

به نام خدا



وزارت علوم، تحقیقات و فناوری
پژوهشگاه علوم و فناوری اطلاعات ایران (ایرانداک)

طرح پژوهشی

**تعیین قواعد صوری همگون گروه‌های
دستوری برای کاربرد در پردازش زبان فارسی**

مجری

سید مهدی سمائی

خرداد ۱۳۹۵

حقوق: پژوهشگاه علوم و فناوری اطلاعات ایران

طرح پژوهشی «تعیین قواعد صوری همگون گروه‌های دستوری برای کاربرد در پردازش زبان فارسی» پیرو قرارداد شماره ۳۳/۹۵۲۶ مورخ ۱۳۹۴/۸/۱۳ میان پژوهشگاه علوم و فناوری اطلاعات ایران (کارفرما) و آقای دکتر سیدمهدی سمایی اجرا شده است. گزارش حاضر یک جلد از مستندات این طرح است. این گزارش و تمامی حقوق مادی آن بر اساس قانون حمایت حقوق مولفان و مصنفان و هنرمندان، مصوب ۱۳۴۸ و اصلاحیه‌های بعدی آن و همچنین آیین‌نامه‌های اجرایی این قانون متعلق به پژوهشگاه علوم و فناوری اطلاعات ایران و هرگونه استفاده از تمامی یا پاره‌ای از آن، شامل: نقل قول، تکثیر، انتشار، کاربرد نتایج، تکمیل و مانند آنها به صورت چاپی، الکترونیکی یا وسایل دیگر فقط با اجازه کتبی پژوهشگاه امکان‌پذیر است. نقل قول در حد هزار واژه در انتشارات علمی مانند کتاب و مقاله با درج اطلاعات کامل کتابشناختی، نیازی به مجوز پژوهشگاه ندارد.

صحت مندرجات گزارش بر عهده مجری طرح پژوهشی است.

صفحه تأييديه ناظر

عنوان طرح

مدیر طرح پژوهشی: سیدمهدی سمایی، پژوهشگاه علوم و فناوری اطلاعات ایران.

نشانی: تهران، خیابان انقلاب، چهارراه فلسطین، شماره ۱۰۹۰، طبقه سوم

صندوق پستی: ۱۳۱۵۷۷۳۳۱۴ تلفن: ۰۲۱-۶۶۴۹۴۹۸۰ (داخلی ۳۲۳)

وبگاه: <http://www.irandoc.ac.ir/Irandoc/files/file/CV/Faculty/samai-resume.pdf>

رایانامه: Samai@irandoc.ac.ir

چکیده

پردازش زبان یکی از حوزه‌های پژوهشی هوش مصنوعی است. ترجمه ماشینی و درک متن دو زمینه اصلی پژوهش در حوزه پردازش زبان‌اند. در دست داشتن قواعد صوتی زبان و تبدیل آن‌ها به قواعد و کدهای قراردادی از ملزومات و داده‌های اصلی در این گونه پژوهش‌ها است. استخراج قواعد دستوری گروه‌های دستوری و عرضه کدها و قواعد قراردادی مقدمه‌ی توصیف صوتی دستور زبان است. تا کنون گروه‌های دستوری زبان فارسی به طور جداگانه بررسی و برای آن‌ها کدها و قواعد قراردادی عرضه شده است. در این طرح کدهایی که تا کنون برای گروه‌های دستوری در طرح‌های جداگانه انتخاب شده بررسی و همگون شده و فهرست درهم کرد و جامع آن‌ها عرضه شده است.

کلیدواژه‌ها: پردازش زبان، هوش مصنوعی، دستور زبان فارسی، قواعد صوتی، گروه دستوری

فهرست مطالب

صفحه	عنوان
	چکیده
	فصل اول
۲	۱-۱- مقدمه
۳	۲-۱- تعریف مسئله
۳	۳-۱- اهداف
۴	۴-۱- تعریف اصطلاحات
۵	۵-۱- پیشینه تحقیق
۶	۱-۵-۱- واژگان در دستورسنگ
۷	۱-۵-۲- واژه شکن فارسی
۹	۱-۵-۳- فعل شکن فارسی
۹	۱-۵-۴- عددشکن فارسی
۱۰	۱-۵-۵- وابسته های پسین گروه اسمی
۱۱	۱-۵-۶- وابسته های پیشین گروه اسمی
	فصل دوم
۱۳	۱-۲- گروه قیدی
۱۴	۲-۲- گروه صنعتی
۱۵	۲-۳- همگون کردن کدهای گروه های دستوری
۱۵	۲-۳-۱- کدهای واژه شکن
۱۶	۲-۳-۲- کدهای وابسته های پیشین و وابسته ای پسین
۱۷	۲-۳-۳- کدهای فعل شکن
۱۸	۲-۳-۴- کدهای عددشکن
۱۸	۲-۳-۵- مقولات واحدها و کدهای متفاوت
۲۰	۲-۳-۶- کدهای درهم کرد

فهرست مطالب

صفحه	عنوان
	فصل سوم
۲۸	خلاصه و نتیجه گیری
۲۹	کتابنامه
۳۱	پیوست ها

فصل اول

۱-۱- مقدمه

ارتباط زبان و رایانه از اواخر دهه چهل میلادی و با تحقیق درباره ساز و کار ترجمه ماشینی آغاز شد.^۱ در آن دوره رایانه را فرهنگ لغتی الکترونیکی می‌پنداشتند که کلمه‌ای را به کلمه‌ای دیگر ترجمه می‌کرد. از دهه پنجاه میلادی و با چامسکی تحلیل‌های نحوی مستقل از معنی آغاز شد. طبق نظر او تحلیل زبان آدمی بدون استفاده از معنی امکانپذیر بود. از دهه شصت میلادی و با نظریه‌های کتز و فودور (۱۹۶۳) و کتز و پستال (۱۹۶۴) معنی‌شناسی نیز وارد کار پردازش زبان شد. در مرحله بعد کاربرد شناسی^۲ به تحلیل‌های زبانی راه یافت. انسان دانشی از عالم واقع دارد که آن را در تعبیر متن به کار می‌بندد. از این پس نظریه‌های گوناگونی از شکل تجربه انسان از عالم خارج ارائه شد. بدین ترتیب متخصصان در آن زمان تصمیم گرفتند که رایانه را با دانش عالم خارج تجهیز کنند. اما چنین دانشی وسیع و نامحدود است و به فرض اینکه ذخیره کردن آن در حافظه رایانه مقدور باشد برای جستجوی آن باید وقت بسیار صرف کرد. عاقبت به این نتیجه رسیدند که بهتر است دانش عالم خارج را محدود کنند. چنین روشی واقع‌بینانه و شدنی و کاربردی است زیرا وقتی انسان نیز با متنی روبرو می‌شود از دانش خاص مربوط به عالم همان متن برای تعبیر آن استفاده می‌کند. به این سبب به جای ذخیره کل دانش در حافظه رایانه مسئله عالم فرعی^۳ و نمایشنامه^۴ و فیلمنامه به وجود آمد که همگی شکل‌های مختلف ذخیره دانش عالم خارج در رایانه است.

زبان‌شناسی رایانه‌ای یک از شاخه‌های زبان‌شناسی است که در آن فنون و مفاهیم علم رایانه را در تحقیقات حوزه‌های مختلف به کار می‌برند. دو حوزه اصلی زبان‌شناسی رایانه‌ای ترجمه ماشینی^۵ و دیگری درک متن^۱ است.

۱. طرح حاضر یکی از مجموعه طرح‌هایی است که نگارنده از سال ۱۳۷۷ تا کنون (مجموعاً ۸ طرح) در پژوهشگاه علوم و فناوری ایران انجام داده است. از آنجاییکه موضوع کلی طرح‌ها یکسان بودن است و هدف مشخصی داشته است مقدمه نیز با تغییراتی از طرح‌های پیشین اقتباس شده است.

2. Pragmatics

3. subworld

4. Scripts

5. Mashin translation

ترجمه ماشینی و درک متن نوعی شبیه‌سازی از مغز انسان است و از این رو یکی از شاخه‌های هوش مصنوعی^۲ است. اساس کار ماشین در این دو نوع فعالیت بر تحلیل داده‌ها است. ماشین مفروض با داده‌هایی که در اختیار دارد متن را ترجمه یا تحلیل و فهم می‌کند و خروجی را در اختیار کاربر قرار می‌دهد. جدا کردن حوزه‌های زبان و توصیف آنها مقدمه ساختن قواعد است. ماشین مذکور باید قواعد زبان مورد تحلیل را در اختیار داشته باشد. قواعدی صوری و استثناء ناپذیر. قواعد زبانی باید دقیق و متقن باشد. استفاده از معنی در ترجمه ماشینی و درک متن در نهایت ضروری است ولی قاعده بعضی از بخش‌های زبان را می‌توان فارغ از معنی عرضه کرد و در حافظه رایانه ذخیره نمود.

