي}}{(\text{الناتج المحلي الإجمالي} - \text{الناتج المحلي الإجمالي})}]$$

ساختار تصویری فعل (اسلامی و علیزاده ۱۳۸۸)

دیگر اقسام کلمه مانند صفت‌ها، قیدها، اعداد، حروف اضافه نیز ساخت تصریفی خاص خود را دارند که در طراحی پارس‌مorf بهطور کامل مورد توجه بوده‌اند.

۳- نظام واژه‌سازی زبان فارسی

نظام واژه‌سازی زبان فارسی نیز همانند نظام تصویری این زبان در طراحی پارس‌مورف با دقت صورت‌بندی شده است. دو شیوه عمده واژه‌سازی یعنی اشتقاق^۲ و ترکیب^۳ در زبان فارسی عمده‌ترین شیوه‌های واژه‌سازی هستند. پارس‌مورف با مراجعه به واژگان زیایی زبان فارسی (اسلامی و همکاران ۱۳۱۳) و نیز با توجه به امکانات اشتقاق و ترکیب به تجزیه کلمات مشتقه و مرکب می‌پردازد.

۱-۳- اشتقاق در زبان فارسی

فهرست تکوازهای اشتراقی با استفاده از ساخت اشتراقی واژه در فارسی امروز (کلیبасی، ۱۳۷۱) و اشتراق پسوندی در زبان فارسی، امروز (کشانی، ۱۳۷۱) و نیز ملاحظات نویسنده‌گان این

¹ complementary distribution

² derivation

³ derivation compounding

"دانشگاهی‌ها"، پارس مورف به هیچ وجه "ی" را "یای" غیر اشتراقی قلمداد نمی‌کند، چون قبل از تکواز جمع آمده است.

۲-۳- ترکیب در زبان فارسی

با توجه به تعریفی که از ترکیب در مقدمه مقاله حاضر ارائه کردیم، اصول و الگوهای ترکیب صورت‌بندی و پارس‌مورف با مراجعه به واژگان زیایی زبان فارسی (اسلامی و همکاران ۱۳۸۳) و با در دست داشتن الگوهای ترکیب (طباطبایی، ۱۳۸۳) کلمات مرکب را شناسایی و اجزای ترکیب را با ذکر مقوله دستوری آنها مشخص می‌کند.

واژه‌های بسیط و واحدهای واژگانی غیرپیچیده (مانند بن‌های مضارع) در چارچوب الگوهای ترکیب (همان) می‌توانند با هم ترکیب شوند و واژه مرکب درست کنند. در اغلب موقع اجزای واژه مرکب خود مدخل‌های واژگانی در واژگان برنامه هستند که در آن صورت مقوله‌های دستوری اجزای ترکیب و نیز مقوله دستوری واژه مرکب مشخص است. کلمات مرکب خارج از واژگان نیز با مراجعه به واژگان برنامه و با درنظر گرفتن الگوهای ترکیب تجزیه می‌شوند.

درخصوص کلمات مرکب - مشتق نیز، پارس‌مورف قواعد اشتراق و ترکیب را هم‌زمان لحاظ می‌کند و این واژه‌ها را به اجزای سازنده‌اش تجزیه می‌کند که در آن نقش و مقوله دستوری هر کدام از اجزا مشخص شده است. به عنوان مثال کلماتی مانند "کارخانه‌چی" و یا "دانشجو" را به ترتیب به صورت "کار(N1) + خانه(N1) + چی(N)=تکواز اشتراقی اسم‌ساز" و "دان(VI) + ش(N=تکواز اشتراقی اسم‌ساز) + جو(VI)" تجزیه می‌کند.

۴- نتیجه‌گیری

در این تحقیق، سامانه تحلیل صرفی کلمه در زبان فارسی با عنوان پارس‌مورف را معرفی کرده‌ایم که بر پایه یک مطالعه دقیق زبان‌شناختی از نظام صرفی زبان فارسی استوار است. پارس‌مورف قادر است ساخت درونی کلمه فارسی را از حیث نظام تصريف و واژه‌سازی تجزیه و تحلیل کند و برای هر کدام از اجزا در درون کلمه عنوان و نقش زبانی خاصی اختصاص دهد. در اجرای این تحقیق پس از صورت‌بندی دقیق نظام تصريف و واژه‌سازی، الگوریتم عملیاتی برنامه تحلیل‌گر فارسی پریزی و تکمیل شد. پارس‌مورف به لحاظ نظری بر پایه علمی استوار است و نسخه اول آن قادر است با دقت بالای ۹۵٪ ساخت درونی کلمات فارسی را تجزیه و تحلیل کند و با

افزومن اطلاعات مربوطه به همنویسه‌ها و افزودن اطلاعات نحوی و نیز رفع خطای برنامه که جزء جدانشدنی فرآیند توسعه نرم‌افزار است، این دقت به نزدیک ۱۰۰٪ خواهد رسید. پارس‌مورف می‌تواند در دو زمینه به طور کامل مجزا به کار گرفته شود و یاریگر محققان در این دو زمینه باشد. در مطالعات زبان‌شناختی محض؛ به عنوان مثال در تهیه منابع زبانی برای زبان فارسی و نیز در زمینه انواع پردازش خودکار زبان فارسی، به طور مشخص در تبدیل متن به گفتار، ترجمه ماشینی، استخراج اطلاعات، خلاصه‌سازی متون و غیره به کار گرفته شود.

