

فرهنگ ظرفیتی فعل: تلاشی در جهت آموزش فارسی به غیر فارسی‌زبانان

* امیر سعید مولودی

** منوچهر کوهستانی

*** محمدصادق رسولی

چکیده

در زبان، برخی از اجزاء سخن ظرفیت دارند. مفهوم ظرفیت که برگرفته از نظریه دستور وابستگی است، به تعداد و نوع پیوندهایی که عناصر نحوی می‌توانند با هم تشکیل دهند، اطلاق می‌شود. در دستورهای مبتنی بر ظرفیت، انگاره‌ای از جملات ارائه می‌شود که در آن، یک عنصر بنیادین (هسته) و تعدادی عنصر وابسته که به آنها متمم یا موضوع می‌گویند، ایفای نقش می‌کنند. نوع و تعداد متمم‌ها توسط ظرفیت منسوب به هسته تعیین می‌شود. در بسیاری از زبان‌ها، متمم‌های عناصر دارنده ظرفیت در فرهنگ‌های عمومی یا ظرفیتی فهرست می‌شوند. فرهنگ‌های ظرفیتی که فهرست کاملی از عناصر دارای ظرفیت به همراه متمم‌هایشان (اعم از اجباری و اختیاری) به دست می‌دهند، می‌توانند در زمینه‌های متفاوتی مورد استفاده‌های زبان‌شناختی قرار گیرند. از جمله این استفاده‌ها می‌توان به کاربرد آنها در آموزش زبان به سخن‌گویان سایر زبان‌ها اشاره کرد. متأسفانه زبان فارسی تاکنون از چنین فرهنگ‌هایی بی‌بهره بوده و آموزش آن زبان به غیرفارسی‌زبانان، به دلیل کمبود چنین مواد آموزشی‌ای، با دشواری‌های فراوانی روبه‌رو بوده است. مقاله حاضر، شرحی است از تولید نخستین فرهنگ ظرفیتی افعال فارسی که بخشی از پروژه دادگان درختی وابستگی است. این فرهنگ، با بهره‌گیری از مفاهیم نظریه دستور وابستگی و شیوه‌های استخراج رایانه‌ای تهیه شده و حاوی 4282 سرواژه فعلی منحصر به فرد و 5429 جفت فعل - متمم است.

کلیدواژه‌ها: فرهنگ ظرفیت، دستور وابستگی، دادگان درختی وابستگی، زبان فارسی.

* دانشجوی دوره دکتری زبان‌شناسی همگانی، دانشگاه تهران

** دانشجوی دوره دکتری زبان‌شناسی همگانی، دانشگاه تربیت مدرس

*** دانشجوی دوره دکتری علوم رایانه‌ای، دانشگاه کلمبیا

1- مقدمه¹

فرهنگ‌های لغت یکی از منابع ارزش‌مند در آموزش زبان به سخن‌گویان سایر زبان‌ها، زبان‌شناسی رایانشی و پردازش زبان طبیعی به‌شمار می‌آیند. یکی از انواع فرهنگ‌ها، فرهنگ‌های ظرفیتی هستند که در آنها کلمات زبان به همراه ظرفیت‌ها (valency) یا متمم‌هایشان (complement) (اعم از اختیاری و اجباری) فهرست شده‌اند. علاوه بر افعال که همگی ظرفیت دارند، برخی اسم‌ها و صفت‌ها نیز دارای ظرفیت هستند که به آنها اسم‌های گزاره‌ای و صفت‌های گزاره‌ای می‌گویند (کریمی دوستان: 198-200). در مقاله حاضر که شرحی از مراحل تهیه فرهنگ ظرفیتی افعال زبان فارسی است، بر معرفی ظرفیت افعال زبان فارسی، مراحل تولید فرهنگ نام‌برده و فواید اطلاع از آن برای غیرفارسی‌زبانان تأکید شده است. ظرفیت از آن جهت اهمیت دارد که ساخت بنیادین جمله‌های هر زبان را تعیین می‌کند و منظور از ساخت بنیادین، عناصری از جمله‌اند که حضورشان به واسطه ویژگی‌های منحصر به فرد فعل مرکزی جمله ضروری (در مورد ظرفیت‌های اجباری) یا توجیه می‌شود (در مورد ظرفیت‌های اختیاری). مفهوم ظرفیت از دستور وابستگی (Dependency Grammar) نشأت می‌گیرد؛ این دستور نظریه‌ای عمدتاً نحوی است که بر اساس آن، ساخت نحوی جملات زبان توسط رابطه یا روابط میان یک هسته و وابسته‌های آن تعیین می‌شود. در زبان فارسی، تعدد افعال مرکب، نیاز به فهرستی جامع از افعال و ظرفیت‌هایشان را افزایش می‌دهد. یکی از دلایل ضرورت چنین فهرستی، نیاز غیرفارسی‌زبانان به آگاهی از متمم‌های خاص هر یک از افعال زبان است. دلیل دیگر این است که شناسایی افعال مرکب و متمم‌هایشان در صورت نبود فهرست کاملی از آنها، برای غیرفارسی‌زبانان بسیار دشوار می‌شود. غیرفارسی‌زبانی که با جمله‌ای حاوی یک فعل و عنصری مشکوک به فعل‌یار یا مفعول مواجه می‌شود، باید تشخیص دهد که عنصر غیرفعلی موجود در جمله فعل‌یار است و در نتیجه توالی آن با فعل، فعل مرکب می‌سازد یا مفعول است و توالی آن با فعل، فعل مرکب تشکیل نمی‌دهد. تمایز میان این دو از آن

روی حائز اهمیت است که برخی فرایندهای نحوی زبان فارسی نسبت به تمایز پیش‌گفته حساسند.