۱-۲- تعریف مسئله

منطق و عملکرد پردازش زبان - چه در ترجمه ماشینی و چه در درک متن - ایجاب می‌کند که قواعد دستوری زبان استخراج و در حافظه رایانه ذخیره شود. قواعد استخراجی قواعدی صوری و بدون استثناء و ماشین خوان است که شم زبانی اهل زبان در تحلیل و فهم آن دخیل نیست و صرفاً برای فهم ماشین استخراج شده است. گستردگی این قواعد ایجاب می‌کند که قواعد حوزه‌های مختلف دستور زبان ابتداء تقسیم‌بندی و حوزه‌بندی شوند و سپس جداگانه استخراج گردند. در طرحهایی که تا کنون در پژوهشگاه علوم و فناوری اطلاعات ایران (ایرانداک) در حوزه پردازش زبان انجام شده کدها و قواعد صوری حوزه کلمات مرکب و مشتق و گروه افعال و قواعد ترکیب اعداد و وابسته‌های پیشین و وابسته‌های پسین گروه اسمی استخراج شده‌اند. در طرح حاضر قواعد صوری گروه صنعتی و گروه قیدی نیز استخراج می‌شوند و در نهایت کدها و قواعد استخراجی همه طرح‌ها با یکدیگر مقایسه و همگونه خواهند شد.

۱-۳- اهداف

الف. همگون کردن کدها و قواعد صوری گروه‌های دستوری زبان فارسی

ب. عرضه مجموعه کدها و قواعد گروه‌های دستوری زبان فارسی در طرح‌های اجرا شده در ایرانداک

۱-۴- تعریف اصطلاحات

الف. پردازش زبان^۱

فعالیتی است که در حوزه زبانشناسی رایانه‌ای انجام می‌شود و در آن از فنون و تکنولوژی رایانه‌ای در تحلیل زبان استفاده می‌شود.

ب. ترجمه ماشینی

ترجمه خود کار متن زبانی به زبان دیگر. زبانی که ترجمه می‌شود زبان مبدا و زبانی که ترجمه به آن تبدیل می‌شود زبان مقصد نام دارد.

ج. درک متن

فهم و ادراک متن زبانی مفروض به صورت خودکار و ماشینی است.

د. گروه اسمی

گروهی نحوی است که از یک کلمه یا گروهی از کلمات ساخته می‌شود. هنگامی که گروه اسمی با گروهی از کلمات ساخته شده باشد یک هسته و تعدادی وابسته دارد. هسته گروه اسمی معمولاً اسم یا ضمیر است. وجود هسته اجباری و وابسته‌ها اختیاری است. وابسته‌ها چنانچه قبل از هسته قرار گیرند وابسته پیشین و اگر بعد از هسته بیایند وابسته پسین نام دارند.

هـ. گروه فعلی

گروهی نحوی است که از یک فعل یا یک فعل و متمم‌هایی ساخته می‌شود.

و. گروه صفتی

گروهی نحوی است که هسته آن صفت است. متمم هسته در این گروه نحوی ممکن است اسم یا گروه اسمی یا گروه حرف اضافه‌ای باشد.

ز. گروه قیدی

گروه نحوی است که هسته آن قید است. متمم قید در این گروه نحوی ممکن است صفت یا گروه صفتی یا قیدی دیگر باشد.

ح. قواعد صوری

قواعدی‌اند که به زبانی قراردادی نوشته می‌شوند تا امکان درج آن در رایانه فراهم شود. این قواعد استثناء ناپذیر و ماشین خوان هستند.

۱-۵- پیشینه تحقیق

پیشینه استخراج قواعد صوری دستور زبان فارسی برای کاربرد در رایانه به طرح‌هایی بر می‌گردد که در فاصله سالهای ۱۳۸۵ تا ۱۳۹۳ در پژوهشگاه علوم و فناوری اطلاعات ایران (ایرانداک) انجام پذیرفته است. طرح‌های مذکور به قرار ذیل است:

- واژه‌شکن زبان فارسی

- عددشکن زبان فارسی

- فعل‌شکن زبان فارسی

- تعیین قاعده‌های صوری وابسته‌های پیشین گروه اسمی و کاربرد آن در ترجمه ماشینی

- تعیین قاعده وابسته‌های پسین گروه اسمی (تمهیداتی زبانی برای پردازش زبان)

به جز این پنج طرح می‌توان به دو طرح دیگر نیز اشاره کرد که به زمینه و موضوعی مشابه انجام شده‌اند. طرح نخست واژگان در «دستورسنج»^۱ و طرح دیگر «شناسایی طرح‌های مرتبط با تحقیق ترجمه ماشینی در زبان فارسی» بوده است.

طرح‌های یاد شده پژوهش‌هایی‌اند که مستقیماً به موضوع طرح حاضر مربوط‌اند. در طرح «تعیین قواعد صوری همگون گروه‌های دستوری برای کاربرد در پردازش زبان فارسی» علاوه بر اینکه قواعد صوری دو گروه صفتی و قیدی استخراج خواهد شد به نوعی سامان‌دهی طرح‌های قبلی و عرضه قواعد صوری همگون برای همه طرح‌هایی است که تا کنون با هدف مشابهی در پژوهشگاه علوم و فناوری اطلاعات ایران انجام شده‌اند.

اینک به بررسی اجمالی طرح‌های سابق‌الذکر می‌پردازیم:

۱-۵-۱- واژگان در دستورسنج

دستورسنج برنامه‌ای رایانه‌ای است که با آن می‌توان ترکیبهای دستوری و غیر دستوری زبان فارسی را از هم تمیز داد. ساز و کار دستورسنج بر طبقه‌بندی دقیق مقوله‌های زبان از روی رفتاری نحوی و صرفی‌شان قرار دارد. مقوله‌های زبان و زیر مقوله‌های آنها با ویژگی‌های صرفی و نحوی مربوط در واژگان^۱ درج می‌شوند. هر یک از این مقوله‌ها حوزه قواعد دارند که در آنها قاعده‌های مربوط به رفتار هر مقوله ثبت شده است. دستورسنج پس از شناسایی مقوله‌ها در واژگان رفتار آنها را با قاعده‌های حوزه مذکور تطبیق می‌دهد و ترکیب دستوری را از ترکیب غیر دستوری متمایز می‌کند.

حوزه‌هایی نیز برای گشتاور و حرف و تطابق نهاد و گزاره و املاء وجود دارد. ساز و کار دستورسنج بر اساس الگوریتم^۲ است و از معنی و اطلاعات بینشی^۳ در آن استفاده نمی‌شود. دستورسنج را می‌توان به بیماران دچار اگنوربا^۴ تشبیه کرد که فقط با استقراء می‌توانند واقعات را تحلیل کنند.

دستورسنج از دو بخش اصلی تشکیل می‌شود. واژگان و حوزه قواعد. در طرح دستورسنج شانزده جدول برای مقولات دستوری انواع فعل و قواعد نحوی و شناسه‌ها و وندها و علائم سجاوندی و عناصر منفرد عرضه شده است. نمونه‌ای از جدول‌ها عرضه شده:

1. lexicon
2. algorithm
3. heuristic
4. agnosia

جائدار / بی جان	اسم معنی / ذات	جنس دستوری	اسم عام / اسم جمع	شمارش پذیر / اسم جنس	با قاعده / بی قاعده	مفرد جمع	شماره معنی	شماره همنویسه	کلمه
جائدار	ذات	مونث	عام	شمارش پذیر	باقاعده	مفرد	1	1	hen
بی جان	ذات	بی جنس	عام	اسم جنس	باقاعده	مفرد	1	1	pasta
انسان	ذات	بی جنس	عام	شمارش پذیر	بی قاعده	مفرد	1	1	person
بی جان	ذات	بی جنس	عام	شمارش پذیر	باقاعده	جمع	1	0	scissors
جدول شماره ۱-۱									

مدخل	ویژگی	ویژگی	ویژگی
پسر	جائدار	بی نشان	شمارش پذیر
خوردن	بی جان	مصدر	شمارش پذیر
درخت	جائدار	بی نشان	شمارش پذیر
دست	جائدار	بی نشان	شمارش پذیر
سیب	بی جان	بی نشان	شمارش پذیر
علی	خاص	شمارش ناپذیر	شمارش پذیر
فرار	بی جان	عمل	شمارش پذیر
همسایه	اسم جائدار	بی نشان	شمارش پذیر
مردم	اسم		شمارش پذیر
جدول شماره ۲-۱			

۱-۵-۲- واژگان فارسی

مهمترین بخش پردازش زبان شامل درک متن و تشخیص گفتار^۱ و ترجمه ماشینی است. درک متن گاهی در سطح نحو است. رایانه در این حالت متن را در سطح جمله و بند و کل متن تحلیل می‌کند. از درک متن در ترجمه ماشینی نیز استفاده می‌شود. تحقیق در این بخش از پردازش زبان از اواخر دهه هشتاد میلادی آغاز شد. متن معنی گسترده‌ای دارد. در سطحی عام حتی نشانه‌های راهنمایی و رانندگی را متن می‌نامند (کریستال، ۱۹۹۲). با این نگاه هر لفظ را می‌توان متن نامید و درک مفهوم متن را به حوزه کلمه نیز تعمیم داد.

کلمات یا بسیط‌اند یا پیچیده. یعنی یا ساده‌اند و از اجزاء معنی‌داری به نام تکواژ^۱ ساخته نشده‌اند و یا از اجتماعی تکواژها ساخته شده‌اند.