۵- منابع

- اسلامی، محرم، شریفی آتشگاه، مسعود، علیزاده لمجیری، صدیقه و زندی، طاهره. ۱۳۸۳. واژگان زیایی زبان فارسی، مجموعه مقالات اولین کارگاه پژوهشی زبان فارسی و رایانه، دانشگاه تهران، تهران.
- اسلامی، محرم و علیزاده، صدیقه. ۱۳۸۸. ساخت تصريفی کلمه در زبان فارسی، زبان و ادب فارسی، نشریه دانشکده ادبیات و علوم انسانی دانشگاه تبریز، سال ۵۲، شماره مسلسل ۲۱، ایران.
- شمس‌فرد، مهرنوش. ۱۳۸۸. STeP1: تهیه متن معیار برای زبان فارسی. آزمایشگاه پژوهش زبان طبیعی، دانشگاه شهید بهشتی، ایران.
- طباطبائی، علاءالدین. ۱۳۷۶. فعل بسیط فارسی و واژه‌سازی، مرکز نشر دانشگاهی تهران، ایران.
- طباطبائی، علاءالدین. ۱۳۸۲. اسم و صفت مرکب در زبان فارسی، مرکز نشر دانشگاهی، تهران، ایران.
- کشانی، خسرو. ۱۳۷۱. اشتراق پسوندی در زبان فارسی امروز، مرکز نشر دانشگاهی تهران، ایران.
- کلباسی، ایران. ۱۳۷۱. ساخت اشتراقی واژه در فارسی امروز. موسسه مطالعات و تحقیقات فرهنگی، تهران، ایران.
- Bauer, Laurie., 1983. English Word- Formation, CUP, Cambridge.
- Booij, Geert., 2007. The Grammar of Words, 2nd edition, Oxford University Press.
- Megerdoomian, Karine., 2000. Persian Computational Morphology, A Unification-Based Approach, NMSU, CRL, Memoranda in Computer and Cognitive Science, (MCCS-00-321).

وحید مواجهی مدرک کارشناسی خود را



در رشته مهندسی کامپیوتر در سال

۱۳۸۳ از دانشگاه علم و صنعت ایران اخذ

نمود. وی در سال ۱۳۸۵ مدرک

کارشناسی ارشد خود را در رشته

مهندسی کامپیوتر از دانشگاه صنعتی شریف دریافت نمود.

وی همچنین از سال ۱۳۸۹ در رشته زبان‌شناسی رایانشی در

مقطع کارشناسی ارشد در دانشگاه صنعتی شریف مشغول به

تحصیل می‌باشد. زمینه‌های تحقیقاتی وی شامل زبان‌شناسی

رایانشی، صرف و مهندسی نرم‌افزار می‌باشد.

نشانی رایانمک ایشان عبارت است از:

vaji@alum.sharif.edu

محرم اسلامی در سال ۱۳۷۰ در رشته



مترجمی زبان انگلیسی از دانشگاه علامه

طباطبایی فارغ‌التحصیل شد. سپس در

سال ۱۳۷۳ در رشته زبان‌شناسی همگانی

دوره کارشناسی ارشد را دانشگاه تهران

به پایان رسانید و در سال ۱۳۷۹ در همان رشته و همان

دانشگاه موفق به اخذ مدرک دکترا شد. از ایشان بیش از ۳۰

مقاله در نشریات و همایش‌های داخلی و خارجی به‌چاپ

رسیده است، و همچنین صاحب تألیفاتی می‌باشد. وی

هم‌اکنون عضو هیئت علمی دانشگاه زنجان می‌باشد و

زمینه‌های تخصصی ایشان، زبان‌شناسی همگانی، واج‌شناسی

آهنگ و آواشناسی و پردازش رایانه‌ای زبان طبیعی (NLP)

می‌باشد.

نشانی رایانمک ایشان عبارت است از:

meslami@znu.ac.ir

بهرام وزیرنژاد به عنوان عضو هیات علمی



در دانشگاه صنعتی شریف مشغول به

فعالیت آموزشی و پژوهشی است. ایشان

دکتراخود را در رشته مهندسی پزشکی

از دانشگاه صنعتی امیرکبیر دریافت نموده

است. وی طی سال‌های ۲۰۰۷ تا ۲۰۰۸ به عنوان پژوهشگر

مهمان در دانشگاه سیدنی مشغول به فعالیت تحقیقاتی بود. از

او حدود ۲۰ مقاله در کنفرانس‌ها و نشریات معتبر داخلی و

خارجی به چاپ رسیده است. تحقیقات ایشان شامل

حوزه‌های پردازش گفتار و زبان طبیعی، تعامل انسان و

ماشین، هوش مصنوعی و زبان‌شناسی رایانشی می‌باشد.

نشانی رایانمک ایشان عبارتست از:

bahram@sharif.edu

فصلنامه
پژوهش‌های

سال ۱۳۹۰ شماره ۱ پیاپی ۱۵