در بخش 2، به معرفی مفهوم ظرفیت و دستور وابستگی به عنوان نظریه مورد استفاده در تهیه فرهنگ ظرفیتی فعل می‌پردازیم و تفاوت ظرفیت و ساخت بنیادین را توضیح می‌دهیم. در بخش 3، به پیشینه پژوهش‌های مشابه در زبان‌های دیگر پرداخته، معایب فرهنگ‌های لغت موجود فارسی را بررسی می‌کنیم. دلایل انتخاب دستور وابستگی، چالش‌های تهیه فرهنگ، و معیارهای تشخیص فعل مرکب را در بخش 4 اجمالاً مرور می‌کنیم و سپس انواع طبقه‌بندی‌های موجود از ساخت‌های ظرفیتی فعل در زبان فارسی را از دیدگاه سایر پژوهشگران معرفی کرده، طبقه‌بندی مورد پذیرش نگارندگان از انواع ساخت‌های ظرفیتی فعل و ساخت‌های بنیادین جملات فارسی را بیان می‌کنیم. در بخش 5، مراحل تولید فرهنگ ظرفیت فعل را معرفی و نهایتاً در بخش 6، آمار به‌دست‌آمده از فرهنگ را ارائه داده، موضوعاتی را برای پژوهش‌های آتی پیشنهاد می‌کنیم.

2- ظرفیت و دستور وابستگی

دستور وابستگی، نوعی دستور صوری است که بیشترین شهرت خود را مرهون پیشرفت‌هایی است که در دهه 1950 میلادی، به‌ویژه در آثار زبان‌شناس فرانسوی، (Tesnière, 1893-1954) به‌ویژه (Tesnière, 1980) داشته است. این دستور انواع متفاوتی از وابستگی‌ها را میان عناصر ساختارهای نحوی برقرار و از این طریق روابط نحوی را تبیین می‌کند. این وابستگی‌های میان کلمات، روابطی دودویی‌اند و درخت‌های وابستگی را می‌سازند. (Kübler, et. al.: 2) ساخت نحوی توسط درخت‌های وابستگی بازنمایی می‌شود. درخت‌های وابستگی مجموعه‌هایی از گره‌ها (node) هستند که ارتباطات میان آنها، روابط ساختی را مشخص می‌کنند. هر درخت حاوی یک هسته و تعدادی وابسته است که هر یک رابطه‌ای خاص نسبت به هسته جمله دارند (Crystal: 137)؛ مثلاً در جمله «کتابی روی میز گذاشتم»، «گذاشتم» به عنوان فعل جمله، هسته «کتابی» و «روی» محسوب

می‌شود و رابطه‌های وابستگی میان فعل جمله و دو عنصر نام‌برده به ترتیب «مفعول» و «متمم حرف اضافه‌ای» نامیده می‌شوند. «روی» نیز به نوبه خود هسته «میز» به حساب می‌آید و رابطه وابستگی این دو «مفعول حرف اضافه» نام می‌گیرد.

2-1- تفاوت ظرفیت و ساخت بنیادین

«در دستور وابستگی فرض بر این است که هر جمله یک فعل مرکزی دارد، و ثانیاً براساس نوع و تعداد متمم‌های اجباری و اختیاری آن فعل مرکزی، می‌توان ساخت بنیادین جمله‌هایی را که آن فعل در آنها به کار رفته است تعیین کرد. در این معنا، ساخت بنیادین چیزی نیست مگر فعل مرکزی جمله به‌علاوه متمم‌های اجباری و اختیاری آن که میان || نشان داده می‌شود (طیب‌زاده: 17). از آنجا که فعل جمله مشخص‌کننده تعداد و نوع متمم‌های موجود یا محتمل است، در دستور وابستگی مرکز ثقل ساختاری جمله محسوب می‌شود؛ به عبارت دیگر، در هر جمله، غالباً فعل هسته و سایر عناصر به طور مستقیم یا غیرمستقیم وابسته فعل جمله به‌شمار می‌آیند. به مثال‌های زیر توجه کنید:

(1) سارا شیرینی تعارف کرد.

فعل مرکزی جمله (1) «تعارف کرد» است. متمم‌های فعل «تعارف کرد» که هسته جمله است، عبارتند از فاعل (سارا) و مفعول (شیرینی). ساخت بنیادین جمله فوق به صورت زیر نشان داده می‌شود:

||فاعل، مفعول||

جمله (1) را می‌توان بسط داد و به صورت (2) در آورد:

(2) سارا که لباس قرمز رنگی به تن داشت، در میهمانی دیشب شیرینی خاهه‌ای تعارف کرد. ساخت بنیادین جمله (2) هم‌چنان همان ||فاعل، مفعول|| است. در این جمله هم‌چنان فعل، هسته است و فاعل و مفعول وابسته‌های فعل محسوب می‌شوند. تنها تفاوت بین دو جمله (1) و (2) این است که در جمله (2)، فاعل و مفعول با وابسته‌هایی گسترش

یافته‌اند و افزوده «در میهمانی دیشب»، به جمله (1) اضافه شده است. به عبارت دیگر، در جمله (2) بند موصولی «که لباس قرمز رنگی به تن داشت» و صفت «خامه‌ای» به ترتیب وابسته فاعل و مفعول محسوب می‌شوند و به طور غیرمستقیم وابسته فعل هستند.

فعل «تعارف کرد» در واژگان، به یک فاعل و مفعول اجباری و متمم حرف اضافه‌ای اختیاری (با حرف اضافه به) نیاز دارد که با توجه به اختیاری بودن متمم حرف اضافه‌ای و امکان ظاهر نشدن آن در جمله، ساخت بنیادین جمله‌های به دست آمده از این فعل می‌تواند |فاعل، مفعول|| (جمله‌های (1) و (2)) یا |فاعل، مفعول، متمم حرف اضافه‌ای به‌ای|| (جمله (3) و (4)) باشد.