تحلیل کلمات در سطح مربوط به حوزه حرف است. کلمات پیچیده را می‌توان به دو دسته کلی تقسیم کرد:

الف. کلماتی که با اشتقاق^۲ ساخته می‌شوند.

ب. کلماتی که با ترکیب^۳ ساخته می‌شوند.

کلمات مشتق از یک پایه^۴ به اضافه‌وند ساخته می‌شوند. کلمات مرکب از اجتماع اسم و صفت و قید و بن فعل با یکدیگر به وجود می‌آیند.

واژه‌شکن فارسی برای تحلیل کلمات پیچیده ساخته شده است.

این ماشین کلمات مشتق و مرکب را تجزیه و مقوله آن‌ها را مشخص می‌کند.

نمونه‌ای از فرمول‌ها و قواعد واژه‌شکن از این قرار است:

formula descriptions	
code	description
s31	پسوند نسبت ۱
s32	پسوند جنس
s33	پسوند فاعلی
s34	پسوند عمل ۲
s35	پسوند مصدر ساز
s36	پسوند فاعلی ۱
s37	پسوند مفعولی
s38	پسوند کنندگی ۲
جدول شماره ۱-۳	

۱-۵-۳- فعل شکن فارسی

فعل شکن زبان فارسی برنامه‌ای رایانه بنیاد است که صیغگان فعل‌های بسیط زبان فارسی را تجزیه و شناسایی می‌کند. فعل شکن قادر به شناسایی و تجزیه ساختمان انواع فعل‌های ساده با قاعده و بی‌قاعده زبان فارسی است. نمونه‌ای از زبان قاعده‌وار و میانجی برای کدگذاری ساختمان افعال زبان فارسی از این قرار است:

شناسه های مضارع HX					
م	اول شخص مفرد	H1	یم	اول شخص جمع	H4
ی	دوم شخص مفرد	H2	ید	دوم شخص جمع	H5
د	سوم شخص مفرد	H3	ند	سوم شخص جمع	H6
جدول شماره ۱-۴					

شناسه های ماضی GX					
م	اول شخص مفرد	G7	یم	اول شخص جمع	G4
ی	دوم شخص مفرد	G8	ید	دوم شخص جمع	G5
	سوم شخص مفرد	G9	ند	سوم شخص جمع	G6
جدول شماره ۱-۵					

۱-۵-۴- عدد شکن فارسی

زبان فارسی همچون بسیاری از زبانهای دیگر دارای قواعد خاصی برای ترکیب اعداد است. گروه اسمی که پایه و اساس نحو فارسی است از یک اسم یا ضمیر به نام هسته و تعدادی وابسته ساخته می‌شود. این وابسته‌ها یا پیش از هسته قرار می‌گیرند یا پس از هسته. این وابسته‌ها را بسته به جایگاه ماقبل و مابعدشان نسبت به هسته وابسته‌های پیشین در گروه اسمی اعداد هستند که با قاعده خاصی با هم ترکیب می‌شوند. در حوزه‌های پردازش زبان نظیر ترجمه ماشینی و درک متن رایانه باید قواعد ترکیب اعداد را شناسایی کند و قادر باشد آنها را تجزیه نماید و ترکیبهای درست را از نادرست تمیز دهد.

عدد شکن فارسی برنامه‌ای رایانه‌ای است که قادر به پردازش قواعد ترکیب اعداد زبان فارسی است و می‌تواند صورتهای صحیح را از صورتهای غیر دستوری تمیز دهد.

نمونه‌ای از قواعد ترکیب اعداد در زبان فارسی:

A	B	C	D	E
الف	ج+عطف+الف	A + ه	A + و	A + ز
ب	د+عطف+الف	A+ه+عطف+A	A+و+عطف+A	A+ز+عطف+A
ج	د+عطف+ب		B+و+عطف+B	B+ز+عطف+B
د	د+عطف+ج	ه+B	A+و+عطف+C	A+ز+عطف+C
			B+و	A+ز+عطف+D
جدول شماره ۶-۱				

۱-۵-۵- وابسته‌های پسین گروه اسمی

بخشی از وابسته‌ها در گروه اسمی زبان فارسی که پس از هسته قرار می‌گیرند علائم جمع و وابسته‌های پسین نام دارند. این وابسته‌ها شامل انواع صفت و اسم و ضمیر و ترکیب اعداد و جمله موصولی‌اند که با قواعد خاصی پس از هسته می‌آیند و تابع محدودیت‌هایی هستند.

هدف از کد گذاری تبدیل قاعده‌های زبانی به قاعده‌هایی صوری و قراردادی است. با کد گذاری قواعد می‌توان به زبانی واسط و مشاین خوان دست یافت که میانجی بین زبان طبیعی و ماشین است. تبدیل دستور زبان فارسی به این زبان صوری مفروض راه رسیدن به پردازش کامل زبان فارسی است.

اینک نمونه‌ای از قواعد وابسته‌های پسین گروه اسمی زبان فارسی:

مثال	قاعده
کتاب مفید	ni+A
شاگرد درس‌خوان	n+A
سپاه آموزش دیده	nl+A
کتاب مدرسه	ni+ni
سپر علی	n+np
سپاه شهر	nl+ni
کتاب هزار و چهارصد و هفتم	ni+ngo
کتابی که دیروز خریدی	ni+nph
شاگردی که دیروز دیدی	n+nph
جدول شماره ۷-۱	

۱-۵-۶- وابسته‌های پیشین گروه اسمی

وابسته‌های پیشین گروه اسمی وابسته‌هایی هستند که پیش از هسته گروه اسمی قرار می‌گیرند. این وابسته‌ها عبارتند از: صفت اشاره بسیط، صفت اشاره مرکب، صفت مبهم بسیط، صفت مبهم مرکب، صفت پرسشی بسیط، صفت پرسشی مرکب، ترکیب اعداد، ممیز، صفت عالی. دو گونه محدودیت بر قرار گرفتن این وابسته‌های قبل از هسته گروه اسمی زبان فارسی مترتب است:

الف. این وابسته از لحاظ تقدم و تأخر محدودیت دارند.

ب. برخی از این وابسته‌ها مانعه‌الجمع‌اند و قرار گرفتن برخی از آنها قبل از هسته مانع از قرار گرفتن وابسته‌ای دیگر می‌شود.

نمونه‌ای از قواعد وابسته‌های پیشین گروه اسمی:

A10	صفت اشاره
A10a	صفت اشاره بسیط
A10b	صفت اشاره مرکب
A11	صفت مبهم
A11a	صفت مبهم بسیط
A11b	صفت مبهم مرکب
A12	صفت پرسشی
A12a	صفت پرسشی بسیط
A12b	صفت پرسشی مرکب
جدول شماره ۸-۱	

فصل دوم

۲-۱- گروه قیدی

تعریف سنتی قید از این قرار است:

«کلمه‌ای است که مضمون جمله، فعل، صفت، قید و کلمات دیگری غیر از اسم و جانشین اسم را مقید سازد، و یا حالت و هیأت فاعل، مفعول بیواسطه و فعل تام را در حین صدور فعل تعیین کند» (معین ۱۳۷۱) قیدی انواعی دارد. قید زمان و مکان و قید وصف و کیفیت. قید مختص و مشترک نیز وجود دارد. قید مختص فقط قید است مثل «پیوسته» و «ظالمانه». قیدیت مشترک «در غیر حالات قید نیز استعمال شود» مثل «خوب» و «بد». قید استفهام، تاکید، تشبیه، تمنی، و نفی از دیگر انواع قیداند (همان).

همسر و کورتمن (۲۰۰۶) نیز به قید گزاره‌ای^۱، قید درجه^۲، قید جمله، قید مکان و زمان، قید محدود کننده^۳ و قید متن^۴ قائل‌اند (قیدی که موجب انسجام^۵ بین جملات است).

گروه قیدی از گروه‌های اصلی در زبان نیست و شانی حاشیه‌ای دارد.

پوستت (۲۰۰۶) نیز گروه اسمی و فعلی را «اجزاء اصلی» کلام می‌داند.

به همین علت همگان بر وجود گروه قیدی به آن صورتی که در گروه اسمی و گروه فعلی وجود دارد تاکید نمی‌کنند و به آن نپرداخته‌اند. معین (۱۳۷۱) اشاره می‌کند که «قیدی بر سر قید یا قیود دیگر افزوده می‌شود». ماهوتیان (۱۹۹۷) نیز عقیده دارد که گروه قیدی علامت صرفی مشخص ندارد و سازه‌آرایی^۶ متمایزی نیز ندارد تا بتوان تعریفی صوری از آن عرضه کرد. گروه قیدی در زبان فارسی می‌تواند فعل و صفت و دیگر قیود را توصیف کند. برخی گروه‌های حرف اضافه^۷ نیز به صورت قید زمان و مکان و حالت به کار می‌روند (همان). بنابراین علاوه بر قید مختص و مشترکی که معین از آنها نام می‌برد گروه‌های حرف اضافه‌ای نیز وجود دارد که نقشی قید گونه در زبان فارسی می‌توانند داشته باشند.

1. predicate adverb

2. Degree adverb

3. Focalizer adverb

4. Text adverb

5. coherence

6. Word order

۷. گروه حرف اضافه‌ای همان گروه اسمی است که حرف اضافه‌ای بر سر آن درآمده باشد.