(3) سارا به دوستانش شیرینی تعارف کرد.

(4) سارا که لباس قرمز رنگی به تن داشت، در میهمانی دیشب به دوستانش شیرینی خامه‌ای تعارف کرد.

طیب‌زاده معتقد است ساخت ظرفیتی فعل نشان‌دهنده تعداد و نوع متمم‌های اجباری و اختیاری آن است که آنها را درون دو قلاب < > نمایش می‌دهند. (طیب‌زاده: 17) بنابراین، در واژگان زبان فارسی نوع و تعداد متمم‌های فعل «تعارف کرد» به صورت زیر مشخص می‌شود:

تعارف کرد: <فاعل، مفعول، متمم حرف اضافه‌ای به‌ای>

می‌توان چنین نتیجه گرفت که تعداد و نوع متمم‌های اجباری و اختیاری فعل درون هر جمله، ساخت بنیادین و تعداد و نوع متمم‌های اجباری و اختیاری فعل در واژگان ذهنی، ساخت ظرفیتی نامیده می‌شود.

3- پیشینه پژوهش

در این بخش از مقاله به بررسی پژوهش‌های انجام شده برای تهیه فرهنگ‌های ظرفیتی در زبان‌های دیگر می‌پردازیم. از آنجا که فرهنگ‌های لغت موجود در زبان فارسی

نشان‌دهنده ساخت ظرفیتی افعال نیستند تنها مروری گذرا بر برخی ویژگی‌های آنها خواهیم داشت.

یکی از نخستین کوشش‌ها در جهت ساخت فرهنگ‌های ظرفیتی، اثر گریشمان و دیگران² است. اثر دیگری که هدف مشابهی را پی می‌گیرد، متعلق به بالدوین و دیگران³ است که در آن یک واژه‌نامه دوزبانه برای ترجمه ماشینی ژاپنی به انگلیسی تهیه شده است. در پژوهش هرست و دیگران⁴ ظرفیت برخی افعال، اسم‌ها و صفات انگلیسی به همراه مثال‌ها و یادداشت‌هایی درباره ویژگی‌های معنی‌شناختی آنها جمع‌آوری شده است. معیار اصلی برای گزینش فعل‌ها، اسم‌ها و صفات گزاره‌ای فهرست‌شده، فراوانی و دشواری آنها برای زبان‌آموزان بیگانه بوده است (Fillmore, 2009: 55). پرویزل و دیگران⁵ بر اهمیت کاربرد فرهنگ‌های ظرفیتی در پردازش زبان طبیعی تأکید می‌کنند. فرهنگ ظرفیتی دیگری در اثر کورهونن و دیگران⁶ تولید شده که در آن، اطلاعات مربوط به زیرمقوله‌بندی برچسب (tag) خورده‌اند. (Fillmore: 55) منبع دیگر در این زمینه، فریم‌نت (FrameNet) فیلمور و دیگران⁷ است که در آن قسمت‌هایی از پیکره ملی بریتانیا (British National Corpus) با اطلاعاتی درباره انواع گروه‌های نحوی، معنی‌شناسی و نقش‌های دستوری برچسب خورده‌اند. پروژه دیگر در همین راستا ورب‌نت (VerbNet) کیپر و دیگران⁸ است که بر طبقه‌های فعلی معرفی‌شده توسط لوین⁹ استوار است. در ورب‌نت، ترجیحات گزینشی (Selectional preferences)، معانی و ساخت ظرفیتی افعال انگلیسی برچسب‌زنی شده‌اند. واژه‌نامه ظرفیتی دیگر در زبان انگلیسی، اثر سمکی و دیگران¹⁰ است که بر پایه دستور نقشی - زایشی (Functional Generative Grammar) تولید شده و از دادگان درختی وابستگی پراگ استخراج شده است.

آثار دیگری از این دست نیز با محوریت زبان‌هایی غیر از انگلیسی، به‌ویژه زبان چکی وجود دارند. در این زبان منابعی همچون ورب‌الکس (VerbaLex) هلاواچکوا و دیگران¹¹ موجودند که دارای برچسب‌هایی مربوط به ظرفیت افعالند. در اثر لویاتکوا و دیگران¹²

برچسب‌های ظرفیتی به اسم‌های دارای ظرفیت نیز گسترش یافته‌اند. کتنروا و دیگران¹³ با در نظر گرفتن نقش‌های معنایی در فریم‌نت به شیوه نیمه‌خودکار نقش‌های معنایی را به فرهنگ ظرفیتی چکی اضافه می‌کند. یک واژه‌نامه ظرفیتی دیگر در زبان چکی پی‌دی‌تی - ولکس (PDT-Vallex) هاجیچ و دیگران¹⁴ است که از دادگان درختی وابستگی پراگ بومووا و دیگران¹⁵ برگرفته شده است. ضمناً منابعی در زبان‌های دیگر وجود دارند که از آن جمله‌اند: فرهنگ زیرمقوله‌بندی (subcategorization) زبان فرانسوی مسیان و دیگران¹⁶، فرهنگ ظرفیت افعال رومانیایی باربو¹⁷، فرهنگ ظرفیت افعال عربی بلیچکی و دیگران¹⁸، فرهنگ کرواتی آگیچ و دیگران¹⁹، فرهنگ آلمانی هینریشس و دیگران²⁰ و بانک ظرفیت‌های روسی لیاشفسکایا²¹.

با بررسی شناخته‌شده‌ترین فرهنگ‌های لغت در زبان فارسی که عبارتند از فرهنگ دهخدا، فرهنگ معین، فرهنگ عمید و فرهنگ سخن، به این نتیجه می‌رسیم که همه آنها دارای نقصی مشابهند؛ به این معنی که در هیچ‌یک از آنها ساخت ظرفیتی فعل‌ها، اسم‌ها و صفت‌ها به شکلی مدون، ارائه نشده و کاربران‌شان مجبورند تنها با رجوع به مثال‌هایی که ممکن است حاوی برخی ظرفیت‌های نمودیافته کلمات متمم‌پذیر باشند، درباره نحوه کاربرد صحیح فعل‌ها و متمم‌هایشان دست به استنتاج‌هایی شهودی و شمی بزنند. اگرچه ممکن است کاربران فارسی‌زبان به واسطه برخورداری از شم زبان فارسی بتوانند استنتاج‌های صحیحی را شکل دهند، مشکل کاربران غیرفارسی‌زبان با اتکا به دانش ناکاملشان از زبان نام‌برده رفع نمی‌شود؛ ضمن این‌که در مورد نحوه استفاده صحیح از فعل‌های کم‌کاربرد زبان، سخن‌گویان بومی نیز دچار مشکل می‌شوند.

4- فرهنگ ظرفیتی فعل در زبان فارسی

با توجه به آنچه تاکنون ذکر شد، مشخص است که یکی از نیازهای عمده فارسی‌زبانان، غیرفارسی‌زبانان و متخصصان حوزه پردازش زبان طبیعی، دسترسی به پیکره‌ای با برچسب‌های نحوی است. یکی از ضروری‌ترین اقدامات در جهت تولید چنین پیکره‌ای،

تهیه فهرستی جامع از فعل‌های زبان فارسی به همراه متمم‌های اجباری و اختیاری آنهاست. علت آن است که برنامه رایانه‌ای مسئول تشخیص ساخت نحوی جملات، پس از یافتن فعل اصلی جمله می‌تواند با دسترسی به ساخت ظرفیتی آن، گزینه‌های خود را برای تشخیص هر یک از عناصر جمله محدود کند و تحلیلی دقیق‌تر از آن ارائه دهد.

دلایل انتخاب نظریه دستور وابستگی از میان تمامی نظریه‌های دستوری موجود، برای بازنمایی ساخت نحوی جملات زبان فارسی، یکی این بوده است که جابه‌جایی عناصر در جملات این زبان بسیار ساده‌تر از زبان‌هایی هم‌چون انگلیسی است که ترتیب خطی اکیدتری (strict word order) را بر عناصر جمله تحمیل می‌کنند. به مثال‌های زیر توجه کنید:

(5) علی دیروز در مدرسه کتاب را به دوستش داد.

(6) دیروز علی در مدرسه کتاب را به دوستش داد.

(7) در مدرسه علی دیروز کتاب را به دوستش داد.

(8) کتاب را علی دیروز در مدرسه به دوستش داد.

(9) به دوستش علی دیروز در مدرسه کتاب را داد.

مثال‌های بالا تنها تعدادی از جابه‌جایی‌های ممکن عناصر درون جمله فارسی را نشان می‌دهند. در نظریه‌ای هم‌چون دستور زایشی، برای تبیین جابه‌جایی‌های درون جملات نیاز به تعریف حرکت (movement) وجود دارد و از آن‌جا که تشخیص حرکت‌های مختلف با انگیزه‌های متفاوت برای ماشین بسیار دشوار است، نظریه دستور وابستگی انتخاب شده است تا تقدم و تأخر وابسته‌ها نسبت به هسته نیازمند یادگیری ماشینی نباشد. نبود ترتیب خطی اکید در زبان‌های دیگر نیز باعث انتخاب دستور وابستگی برای بازنمایی ساخت نحوی جملات آن زبان‌ها شده است که از آن جمله می‌توان به چکی (Bömová, et. al., 2003)، ترکی (Van der Beek, et. al., 2002)، دانمارکی (Kromann, 2003)، آلمانی (Oflazer, et. al., 2003) و عربی (Hajič, et. al., 2004) اشاره کرد.

دلیل دوم برای انتخاب دستور وابستگی برای بازنمایی ساخت نحوی جملات این است که نگارندگان مقاله حاضر، همچون کوپلر و دیگران (Kübler et. al.: 95) بر این باورند که شیوه دستور وابستگی برای نشان دادن روابط میان عناصر سازنده جملات زبان به واقعیت پردازش آنها در ذهن بشر نزدیک‌تر است و با شم زبانی سخن‌گویان هماهنگی بیشتری دارد.

دلیل سوم برای این انتخاب، سهولت تبدیل خودکار برچسب‌گذاری وابستگی به شیوه برچسب‌گذاری گروه‌بنیان (phrase-based) است؛ در حالی که عکس این موضوع به راحتی و به طور کامل امکان‌پذیر نیست. (Johnson: 175)

یکی از چالش‌های پیش رو در مسیر جمع‌آوری داده‌های خام برای پیکره وابستگی زبان فارسی، این بوده است که متوازن (balanced) ساختن داده‌ها برای افزایش ویژگی نمایندگی (representativeness) آنها همواره کار دشواری است. با توجه به اهمیت ایفای سهم فعل نسبت به ساخت کلی جمله، لازم است داده‌ها بر اساس توزیع فراوانی ظرفیت فعل متوازن شوند. فقدان هر گونه فرهنگ ظرفیتی برای فعل‌های زبان فارسی، به این دشواری دامن می‌زند.