کریستال (۱۹۹۷) نیز به ترکیبات ساده‌ای مثل *very important* اشاره می‌کند و آن را گروه قیدی می‌داند و تراسک (۱۹۹۳) نیز گروه قیدی را گروهی می‌داند که هسته‌اش قید است. مانند *very quietly* بنابراین قواعد پیچیده و چند لایه‌ای که در مورد گروه‌های اصلی زبان نظیر گروه اسمی عرضه شده در مورد گروه قیدی نمی‌تواند ارائه شود. چنانچه قید را با «Adv» نشان دهیم، تنها قاعده مربوط به گروه قیدی قاعده *Adv + Adv* خواهد بود.^۱

۲-۲- گروه صفتی

به گفته پوست (۲۰۰۶) صفت نیز همانند اسم و فعل یکی از اجزاء اصلی کلام است. صفت از لحاظ معنایی در همه زبانها وجود دارد ولی ممکن است در همه زبانها صورت مستقل نداشته باشد. یعنی از لحاظ واژنحوی^۲ به شکل مستقل دیده نشود. بدین معنی که تابع ویژگیهای جنس^۳ و شمار شمار اسم باشد و متناسب با آن تصریف^۴ شود یا تابع شخص و شمار و زمان فعل باشد و با وند نشان داده شود. (تعریف با اسم و وندپذیری با فعل).

صفت ممکن است ویژگی مدرج بودن^۵ داشته باشد. در زبان انگلیسی این ویژگی را با وند یا قید نشان می‌دهند: *shorter* ، *more intelligent* ، *most intelligent*. در زبان فارسی این ویژگیها با وند نشان داده می‌شوند. در زبانهای خاصی نظیر زبان باری^۶ مدرج بودن در فعل نشان داده می‌شود.

زبانها از لحاظ تعداد صفت نیز با یکدیگر تفاوت دارند. در زبان ایگبو^۷ فقط هشت صفت وجود دارد و زبان پنگو^۸ بیست صفت دارد.

۱. هدف اصلی طرح هنگون کردن قواعد گروه‌های دستوری است. از آنجاییکه دو گروه یعنی گروه قیدی و گروه صفتی تا کنون بررسی نشده بودند به ناچار اشاره‌ای نیز به این دو گروه و قواعد آن می‌شود. همه گروه‌های دستوری زبان فارسی بررسی شده باشند.

2. morph syntactically

3. gender

4. inflection

5. graduability

۶. Bari: زبانی که در جنوب سودان بدان تکلم می‌شود.

۷. Igbo: زبانی که در جنوب شرق نیجریه بدان تکلم می‌شود.

8. pengo

به نظر ماهوتیان گروه صنعتی در زبان فارسی را فقط با معیار صرفی و ترتیب کلمات نمی‌توان تعریف کرد. گروه صفتی در فارسی با صفت آغاز می‌شود. یعنی هسته صفت است و در آغاز قرار می‌گیرد و می‌توان در سمت چپ آن وابسته‌ها را قرار داد. گروه صفتی توزیعی^۱ همانند صفت دارد و از نظر معنایی تعیین کننده^۲ اسم است. گروه صفتی در زبان فارسی جزء وابسته‌های پیشین یا پسین گروه اسمی می‌تواند قرار گیرد:

بزرگتر از + گروه اسمی، بزرگترین + گروه اسمی

گروه اسمی + صفت تفضیلی، گروه اسمی + صفت برترین

برخی صفتها نیز خود متمم می‌گیرند: شیفته روی دوست

بنابراین قواعد گروه صفتی را می‌توان چنین ارائه کرد:

الف: صفت + متمم

ب: صفت تفضیلی + گروه اسمی = بزرگتر از خیابان کوچه شما

ج: صفت عالی + گروه اسمی = بزرگترین خیابان شهر

د: صفت + ساده + گروه اسمی = شیفته روی دوست

چنانچه صفت ساده را با A نشان دهیم و نشان تفضیلی و برتر را به ترتیب با A1 و A2 قواعد آن از این قرار خواهد بود:

A1 + NP

A2 + NP

۲-۳- همگون کردن کدهای گروه‌های دستوری

۲-۳-۱- کدهای واژه‌شکن

یکی از کامل‌ترین قواعد و کدهای قراردادی در طرح واژه شکن استفاده شده است. در این طرح هم برای وندهای متعدد زبان فارسی - پیشوندها و پسوندها - و هم برای دیگر اجزاء کلام کدهای قراردادی انتخاب شده است. در بخش اشتقاقی‌ها هم برای پایه‌ای که با وند ترکیب شده و هم برای وندهایی که ویژگیهای معنایی خاص خود را دارند علائم و کدهای قراردادی پیشنهاد شده است.

1. distribution
2. modifier

در بخش کلمات مرکب نیز اسامی و صفات و قیود و ضمائر بر اساس معنی نام‌گذاری و تعیین کد شده است. طبقه‌بندی پیچیده معنایی اجزاء تشکیل دهنده صورتهای اشتقاقی و مرکب باعث شده که کامل‌رین و جامع‌ترین کدها در مجموعه طرح‌های اجرا شده مربوط به طرح واژه شکن باشد.

در طرح واژه شکن ۱۳ نوع صفت و ۲ نوع قید و ۳۷ نوع اسم و ۵ نوع پیشوند و ۳۸ نوع پسوند از لحاظ معنایی از یکدیگر متمایز و برای هر یک کد قراردادی جداگانه اختیار شده است. فهرست نهایی کدهای قراردادی برای همه مقوله‌ها ۱۰۰ عدد است. تمامی این مقوله‌ها از بافتی که در آن قرار می‌گیرند و از روی رفتار صرفی‌شان استخراج شده است. کدهای استخراجی از این قرار است: ۸۶ تا ۱۰۰

۲-۳-۲- کدهای وابسته‌های پیشین و وابسته‌های پسین

n, ni, na, nl	مضاف‌الیه =
np	بدل =
ng	اعداد ترتیبی =
p	ضمیر =

کدها و قواعد مربوط به تلفیق کدها در وابسته‌های پیشین از این قرار است: (۱۱۶ تا ۱۱۹)

کدها و قواعد مربوط به تلفیق کدها در وابسته‌های پسین به قرار ذیل است:

n	اسم =
np	اسم خاص =
nl	اسم جمع =
na	اسم جاندار =
ni	اسم بی جان =
ind	ی نکره =
a	صفت بیانی =
ng	گروه اعداد =
s	جمله ربطی =
r	جمله موصولی =

قواعد صوری و مثال‌ها: (p2, p1)

۲-۳-۳- کدهای فعل شکون

کدها:

s1 = بن ماضی	Vkx	حال التزامی =
s2 = بن مضارع	Vlx	امری =
x (کوچک) = متغیر	Vjx	نهی =
H6, H5, H4, H3, H2, H1 = شناسه‌های مضارع	p1	می =
G6, G5, G4, G3, G2, G1 = شناسه‌های ماضی	p2	پ =
N6, N5, N4, N3, N2, N1 = شناسه‌های ماضی نقلی	p3	ن =
	p4	ه =
	VNyx	فعلی منفی =
	VNkx	حال التزامی منفی =
	A	صفت مفعولی =
	F1	بود =
	F2	باش =
	F3	خواه =
		زمان ها
VAX		حال ساده =
VBX		ماضی نقلی =
VCX		گذشته ساده =
VDX		گذشته استمراری =
VEX		ماضی بعید =
VFX		ماضی التزامی =
VGX		آینده

۲-۳-۴- کدهای عددشکن

عدد شکن به خلاف دیگر طرح‌ها دارای کدهای قراردادی کامل نسبت بلکه قواعد آن عمدتاً به زبان طبیعی عرضه شده و بر اساس زبان طبیعی برای آن برنامه رایانه‌ای نوشته شده است. (ر.ک. ۱-۵-۴)

۲-۳-۵- مقولات واحد و کدهای متفاوت

مجموع طرح‌های واژه شکن و وابسته‌های پیشین و وابسته‌های پسین و فعل شکن و عدد شکن را به دو دسته عمده می‌توان تقسیم کرد. در یک سو طرح واژه شکن قرار می‌گیرد که طرحی در حوزه صرف است و در سوی دیگر چهار طرح وابسته‌های پیشین و وابسته‌های پسین و فعل شکن و عدد شکن که جمله‌گی مربوط به شیوه ترکیب مقولات و عناصر زبانی با یکدیگر در حوزه نحو اند.^۱ مجزا بودن حوزه‌های دستوری در طرح‌های سابق‌الذکر باعث می‌شود که کدهای عرضه شده با وجود همپوشی‌هایی که دارند تا در مواردی متفاوت باشند. بدین معنی که مقوله دستوری واحدی به این سبب که در ساختمان داخلی کلمات به کار می‌رود کد خاصی می‌گیرد و به علت به کار رفتن در ترکیب کلمات در حوزه نحو به صورت دیگری برچسب دهی می‌شود. از این رو است که مثلاً در طرح واژه شکن ۳۴ نوع اسم و ۱۳ نوع صفت از یکدیگر تمیز داده شده‌اند ولی در طرح وابسته‌های پیشین صرفاً چهار نوع اسم (اسم خاص، اسم جمع، اسم جاندار، اسم بی‌جان) تشخیص داده شده است یا در طرح وابسته‌های پیشین ۱۲ نوع صفت فهرست شده است.