دشواری دیگری که نگارندگان با آن مواجه بوده‌اند، نبود فهرستی جامع از فعل‌های مرکب زبان فارسی است که بر مبنای معیارهای زبان‌شناختی تهیه شده باشد. طباطبایی تعریفی زبان‌شناختی از فعل مرکب در فارسی ارائه می‌دهد:

فعل مرکب به فعلی اطلاق می‌شود که از دو واژه مستقل ترکیب یافته است؛ واژه اول اسم یا صفت یا قید و واژه دوم فعل است. مانند «اجرا کردن»، «دراز کشیدن»، و «حدس زدن». جزء اول را فعل‌یار و جزء دوم را همکرد می‌نامیم. از مجموع فعل‌یار و همکرد معنی واحدی دریافته می‌شود. هرگاه دو کلمه از این انواع، دو معنی را به ذهن القا کنند، یعنی هر یک از اجزاء معنی مستقل و اصلی خود را حفظ کرده باشد، اطلاق فعل مرکب به آن‌ها درست نیست. وجه مشخص همکردها این است که معنی اصلی خود را از دست

می‌دهند یا معنی‌شان کم‌رنگ می‌شود و عمدتاً همچون عنصری صرفی که به ترکیب هویت فعلی می‌بخشد، به کار می‌روند؛ برای مثال در فعل‌های «اتو کردن» و «رنگ کردن»، جزء فعلی از نظر معنی تهی است و نقش اصلی آن این است که به کل عبارت، هویت فعلی می‌دهد. (طباطبائی: 26-27)

همچنین طیب‌زاده معتقد است «رابطه میان اجزاء فعل‌یار و همکرد از نوع رابطه نحوی نیست، و هیچ‌یک از آنها وابسته یا متمم دیگری محسوب نمی‌شود». (طیب‌زاده: 84) بنابراین توالی‌های «غذا خورد» و «فیلم دید» فعل مرکب محسوب نمی‌شوند؛ زیرا در این ترکیب‌ها، «غذا» و «فیلم» مفعول‌هایی بدون نقش‌نما برای فعل‌های متعدی «خورد» و «دید» هستند.

نگارندگان با استفاده از ادبیات موجود درباره فعل مرکب و نیز افزودن معیارهایی جدید، از میان توالی‌هایی که با استخراج رایانه‌ای، نامزد قرارگرفتن در زمره فعل‌های مرکب بودند، فعل‌های مرکب واقعی را به طور دستی شناسایی و فهرست کردند. تنها به عنوان نمونه‌ای از اهمیت بازشناسی فعل مرکب از سایر توالی‌های دوکلمه‌ای برای آموزنده غیرفارسی‌زبان، می‌توان مثال‌های زیر را در نظر گرفت:

(10) او غذا را سرد کرد.

(11) او بازی را برگزار کرد.

در جمله (10)، «سرد» مسند و «کرد» فعل سببی است؛ در حالی که در جمله (11)، «برگزار» اگرچه به طبقه صفت تعلق دارد، نمی‌تواند به عنوان صفت پسین اسم به کار رود.

(12) غذای سرد را خوردم.

(13) بازی برگزار را دیدم.^{22*}

غیرفارسی‌زبان برای اطلاع از چگونگی کاربرد توصیفی صفات، نیازمند داشتن فهرستی از فعل‌های مرکب زبان است تا با مراجعه به آن، از استفاده توصیفی صفات به‌کاررفته در افعال مرکب دارای همکرد «کردن» در بافت‌های دیگر پرهیز کند.

4-1- انواع ظرفیت‌های فعل در زبان فارسی

همان‌طور که اشاره شد، ساخت ظرفیتی فعل به مجموع ظرفیت‌هایی که یک فعل می‌تواند بپذیرد، اطلاق می‌شود و ساخت بنیادین که مفهومی است عینی‌تر، به تحقق ظرفیت‌های فعل در جمله گفته می‌شود. تاکنون معدودی از زبان‌شناسان به بررسی انواع ساخت‌های ظرفیتی فعل در زبان فارسی پرداخته‌اند. پژوهش مهاجر قمی (1978) اولین اثری است که ساخت‌های ظرفیتی فعل فارسی را از دیدگاه نظریه ظرفیت واژگانی بررسی می‌کند. وی 10 نوع متمم را برای فعل‌های فارسی برمی‌شمرد که عبارتند از مفعول‌نومیناتیو، مفعول‌آکوزاتیو، مفعول‌گنتیو، مفعول‌داتیو، مفعول‌حرف‌اضافه‌ای، قید مقدار، قید جهت، شمار، مقایسه، و متمم فعلی. (طیب‌زاده: 51-52)

اثر دیگری که به متمم‌های نحوی افعال در فارسی می‌پردازد، متعلق به احدی (2001) است. وی 11 متمم نحوی را برای افعال ذکر می‌کند که عبارتند از فاعل، متمم مستقیم، متمم پیش‌اضافه‌ای، متمم اضافه‌ای، متمم اضافه‌ای رایبی، متمم پی‌بستی، متمم پی‌بستی، متمم مقداری، متمم اسمی، متمم صفتی، و متمم فعلی. (همان: 56-57)

طیب‌زاده نیز به متمم‌های نحوی فعل در زبان فارسی در چارچوب دستور وابستگی می‌پردازد و هشت نوع متمم را برای فعل فارسی شناسایی می‌کند: فاعل، مفعول، مفعول حرف‌اضافه‌ای، مفعول نشانه‌اضافه‌ای، بند متممی، مسند، تمیز، و متمم قیدی. نگارندگان ضمن پذیرش طبقه‌بندی طیب‌زاده (همان)، یک متمم دیگر را به فهرست مذکور اضافه کرده‌اند و آن «مفعول دوم» است. به مثال زیر توجه کنید:

(14) او کتاب را به من هدیه داد.

با توجه به این که به لحاظ معنی‌شناختی، هم «کتاب» و هم «هدیه» را می‌توان مفعول فعل «داد» در نظر گرفت، به وجود متممی تحت عنوان مفعول دوم قائل شده‌ایم. مفعول دوم فقط هنگامی تظاهر پیدا می‌کند که در جمله مفعولی به همراه «را» یا بدون آن نیز وجود داشته باشد؛ ضمناً از ویژگی‌های مفعول دوم این است که همیشه الزاماً بدون «را» تظاهر می‌شود.

در جمله (14)، کلمه «هدیه» مفعول دوم است و بدون «را» تظاهر پیدا کرده است. بدساختی جمله‌های زیر نشان می‌دهد که افزودن «را» به مفعول دوم ناممکن است:

(15) *علی کتاب را به مریم هدیه را داد.

(16) *علی کتاب به مریم هدیه را داد.

طیب‌زاده 23 ساخت بنیادین را برای جملات زبان فارسی برمی‌شمارد (همان) که نگارندگان چهار ساخت دیگر به آن مجموعه افزوده‌اند. جدول (1) ساخت‌های بنیادین را نشان می‌دهد که چهار ساخت پایانی توسط نگارندگان شناسایی شده‌اند.

جدول 1. ساخت‌های بنیادین ممکن در زبان فارسی

1.	فاعل
2.	تهی، بند
3.	فاعل، مفعول
4.	فاعل، متمم حرف اضافه‌ای
5.	فاعل، متمم نشانه اضافه‌ای
6.	فاعل، بند
7.	فاعل، مسند
8.	فاعل، متمم قیدی
9.	فاعل، مفعول، متمم حرف اضافه‌ای
10.	فاعل، مفعول، متمم نشانه اضافه‌ای
11.	فاعل، مفعول، بند
12.	فاعل، مفعول، تمیز
13.	فاعل، مفعول، مسند
14.	فاعل، مفعول، متمم قیدی
15.	فاعل، متمم حرف اضافه‌ای، متمم حرف اضافه‌ای
16.	فاعل، متمم حرف اضافه‌ای، متمم نشانه اضافه‌ای

17.	فاعل، متمم حرف اضافه‌ای، متمم قیدی
18.	فاعل، متمم حرف اضافه‌ای، بند
19.	فاعل، متمم حرف اضافه‌ای، تمیز
20.	فاعل، متمم نشانه اضافه‌ای، بند
21.	تهی، بند، متمم حرف اضافه‌ای
22.	فاعل، مفعول، متمم حرف اضافه‌ای، متمم حرف اضافه‌ای
23.	فاعل، متمم حرف اضافه‌ای، متمم حرف اضافه‌ای، متمم قیدی
24.	فاعل، مفعول، متمم حرف اضافه‌ای، مفعول دوم
25.	فاعل، متمم قیدی، متمم نشانه اضافه‌ای
26.	فاعل، مفعول، متمم حرف اضافه‌ای، تمیز
27.	فاعل، مفعول، متمم حرف اضافه‌ای، متمم قیدی

با توجه به جدید بودن چهار ساخت آخر، برای هریک به ترتیب نمونه‌ای ذکر می‌شود:

(17) مریم توپ را به مهرناز هدیه داد.

(18) علی اینجا ثبت نام کرد.

(19) علی حسین را به ما انسانی صادق معرفی کرد.

(20) علی کتاب را از مدرسه اینجا آورد.

5- مراحل تولید فرهنگ ظرفیت فعل

چندین مرحله پیش‌پردازش برای تولید فرهنگ ظرفیت فعل فارسی پشت سر گذاشته شد. در مرحله نخست یکه برنامه برچسب‌زنی اجزای سخن از داده‌های پیکره بی‌جن‌خان ساخته شد. در مرحله دوم، تمام متون پیکره بی‌جن‌خان و پیکره همشهری (Aleahmad, et. al., 2009) برچسب‌زنی و سرواژه‌یابی (lemmatization) شدند. در مرحله سوم، پس از سرواژه‌یابی افعال و تبدیل آنها به سرواژه‌های خود، بر اساس اثر (Rasooli, et. al., 2011)، نامزدهای فعل مرکب در فارسی شناسایی شدند. پس از تکمیل مراحل پیش‌گفته، نامزدهای فعل مرکب به طور دستی و توسط گروهی برچسب‌زن برای

اطمینان خاطر از صحت نتایج خروجی، بازبینی شدند. هم اکنون فرهنگ ظرفیت افعال فارسی به صورت یک برنامه رایانه‌ای و به عنوان یکی از مراحل تولید پیکره دادگان وابستگی آماده استفاده کاربران است.²³

6- نتیجه‌گیری و پژوهش‌های آتی

با افزودن اطلاعات معنی‌شناختی به افعال و اضافه کردن جملاتی برای مثال از پیکره به فرهنگ ظرفیت، نسخه فعلی را می‌توان غنی‌تر ساخت. این امکان نیز هست که فرهنگ نام‌برده را به شکل کتابی مدون و جامع روانه بازار کتاب ایران کرد؛ هم‌چنین می‌توان فرهنگ‌های ظرفیتی مشابهی را برای اسم‌ها و صفت‌های گزاره‌ای فارسی تهیه و بدین ترتیب منابع ارزشمندی برای علاقه‌مندان به فراگیری زبان فارسی و متخصصان حوزه پردازش زبان طبیعی فراهم کرد.

یافته‌های آماری این پژوهش را می‌توان بدین شکل گزارش کرد: تعداد افعال منحصر به فرد: 4282، تعداد ساخت‌های ظرفیتی: 5429، متوسط تعداد ساخت ظرفیتی به ازای هر فعل: 1/268، بیشینه تعداد ساخت ظرفیتی برای فعل: 5.

پی‌نوشت‌ها

1- پژوهش حاضر یکی از دستاوردهای پروژه «پیکره دادگان وابستگی» مرکز تحقیقات کامپیوتری علوم اسلامی (نور) است و توسط مرکز نامبرده تأمین مالی شده است.

2- Grishman, Ralph, Catherine Macleod, and Adam Meyers: *Complex Syntax: building a computational lexicon*.