از طرفی برای طرح‌های حوزه نحو بعضاً کدهایی با ویژگیهای نحوی در اختیار شده که طبیعتاً مربوط به ویژگیهای ترکیبی عناصر در زنجیره کلام است. مثلاً در طرح وابسته‌های پسین به چهار نوع مضاف الیه و بدل و جمله ربطی و جمله موصولی قائل گردیده‌اند و نیز در طرح وابسته‌های پیشین به عناصری مثل صفت عالی و ممیز اشاره شده است.

از طرح‌هایی که در حوزه نحو اجرا شده‌اند طرح فعل شکن است که مقوله‌ای کاملاً مجزا است و شباهتی به دیگر طرح‌ها ندارد و کدهای انتخاب شده برای آن هیچگونه همپوشی با کد دیگر طرح‌ها ندارد.

کدهای عدد شکن با اینکه تا حدی با طرح واژه شکن همپوشی دارد ولی چون مربوط به شیوه ترکیب اعداد با یکدیگر است در آن از کدهایی متفاوت استفاده شده است. قواعد و کدهای عدد شکن از طرفی در طرح وابسته‌های پیشین و وابسته‌های پسین کاربرد دارد زیرا بخشی از وابسته‌های پیشین یا پسین هسته گروه اسمی است.

مقوله اسم در طرح واژه شکن و طرح وابسته پسین فقط در یک مورد مشترک است و آن اسم جاندار است. این مقوله در وابسته پسین با علامت na و در طرح واژه شکن با علامت قراردادی

no2 نشان داده شده است.^۱ مقوله اسم در طرح وابسته پسین از آنجاییکه با توجه به کارکرد نحوی و دریافت است.

در نقش مضاف‌الیه و به صورت زیر نشان داده شده است.

n	اسم
nl	اسم جمع
na	اسم جاندار
ni	اسم بی جان

در وابسته پسین فقط دو نوع صفت به کار رفته (صفت و صفت عالی) که به ترتیب با علامت A13 و نمایش داده شده است. کدهایی متفاوت از کدهای صفت در واژه شکن. در طرح وابسته‌های پیشین ۱۲ صفت وجود دارد که نامهایی متفاوت با صفت‌های واژه شکن نیز (۱۳ صفت) دارند و کدهای متفاوتی نیز دارند: کدهایی که با علامت A و متغیرهای عددی نشان داده شده است. نظیر A10a و A12b و غیره.

مقولات واحدی نظیر اسم و صفت در طرح‌های واژه شکن و وابسته پیشین و وابسته پسین با وجود نامهای متفاوتی که بسته به بافت برای آنها اختیار شده اشتراکاتی نیز دارند. بدین معنی که مثلاً برای اسم و صفت به ترتیب کد n و A انتخاب شده یا برای ضمیر و قید به ترتیب علامت p و Ad و برای گروه صفتی علامت AP و برای گروه قیدی علامت Adp.

بدین معنی که کد پایه در تمامی موارد یکسان است و تفاوت در کد متغیرها یا وابسته‌ها است. مثلاً اعدادی که متغیر مقوله صفت در طرح وابسته پیشین‌اند (نظیر A12a) و وابسته‌هایی که در طرح واژه شکن به همراه صفت آمده‌اند (نظیر A51 برای صفت مرکب اشتقاقی)^۲.

۲-۳-۶- کدهای درهم کرد.

در این قسمت کد جامع طرح‌های اجرا شده (واژه شکن، وابسته‌های پیشین، وابسته‌های پسین، فعل شکن) عرضه می‌شود:

n	اسم
np	اسم خاص

۱. در طرح واژه شکن اسم جاندار مرکب و اشتقاقی به ترتیب با Nog و N14 نشان داده شده است.
 ۲. علامت یکی از انواع اسم که مضاف‌الیه قرار می‌گیرد و علامت اسم جمع همپوشی دارد که در فهرست کدهای جامع اصلاح می‌شود.

nl	اسم جمع
na	اسم جاندار
ni	اسم بیجان
ind	ی نکره
A	صفت (بیانی)
ng	گروه اعداد
¹ n, nx, , na, ni	مضاف‌الیه

p	ضمیر
s	جمله ربطی
r	جمله موصولی
A10	صفت اشاره
A10a	صفت اشاره بسیط
A10b	صفت اشاره مرکب
A11	صفت مبهم
A11a	صفت مبهم بسیط
A11b	صفت مبهم مرکب
A12	صفت پرسشی
A12a	صفت پرسشی بسیط
A12b	صفت پرسشی مرکب
A13	صفت عالی ^۲
A1	صفت تفضیلی
M	اعداد
Mo1	اعداد اصلی
Aob (mol s 25)	اعداد ترتیبی
Q	ممیز ^۳
Vax	زمان حال ساده
Vbx	ماضی نقلی
VGx	گذشته ساده
VDx	گذشته استمراری
Vex	ماضی بعید
Vfx	ماضی التزامی

V_{Gx}

آینده

۱. به سبب مشابه بودن کد این نوع مضاف‌الیه و اسم جمع کد مضاف‌الیه به nx تغییر می‌یابد.
 ۲. صفت عالی در بخش گروه صفتی (۲-۲) با کد A2 نشان داده شده است.
 این صفت در طرح وابسته‌های پیشین با کد A01S39 نیز نشان داده شده است.
 ۳. صفت‌های «چند/چندمین» و «کدام» و «همه/چند» به ترتیب با کدهای A12a1 و A12a2 و A11a1 نشان داده شده‌اند.

V_{kx}	حال التزامی
V_{Lx}	امر
V_{Jx}	نهری
P1	می
P2	ب
P3	ن
P4	ه
Vn_{yx}	فعل منفی
VN_{kx}	حال التزامی منفی
A0	صفت مفعولی
F1	بود ^۱
F2	باش
F3	خواه
S1	بن ماضی
S2	بن مضارع

۱. صورتهای استثناء بی‌قاعده

H1	م ^۱
H2	ی
H3	د
H4	یم
H5	ید
H6	ند
N1	ام ^۲
N2	ای
N3	است
N4	ایم
N5	اید
N6	اند
G1	م ^۳
G2	ی
G3	φ
G4	یم
G5	ید
G6	ند

Code	Description
A01	صفت
A02	صفت مفعولی
A03	صفت مفعولی منفی
A04	صفت مبالغه
A05	صفت مبالغه منفی
A06	صفت مرکب
A07	صفت بن ساختی
A08	صفت ترکیبی
A09	صفت اشتقاقی
A51	صفت مرکب اشتقاقی
A52	صفت بن ساختی اشتقاقی
A53	صفت مرکب بن ساختی

۱. شناسه‌های مضارع
 ۲. شناسه ماضی نقلی
 ۳. شناسه‌های ماضی

A54	صفت قید اشتقاقی
B01	قید
B02	قید جهت
C01	ضمیر
I01	بن مضارع
I02	بن ماضی
M01	عدد
N01	اسم
N02	اسم جاندار
N03	اسم معنی
N04	اسم عمل

Code	Description
N05	اسم زمان
N06	اسم خویش
N07	اسم مصدر
N08	اسم مرکب
N09	اسم جاندار مرکب
N10	اسم معنی اشتقاقی
N11	اسم بن ساختی
N12	اسم جاندار بن ساختی
N13	اسم خویش مرکب
N14	اسم جاندار اشتقاقی
N15	اسم عمل اشتقاقی
N16	اسم اشتقاقی
N17	اسم خویش اشتقاقی
N18	اسم ظرف اشتقاقی
N19	اسم مکان اشتقاقی
N20	اسم مجموعه
N21	اسم زمان اشتقاقی
N51	اسم مرکب اشتقاقی
N52	اسم بن ساختی اشتقاقی
N53	اسم جاندار مرکب اشتقاقی
N54	اسم جاندار بن ساختی اشتقاقی
N55	اسم عمل مرکب اشتقاقی
N56	اسم عمل بن ساختی اشتقاقی

Code	Description
N59	اسم مکان مرکب اشتقاقی

N60	اسم مکان بن ساختی
N61	اسم مجموعه مرکب
N62	اسم مجموعه بن ساختی
N63	اسم خویش مرکب اشتقاقی
N64	اسم مرکب بن ساختی
N65	اسم جاندار مرکب بن ساختی
N66	اسم خویش مرکب بن ساختی
N67	اسم معنی مرکب اشتقاقی
N68	اسم معنی بن ساختی اشتقاقی
P01	پیشوند مبالغه
P02	پیشوند دارندگی
P03	پیشوند منفی ساز
P04	پیشوند منفی ساز ۱
P05	پیشوند منفی ساز ۲
P06	پیشوند اشتراک
S01	پسوند شباهت
S02	پسوند آغشتگی
S03	پسوند اسم معنی اشتقاقی ساز
S04	پسوند نسبت
S05	پسوند مبالغه
S06	پسوند تقلید
S07	پسوند محافظت

Code	Description
S08	پسوند عمل
S09	پسوند تصغیر
S10	پسوند ظرف
S11	پسوند مکان
S12	پسوند شباهت ۱
S13	پسوند مکان ۱
S14	پسوند شباهت ۲
S15	پسوند تصغیر ۱
S16	پسوند مکان ۲
S17	پسوند مکان ۳
S18	پسوند مجموعه
S19	پسوند کنندگی
S20	پسوند عمل ۱
S21	پسوند شباهت ۳
S22	پسوند دارندگی
S23	پسوند ترتیب
S24	پسوند دارندگی ۱
S25	پسوند ترتیب ۱
S26	پسوند دارندگی ۲
S27	پسوند شباهت ۴
S28	پسوند دارندگی ۳
S29	پسوند شباهت ۵
S30	پسوند شباهت ۶

Code	Description
S31	پسوند نسبت ۱
S32	پسوند جنس
S33	پسوند فاعلی
S34	پسوند عمل ۲
S35	پسوند مصدر ساز
S36	پسوند فاعلی ۱
S37	پسوند مفعولی
S38	پسوند کنندگی ۲

فصل سوم

خلاصه و نتیجه‌گیری

در فصل اول نگاهی اجمالی به طرح‌هایی شده که تاکنون در پژوهشگاه علوم و فناوری اطلاعات ایران (ایرانداک) درباره‌ی پردازش زبان فارسی اجرا شده است. این طرح‌ها شامل دستورسنج و واژه‌شکن و تعیین قواعد صوری وابسته‌های پیشین زبان فارسی و تعیین قواعد صوری وابسته‌های پسین زبان فارسی و فعل‌شکن و عدد‌شکن بوده‌اند. این طرح‌ها به دو دسته عمده‌اند. طرح واژه‌شکن در حوزه حرف و دیگر طرح‌ها در حوزه نحو انجام شده‌اند.

در فصل دوم ابتدا کدهایی برای دو گروه دستوری زبان فارسی یعنی گروه قیدی و گروه صفتی - که از گروه‌های فرعی زبان فارسی هستند - عرضه شد. سپس مقایسه کدها انجام شد. چند کد با یکدیگر همپوشی داشتند که یک کد واحد برای آنها انتخاب شد. دیگر کدها عمدتاً مکمل یکدیگر بودند. مثلاً کدهای اسامی در طرح وابسته‌های پسین (کد اسم جاندار و اسم خاص و اسم جمع و اسم بی‌جان) مکمل کدهای واژه‌شکن بودند (مکمل ۳۴ کد موجود در واژه‌شکن) یا تعداد ۱۲ کد صفت‌های وابسته‌های پیشین مکمل کدهای صفت‌های استخراجی واژه‌شکن بودند (مکمل ۱۳ صفت موجود در واژه‌شکن).

لازم به ذکر است کدهای مقولات اصلی نظیر کد مقوله اسم و مقوله صفت کد واحدی بودند و پایه آنها کد مشابهی داشت. مثلاً کد اسم با علامت Π و صفت با علامت A بود اما متغیرهای آنها بسته به نوع اسم یا نوع صفت متفاوت بود.

در نهایت فهرست درهم کرد و جامع کد تمامی طرح‌ها عرضه شد.

کتابنامه

- _____، ۱۳۸۱، پردازش گروه اسمی، فصل‌نامه اطلاع‌رسانی، دوره ۱۸، ص ۳۴-۴۱.
- _____، ۱۳۸۵، واژه‌شکن زبان فارسی، پژوهشگاه علوم و فناوری اطلاعات ایران.
- _____، ۱۳۸۹، تعیین قاعده‌های صوری وابسته‌های پیشین گروه اسمی و کاربرد آن در ترجمه ماشینی، پژوهشگاه علوم و فناوری اطلاعات ایران.
- _____، ۱۳۹۲، شناسایی طرح‌های مرتبط با تحقیق ترجمه ماشینی در زبان فارسی، پژوهشگاه علوم و فناوری اطلاعات ایران.
- _____، ۱۳۹۳، تعیین قاعده وابسته‌های پسین گروه اسمی (تمهیداتی زبانی برای پردازش زبان)، طرح پژوهشی، پژوهشگاه علوم و فناوری اطلاعات ایران.
- _____، ۱۳۸۶، عدد شکن زبان فارسی، پژوهشگاه علوم و فناوری اطلاعات ایران.
- _____، ۱۳۸۸، فعل شکن زبان فارسی، پژوهشگاه علوم و فناوری اطلاعات ایران.
- سمائی، سید مهدی، ۱۳۷۷، واژگان در دستورسنج، رساله دکتری زبان‌شناسی، دانشگاه تهران.
- کانلن، اسی، پی، ان و دیگران، ۱۳۵۷، تهیه بانک پر حجم واژگان برای کاربرد در نظام‌های متن‌سازی و سازه‌یابی و بازیابی اطلاعات، ترجمه سید مهدی سمائی، نشریه اطلاع‌رسانی و بازیابی اطلاعات، ترجمه سید مهدی سمائی، نشریه اطلاع‌رسانی، دوره ۱۲، ص ۴۸-۷۳.
- معین، محمد، ۱۳۷۱، فرهنگ فارسی معین، انتشارات امیر کبیر.

Anderson, S.R.2004, towards a less syntactic morphology and a more morphological syntax, in P.V.strekenberg (ed) Linguistics today:

Boot, K.H.V.1967, machine aided translation with a post – editor, in A.D. Booth (ed), machine translation, north Holland publishing company pp.53-67.

Crystal, D. 1992, an encyclopedic dictionary of language and languages, Blackwell.

Facing a greater challenge, John Benjamin's publishing company, pp.31-45

Haser,V and B.kortman, 2006, adverbs, in Keith Brown, encyclopedia of language and linguistics, Elsevier, pp.66-69.

Katz, J & Fosor,J. 1963, the structure of a semantic theory, language,39,pp.17-210.

Katz,j & postal, p.m., 1964, an integrated theory of linguistic descriptions, MIT press.

Mahootian, sh. 1997, Persian, routledge.

Pustet, R. 2006, adjective, in Keith brown, encyclopedia of language and linguistics, Elsevier, pp.60-63.

Ramat, P and D. Ricca, 1994, prototypical adverbs: on the secularity / radiality of the notion of adverb, Rivista di linguistica, 6(2), pp289 - 326.

Trask, R.L. 1993, a dictionary of grammatical terms in linguistics, rout ledge.

پیوست‌ها:

نمونه‌هایی از فهرست قواعد طرح‌ها

formula	small formula
N14I01	N54
N14I02	N54
N14N02	N53
N14N03	N53
N14N04	N53
N14s04	A54
N15I01	N54
N15N01	N51
N16A01	N53
N16I01	N52
N16I01	N52
N16I02	N53
N16N01	N51
N16N02	N53
N16N03	N51
N16N04	N51
N16N05	N51
P01I01	A04
P02n03	N14
P03n03	N14
P04A02	A03
P04L01	A05
P05A02	A03
P05N02	N14
P06N01	N14

Formula	small formula
N09S04	A51
N09S05	N53
N09S06	N55
N09S12	A51
N09S13	N59
N09S16	N59

N09S17	N59
N09S20	N55
N09S27	A51
N09S29	A51
N09S31	A51
N10I10	N67
N11S16	N54
N11S17	N60
N11S27	A52
N11S28	N54
N11S29	A52
N11S31	A52
N12S01	A51
N12S04	A52
N12S05	N54
N12S06	N56
N12S12	A52
N12S13	N60
N12S16	N60
N12S17	N60
N12S21	A52
N12S27	A52
N12S29	A52
N12S31	N68
N13I01	N66

Formula	small formula
N06N06	N13
N06S04	A09
N06S05	N14
N06S09	N17
N06S15	N17
N06S27	A09
N06S31	N10
N07I01	N12
N07S17	N19
N07S19	N14
N08I01	N64
N08I02	N64
N08S01	A51
N08S02	A51
N08I05	N53
N08S07	N53

N08S08	N55
N08S11	N59
N08S12	A51
N08S13	N59
N08S16	N59
N08S17	N59
N08S19	N53
N08S21	A51
N08S26	A01
N08S27	A51
N08S28	N53
N08S29	A51
N08S31	A01
N08S38	N53
N09I01	N65
N09I02	N65
N09S01	A51

Formula	small formula
N03I02	N12
N03N01	N08
N03S02	A09
N03S07	N14
N03S10	N18
N03S13	N19
N03S16	N19
N03S17	N19
N03S18	N20
N03S19	N14
N03S22	N14
N03S24	N14
N03S26	A09
N03S28	N14
N03S31	A09
N04I01	N12
N04N01	N08
N04S05	N14
N04S12	A09
N04S13	N19
N04S17	N19
N04S19	N14
N04S20	N15
N05I01	N12

N05N01	N08
N05N02	N09
N05S04	A09
N05S07	N14
N05S14	A09
N05S17	N21
N05S21	A09
N05S29	A09
N06I01	N12

Formula	small formula
N01S27	A09
N01S28	N14
N01S29	A09
N01S30	N16
N01S31	A09
N01S32	A09
N01S38	N14
N02I01	N12
N02N01	N09
N02N02	N09
N02N03	N09
N02N04	N09
N02N09	N14
N02S01	A09
N02S04	A09
N02S05	N14
N02S06	N15
N02S07	N14
N02S09	N14
N02S10	N18
N02S12	A09
N02S13	N19
N02S14	A09
N02S15	N14
N02S16	N19
N02S17	N19
N02S18	N20
N02S20	N15
N02S21	A09
N02S27	A09
N02S29	A09
N02S31	N10