3- Baldwin, Timothy, Francis Bond, and Ben Hutchinson: *A valency dictionary architecture for machine translation*.

4- Herbst, T., David Heath, and Ian F. Roe: *A Valency Dictionary of English: A Corpus-Based Analysis of the Complementation Patterns of English Verbs, Nouns, and Adjectives*.

5- Proisl, T., and B. Kabashi: *Using High-Quality Resources in NLP: The Valency Dictionary of English as a Resource for Left-Associative Grammars*.

6- Korhonen, Anna, Yuval Krymolowski, and Ted Briscoe: *A large subcategorization lexicon for natural language processing applications*.

7- Fillmore, C.J., C.R. Johnson, and M.R.L. Petruck: *Background to FrameNet*.

- 8- Kipper, K., H.T. Dang, and M. Palmer: Class-based construction of a verb lexicon.
- 9- Levin, B.: English verb classes and alternations.
- 10- Semecký, Jiří, and Silvie Cinková: Constructing an English valency lexicon.
- 11- Hlaváčková, Dana, Aleš Horák, and Vladimír Kadlec: Exploitation of the VerbaLex Verb Valency Lexicon in the Syntactic Analysis of Czech.
- 12- Lopatková, Markéta, Veronika Řezníčková, and Zdeněk Žabokrtský: Valency Lexicon for Czech: From Verbs to Nouns.
- 13- Kettnerová, Václava, Markéta Lopatková, and Klára Hrstková: Semantic Roles in Valency Lexicon of Czech Verbs: Verbs of Communication and Exchange.
- 14- Hajič, J., J. Panevová, Z. Urešová, A. Bémová, V. Kolárová, and P. Pajas: PDT-VALLEX: Creating a Large-coverage Valency Lexicon for Treebank Annotation.
- 15- Bömová, Alena, Jan Hajič, Eva Hajičová, and Barbora Hladká: The Prague Dependency Treebank: A three-level annotation scenario.
- 16- Messiant, Cédric, Anna Korhonen, and Thierry Poibeau: LexSchem: A Large Subcategorization Lexicon for French Verbs.
- 17- Barbu, Ana-Maria: First Steps in Building a Verb Valency Lexicon for Romanian.
- 18- Bieličický, V., and O. Smrz: Building the Valency Lexicon of Arabic Verbs.
- 19- Agić, Z., K. Ojat and M. Tadic: An experiment in verb valency frame extraction from Croatian Dependency Treebank.
- 20- Hinrichs, E., and H. Telljohann: Constructing a Valence Lexicon for a Treebank of German.
- 21- Lyashevskaya, O.: Bank of Russian Constructions and Valencies.

22- وجود نویسه * در ابتدای جملات به معنی بدساختی آنها است.

23- فرهنگ نامبرده در نشانی www.dadegan.ir قابل بارگیری است.

منابع

- بی جن خان، محمود: «نقش پیکره‌های زبانی در نوشتن دستور زبان: معرفی یک نرم‌افزار رایانه‌ای»، مجله زبان‌شناسی، سال نوزدهم، 2، 48-67.
- دهخدا، علی‌اکبر: فرهنگ متوسط دهخدا، تهران: دانشگاه تهران، مؤسسه انتشارات و چاپ، 1386.
- طباطبایی، علاء‌الدین: «فعل مرکب در زبان فارسی»، نامه فرهنگستان، دوره 2، شماره 7: 34-26.

طیب‌زاده، امید: ظرفیت فعل و ساخت‌های بنیادین جمله در فارسی امروز، تهران: نشر مرکز، 1385.

عمید، حسن: فرهنگ عمید، تهران: شهرزاد، 1389.

کریمی‌دوستان، غلامحسین: «اسامی و صفات گزاره‌ای در زبان فارسی» در دستور، 1386، شماره 3: 187-202.

معین، محمد: فرهنگ معین، تهران: بوستان توحید، 1383.

Agic, Z., K. Ojat, Tadic, (2010). "An experiment in verb valency frame extraction from Croatian Dependency Treebank". Paper read at Information Technology Interfaces (ITI), 2010 32nd International Conference on, 21-24 June 2010.

Ahadi, Shahram. (2001). *verbergänzungen und zusammengesetzte in Persischen, Eine Valenzthoretische Analyse*: Dr. Ludvig Reichert, Verlag, Wiesbaden.

Aleahmad, Abolfazl, Hadi Amiri, Masoud Rahgozar, and Farhad Orumchian. (2009). "Hamshahri: A Standard Persian Text Collection". *Knowledge-Based Systems* 22 (5):382-387.

Baldwin, Timothy, Francis Bond, and Ben Hutchinson. (1999). "A valency dictionary architecture for machine translation". In *Eighth International Conference on Theoretical and Methodological Issues in Machine Translation: TMI-99*. Chester, UK.

Barbu, Ana-Maria. (2008). "First Steps in Building a Verb Valency Lexicon for Romanian". In *Text, Speech and Dialogue*, edited by P. Sojka, A. Horák, I. Kopecek and K. Pala: Springer Berlin / Heidelberg.

Bielický, V., and O. Smrz. (2008). "Building the Valency Lexicon of Arabic Verbs". In *6th Conference on Language Resources & Evaluation (LREC 2008)*. Marrakech, Morocco.

Bömová, Alena, Jan Hajič, Eva Hajičová, and Barbora Hladká. (2003). "The Prague Dependency Treebank: A three-level annotation scenario". *Treebanks: Building and Using Parsed Corpora*:103-127.

Crystal, D. (2008). *Dictionary of Linguistics and Phonetics*. Oxford: Blackwell.

Fillmore, C.J. (2009). "A Valency Dictionary of English". *International Journal of Lexicography* 22 (1):55.