N03I01	N12
Formula	small formula
C01A02	N09
C01I01	N12
101S35	N07
101S36	N14
101S37	A02
M01S23	A08
M01S25	A08
N01A01	N09
N01I01	N12
N01I02	N11
N01N01	N08
N01N02	N09
N01N03	N08
N01N04	N08
N01N05	N08
N01S01	A09
N01S02	A09
N01S05	N14
N01S07	N14
N01S08	N15
N01S09	N16
N01S10	N18
N01S11	N19
N01S12	A09
N01S13	N19
N01S14	A09
N01S15	N16
N01S16	N19
N01S17	N19
N01S18	N20
N01S19	N14
N01S21	A09
N01S26	A09
Formula	small formula
A01I01	N12
A01I02	A07
A01N01	N09
A01N02	N09
A01N03	N09

A01N04	N09
A01N05	N09
A01N07	N09
A01S03	N10
A01S14	A09
A01S17	N19
A01S26	A09
A01S31	N10
A01S38	N14
A06I01	N65
A06I02	A53
A06S17	N59
A06S17	N60
A06S31	N67
A07S31	N68
A09I01	N54
A09I02	A52
B01I01	N12
B01N01	N08
B01S17	N19
B02I01	N11
B02I02	N11
B02N01	N08
B02N01	N08
B02N04	N08
B02S17	N19

صفت پرسشی مرکب + صفت اشاره بسیط + همه / چند A12b + A10a + A11a1	صفت پرسشی مرکب + صفت اشاره بسیط A12b + A10a
صفت پرسشی مرکب + صفت اشاره بسیط + اعداد اصلی A10a + M01 + A12b	صفت اشاره بسیط + اعداد اصلی A10a + m01
صفت پرسشی مرکب + صفت اشاره بسیط + اعداد اصلی + ممیز A12b + A10a + M01+Q	صفت اشاره بسیط + همه / چند A10a + A11a1
صفت پرسشی مرکب + صفت اشاره بسیط + صفت عالی A12b + A10a + A13	صفت اشاره بسیط + صفت عالی A10a + A13
صفت پرسشی مرکب + اعداد اصلی + ممیز A12b + M01+ Q	صفت پرسشی مرکب + اعداد اصلی A12b + M01
	صفت پرسشی مرکب + صفت عالی A12b + A13
	صفت مبهم مرکب + صفت اشاره بسیط A11b + A10a
کدام + اعداد + ممیز A12a 2+ M+Q (کدام + اعداد اصلی + ممیز) (A12a 2 + M01+ Q) (کدام + اعداد ترتیبی + ممیز) (A12a 2 + M01 + A08)	M+Q اعداد + ممیز (M+Q) (اعداد اصلی + ممیز) (A08 + Q) (اعداد ترتیبی + ممیز)
	A12a 2 + M کدام + اعداد (A12a 2+ M01)(اعداد اصلی کدام + اعداد) (کدام + اعداد ترتیبی)

در جدول های زیر قواعد و صورت کد نویسی شده آن ها نشان داده شده است.

S1+P4 = A	بن ماضی + ه = صفت مفعولی	1
P1+S2+HX=VAX	می + بن مضارع + شناسه مضارع = حال ساده	2
P2+S2+HX=VKX	ب + بن مضارع + شناسه مضارع = حال التزامی	3
A+NX = VBX	استثناء: ب + ی + بن مضارعی که با آ آغاز می شود + شناسه مضارع = فعل حال التزامی	1-3
	صفت مفعولی + شناسه نقلی = ماضی نقلی	4

S1 + GX = VCX	بن ماضی + شناسه ماضی = گذشته ساده	5
P1+S1+GX=VDX	می + بن ماضی + شناسه ماضی = گذشته استمراری	6
A+F1+GX=VEX	صفت مفعولی + بود + شناسه ماضی = ماضی بعید	7
A+ F2+ HX = VFX	صفت مفعولی + باش + شناسه مضارع + ماضی التزامی	8
P2+S2=VL2	ب + بن مضارع = امر دوم شخص مفرد	9
P3+S2 = VJ2	ن + بن مضارع = نهی دوم شخص مفرد	10
F3+HX=S1=VGX	خواه + شناسه مضارع + بن ماضی = آینده	11
P3+VYX=VNYX	ن + فعل = فعل منفی	12
P3+VaX=VNAX	ن+ حال ساده = حال ساده منفی	1-12
P3+ A + NX = VNBX	ن + صفت مفعولی + شناسه نقلی = ماضی نقلی منفی	2-12
P3+S1+GX=VNCX	ن+ بن ماضی + شناسه ماضی = گذشته ساده منفی	3-12
P3+P1+S1=VNDX	ن + می + بن ماضی + شناسه ماضی = گذشته استمراری منفی	4-12
P3+A+F1+GX=VNEX	ن+ صفت مفعولی + بود + شناسه مضارع = ماضی بعید منفی	5-12
P3+A+F2+HX=VNFx	ن+ صفت مفعولی + باش + شناسه مضارع = ماضی التزامی منفی	6-12
P3+F3+HX+S1=VNGX	ن + خواه + شناسه مضارع + بن ماضی = آینده منفی	7-12
		13
P3+S2+HX=VNKX	ن + بن مضارع + شناسه مضارع = حال التزامی منفی	

آینده	خواه + م + خورد = خواهم خورد	
	اگر HX = H2 (یعنی شناسه مضارع دوم شخص مفرد) آنگاه VGX = VG2 (یعنی آینده دوم شخص مفرد): خواه + ی + خورد = خواهی خورد	
	اگر HX = H3 (یعنی شناسه مضارع سوم شخص مفرد) آنگاه VGX = VG3 (یعنی آینده سوم شخص مفرد): خواه + د + خورد = خواهد خورد	
	اگر HX = H4 (یعنی شناسه مضارع اول شخص جمع) آنگاه VGX = VG4 (یعنی آینده اول شخص جمع): خواه + یم + خورد = خواهیم خورد	
	اگر HX = H5 (یعنی شناسه مضارع دوم شخص جمع) آنگاه VGX = VG5 (یعنی آینده دوم شخص جمع): خواه + ید + خورد = خواهید خورد	
	اگر HX = H6 (یعنی شناسه مضارع سوم شخص جمع) آنگاه VGX = VG6 (یعنی آینده سوم شخص جمع): خواه + ند + خورد = خواهند خورد	

P3+VYX = VNYX	ن + فعل = فعل منفی	12
---------------	--------------------	----

$P3+VAX=VNAX$ ن + حال ساده = حال ساده منفی	اگر $VAX = VA1$ (یعنی حال ساده اول شخص مفرد) آنگاه $VNA1 = VNAX$ (یعنی حال ساده منفی اول شخص مفرد): ن + می خورم = نمی خورم	
	اگر $VAX = VA2$ (یعنی حال ساده دوم شخص مفرد) آنگاه $VNA2 = VNAX$ (یعنی حال ساده منفی دوم شخص مفرد): ن + می خوری = نمی خوری	
	اگر $VAX = VA3$ (یعنی حال ساده سوم شخص مفرد) آنگاه $VNA3 = VNAX$ (یعنی حال ساده منفی سوم شخص مفرد): ن + می خورد = نمی خورد	
	اگر $VAX = VA4$ (یعنی حال ساده اول شخص جمع) آنگاه $VNA4 = VNAX$ (یعنی حال ساده منفی اول شخص جمع): ن + می خوریم = نمی خوریم	
	اگر $VAX = VA5$ (یعنی حال ساده دوم شخص جمع) آنگاه $VNA5 = VNAX$ (یعنی حال ساده منفی دوم شخص جمع): ن + می خورید = نمی خورید	
	اگر $VAX = VA6$ (یعنی حال ساده سوم شخص جمع) آنگاه $VNA6 = VNAX$ (یعنی حال ساده منفی سوم شخص جمع): ن + می خورند = نمی خورند	