Fillmore, C.J., C.R. Johnson, and M.R.L. Petruck. (2003). "Background to FrameNet". *International Journal of Lexicography* 16 (3):235.

- Grishman, Ralph, Catherine Macleod, and Adam Meyers. (1994). "Complex Syntax: building a computational lexicon". In 15th conference on Computational linguistics - Volume 1. Kyoto, Japan: Association for Computational Linguistics.
- Hajič, J., J. Panevová, Z. Urešová, A. Bémová, V. Kolárová, and P. Pajas. (2003). "PDT-VALLEX: Creating a Large-coverage Valency Lexicon for Treebank Annotation". Paper read at The Second Workshop on Treebanks and Linguistic Theories .
- Hajič, Jan, Otakar Smrž, Petr Zemánek, Jan Šnidauf, and Emanuel Beška. (2004). "Prague Arabic Dependency Treebank: Development in data and tools". In Proceedings of the NEMLAR 2004 International Conference on Arabic Language Resources and Tools. Cairo, Egypt.
- Herbst, T., David Heath, and Ian F. Roe. (2004). *A Valency Dictionary of English: A Corpus-Based Analysis of the Complementation Patterns of English Verbs, Nouns, and Adjectives*. Edited by D. Götz. Vol. 40: Walter de Gruyter.
- Hinrichs, E., and H. Telljohann. (2009). "Constructing a Valence Lexicon for a Treebank of German". In Seventh International Workshop on Treebanks and Linguistic Theories. Utrecht, Netherlands.
- Hlaváčková, Dana, Aleš Horák, and Vladimír Kadlec. (2006). "Exploitation of the VerbaLex Verb Valency Lexicon in the Syntactic Analysis of Czech". In Text, Speech and Dialogue, edited by P. Sojka, I. Kopeček and K. Pala: Springer Berlin / Heidelberg.
- Johnson, Mark. (2007). "Transforming projective bilocal dependency grammars into efficiently parseable CFGs with unfold-fold". In Proceeding of the 45th Annual Meeting of the Association of Computational Linguistics. Prague, Czech Republic: Association of Computational Linguistics.
- Kettnerová, Václava, Markéta Lopatková, and Klára Hrstková. (2008). "Semantic Roles in Valency Lexicon of Czech Verbs: Verbs of Communication and Exchange". In Advances in Natural Language Processing, edited by B. Nordström and A. Ranta: Springer Berlin / Heidelberg.
- Kipper, K., H.T. Dang, and M. Palmer. (2000). "Class-based construction of a verb lexicon". In Seventeenth National Conference on Artificial Intelligence (AAAI-2000). Austin, Texas.
- Korhonen, Anna, Yuval Krymolowski, and Ted Briscoe. (2006). "A large subcategorization lexicon for natural language processing applications". In 5th International Conference on Language Resources and Evaluation (LREC'05).

- Kromann, M.T. (2003). "The Danish Dependency Treebank and the DTAG treebank tool". In Second Workshop on Treebanks and Linguistic Theories (TLT 2003). Växjö, Sweden.
- Kübler, Sandra, Ryan McDonald, and Joakim Nivre. (2009). *Dependency Parsing*. Edited by G. Hirst, SYNTHESIS LECTURES ON HUMAN LANGUAGE TECHNOLOGIES: Morgan & Claypool.
- Levin, B. (1993). *English verb classes and alternations*, Chicago. Chicago: Chicago University Press.
- Lopatková, Markéta, Veronika Řezníčková, and Zdeněk Žabokrtský. (2006). "Valency Lexicon for Czech: From Verbs to Nouns". In Text, Speech and Dialogue, edited by P. Sojka, I. Kopecek and K. Pala: Springer Berlin / Heidelberg.
- Lyashevskaya, O. (2010). "Bank of Russian Constructions and Valencies". In Seventh conference on International Language Resources and Evaluation (LREC'10). Valletta, Malta.
- Messiant, Cédric, Anna Korhonen, and Thierry Poibeau. (2008). LexSchem: "A Large Subcategorization Lexicon for French Verbs". In Language Resources and Evaluation Conference (LREC). Marrakech, Morocco.
- Mohadjer-Ghomi, Siamak. (1978). *Eine kontrastive Untersuchung der Satzbaupläne im Deutschen und Persischen*: Burg-Verlag.
- Oflazer, K., B. Say, D.Z. Hakkani-Tür, and G. Tür. (2003). "Building a Turkish treebank. Treebanks: Building and Using Parsed Corpora". 20:261-277.
- Proisl, T., and B. Kabashi. (2010). "Using High-Quality Resources in NLP: The Valency Dictionary of English as a Resource for Left-Associative Grammars". In Seventh conference on International Language Resources and Evaluation (LREC'10), edited by N. Calzolari, K. Choukri, B. Maegaard, J. Mariani, J. Odjik, S. Piperidis, M. Rosner and D. Tapias. Valletta, Malta: European Language Resources Association (ELRA).
- Rasooli, Mohammad Sadegh, Hesham Faily, and Behrouz Minaei-Bidgoli. (2011). "Unsupervised Identification of Persian Compound Verbs". In Submitted in 10th Mexican International Conference on Artificial Intelligence (MICAI).
- Semecký, Jiří, and Silvie Cinková. (2006). "Constructing an English valency lexicon". In Workshop on Frontiers in Linguistically Annotated Corpora 2006. Sydney, Australia: Association for Computational Linguistics.
- Tesnière, Lucien. (1980). *Grundzüge der Strukturalen Syntax*. Edited by H. V. U. Engel. Stuttgart: KlettCotta.
- Van der Beek, L., G. Bouma, R. Malouf, and G. Van Noord. (2002). "The Alpino dependency Treebank". Language and Computers 45 (1):8-22.