$P3+A+NX=VNBX$ ن + صفت مفعولی + شناسه نقلی = ماضی نقلی منفی	اگر $NX = N1$ (یعنی شناسه ماضی نقلی اول شخص مفرد) آنگاه $VNB1 = VNBX$ (یعنی فعل ماضی نقلی منفی اول شخص مفرد): ن + خورد+ام = نخوردم	
	اگر $NX = N2$ (یعنی شناسه ماضی نقلی دوم شخص مفرد) آنگاه $VNB2 = VNBX$ (یعنی فعل ماضی نقلی منفی دوم شخص مفرد): ن + خورد+ای = نخوردای	
	اگر $NX = N3$ (یعنی شناسه ماضی نقلی سوم شخص مفرد) آنگاه $VNB3 = VNBX$ (یعنی فعل ماضی نقلی منفی سوم شخص مفرد): ن + خورد+است = نخورداست	
	اگر $NX = N4$ (یعنی شناسه ماضی نقلی اول شخص جمع) آنگاه $VNB4 = VNBX$ (یعنی فعل ماضی نقلی منفی اول شخص جمع): ن + خورد+ایم = نخوردایم	
	اگر $NX = N5$ (یعنی شناسه ماضی نقلی دوم شخص جمع) آنگاه $VNB5 = VNBX$ (یعنی فعل ماضی نقلی منفی دوم شخص جمع): ن + خورد+اید = نخورداید	
	اگر $NX = N6$ (یعنی شناسه ماضی نقلی سوم شخص جمع) آنگاه $VNB6 = VNBX$ (یعنی فعل ماضی نقلی منفی سوم شخص جمع): ن + خورد+اند = نخورداند	
$P3+S1+GX=VNCX$ ن + بن ماضی + شناسه ماضی = گذشته ساده منفی	اگر $GX = G1$ (یعنی شناسه ماضی اول شخص مفرد) آنگاه $VNC1 = VNCX$ (یعنی فعل گذشته ساده منفی اول شخص مفرد): ن + خورد+م = نخوردم	
	اگر $GX = G2$ (یعنی شناسه ماضی دوم شخص مفرد) آنگاه $VNC2 = VNCX$ (یعنی فعل گذشته ساده منفی دوم شخص مفرد): ن + خورد+ی = نخوردی	
	اگر $GX = G3$ (یعنی شناسه ماضی سوم شخص مفرد) آنگاه $VNC3 = VNCX$ (یعنی فعل گذشته ساده منفی سوم شخص مفرد): ن + خورد+Φ = نخورد	
	اگر $GX = G4$ (یعنی شناسه ماضی اول شخص جمع) آنگاه $VNC4 = VNCX$ (یعنی فعل گذشته ساده منفی اول شخص جمع): ن + خورد+ایم = نخوردیم	
	اگر $GX = G5$ (یعنی شناسه ماضی دوم شخص جمع) آنگاه $VNC5 = VNCX$ (یعنی فعل گذشته ساده منفی دوم شخص جمع): ن + خورد+اید = نخوردید	
	اگر $GX = G6$ (یعنی شناسه ماضی سوم شخص جمع) آنگاه $VNC6 = VNCX$ (یعنی فعل گذشته ساده منفی سوم شخص جمع): ن + خورد+اند = نخوردند	

	ساده منفی سوم شخص جمع: ن + خورد+ند= نخوردند
$P3+P1+S1+GX=VNDX$ بن ماضی + شناسه ماضی = گذشته استمراری منفی	اگر $GX = G1$ (یعنی شناسه ماضی اول شخص مفرد) آنگاه $VNDX = VND1$ (یعنی فعل گذشته استمراری منفی اول شخص مفرد: ن + می + خورد+م= نمی خوردم)
	اگر $GX = G2$ (یعنی شناسه ماضی دوم شخص مفرد) آنگاه $VNDX = VND2$ (یعنی فعل گذشته استمراری منفی دوم شخص مفرد: ن + می + خورد+ی= نمی خوردی)
	اگر $GX = G3$ (یعنی شناسه ماضی سوم شخص مفرد) آنگاه $VNDX = VND3$ (یعنی فعل گذشته استمراری منفی سوم شخص مفرد: ن + می + خورد+Φ= نمی خورد)
	اگر $GX = G4$ (یعنی شناسه ماضی اول شخص جمع) آنگاه $VNDX = VND4$ (یعنی فعل گذشته استمراری منفی اول شخص جمع: ن + می + خورد+یم= نمی خورديم)
	اگر $GX = G5$ (یعنی شناسه ماضی دوم شخص جمع) آنگاه $VNDX = VND5$ (یعنی فعل گذشته استمراری منفی دوم شخص جمع: ن + می + خورد+ید= نمی خورديد)
	اگر $GX = G6$ (یعنی شناسه ماضی سوم شخص جمع) آنگاه $VNDX = VND6$ (یعنی فعل گذشته استمراری منفی سوم شخص جمع: ن + می + خورد+ند= نمی خوردند)
$P3+A+F1+GX=VNEX$ بن +صفت مفعولی+ بود + شناسه ماضی = ماضی بعید منفی	اگر $GX = G1$ (یعنی شناسه ماضی اول شخص مفرد) آنگاه $VNEX = VNE1$ (یعنی فعل ماضی بعید منفی اول شخص مفرد: ن + خورد+م=م نخورده بودم)
	اگر $GX = G2$ (یعنی شناسه ماضی دوم شخص مفرد) آنگاه $VNEX = VNE2$ (یعنی فعل ماضی بعید منفی دوم شخص مفرد: ن + خورد+ی=ی نخورده بودی)
	اگر $GX = G3$ (یعنی شناسه ماضی سوم شخص مفرد) آنگاه $VNEX = VNE3$ (یعنی فعل ماضی بعید منفی سوم شخص مفرد: ن + خورد+د=د نخورده بودند)
	اگر $GX = G4$ (یعنی شناسه ماضی اول شخص جمع) آنگاه $VNEX = VNE4$ (یعنی فعل ماضی بعید منفی اول شخص جمع: ن + خورد+یم=یم نخورده بوديم)
	اگر $GX = G5$ (یعنی شناسه ماضی دوم شخص جمع) آنگاه $VNEX = VNE5$ (یعنی فعل ماضی بعید منفی دوم شخص جمع: ن + خورد+ید=ید نخورده بوديد)
	اگر $GX = G6$ (یعنی شناسه ماضی سوم شخص جمع) آنگاه $VNEX = VNE6$ (یعنی فعل ماضی بعید منفی سوم شخص جمع: ن + خورد+ند=ند نخورده بودند)

$P_3 + S_2 = H_X = VN_{KX}$ ن + بن مضارع + شناسه مضارع = حال التزامی منفی	اگر $H_1 = H_X$ (یعنی شناسه مضارع اول شخص مفرد) آنگاه $VN_{KX} = VN_{K1}$ (یعنی فعل حال التزامی منفی اول شخص مفرد) ن+خور+م= نخورم
	اگر $H_2 = H_X$ (یعنی شناسه مضارع دوم شخص مفرد) آنگاه $VN_{KX} = VN_{K2}$ (یعنی فعل حال التزامی منفی دوم شخص مفرد) ن+خور+ی= نخوری
	اگر $H_3 = H_X$ (یعنی شناسه مضارع سوم شخص مفرد) آنگاه $V_{K3} = V_{KX}$ (یعنی فعل حال التزامی منفی سوم شخص مفرد) ن+خور+د= نخورد
	اگر $H_4 = H_X$ (یعنی شناسه مضارع اول شخص جمع) آنگاه $V_{G4} = V_{GX}$ (یعنی فعل حال التزامی منفی اول شخص جمع) ن+خور+یم= نخوريم
	اگر $H_5 = H_X$ (یعنی شناسه مضارع دوم شخص جمع) آنگاه $V_{G4} = V_{GX}$ (یعنی فعل حال التزامی منفی دوم شخص جمع) ن+خور+ید= نخوريد

	اگر $H_6 = H_X$ (یعنی شناسه مضارع سوم شخص جمع) آنگاه $V_{GX} = V_{G4}$ (یعنی فعل حال التزامی منفی سوم شخص جمع) ن+خور+ند= نخورند
--	--

$P3+A+F2+HX = V_{NFX}$ ن + صفت مفعولی + باش + شناسه مضارع = ماضی التزامی منفی	اگر $H_1 = H_X$ (یعنی شناسه مضارع اول شخص مفرد) آنگاه $V_{NFX} = V_{NF}$ (یعنی فعل ماضی التزامی منفی اول شخص مفرد): ن + خورده + باش + م = نخورده باشم	۶-۱۲
	اگر $H_2 = H_X$ (یعنی شناسه مضارع دوم شخص مفرد) آنگاه $V_{NFX} = V_{NF}$ (یعنی فعل ماضی التزامی منفی دوم شخص مفرد): ن + خورده + باش + ی = نخورده باشی	
	اگر $H_3 = H_X$ (یعنی شناسه مضارع سوم شخص مفرد) آنگاه $V_{NFX} = V_{NF}$ (یعنی فعل ماضی التزامی منفی سوم شخص مفرد): ن + خورده + باش + د = نخورده باشد	
	اگر $H_4 = H_X$ (یعنی شناسه مضارع اول شخص جمع) آنگاه $V_{NFX} = V_{NF}$ (یعنی فعل ماضی التزامی منفی اول شخص جمع): خورده + باش + یم = نخورده باشیم	
	اگر $H_5 = H_X$ (یعنی شناسه مضارع دوم شخص جمع) آنگاه $V_{NFX} = V_{NF}$ (یعنی فعل ماضی التزامی منفی دوم شخص جمع): خورده + باش + ید = نخورده باشید	
	اگر $H_6 = H_X$ (یعنی شناسه مضارع سوم شخص جمع) آنگاه $V_{NFX} = V_{NF}$ (یعنی فعل ماضی التزامی منفی سوم شخص جمع): خورده + باش + ند = نخورده باشند	
$P3+F3+HX+S1 = V_{NGX}$ ن + خواه + شناسه مضارع + بن ماضی = آینده منفی	اگر $H_1 = H_X$ (یعنی شناسه مضارع اول شخص مفرد) آنگاه $V_{NGX} = V_{NG}$ (یعنی فعل آینده منفی اول شخص مفرد): ن + خواه + م + خورد = نخواهم خورد	۵-۱۲
	اگر $H_2 = H_X$ (یعنی شناسه مضارع دوم شخص مفرد) آنگاه $V_{NGX} = V_{NG}$ (یعنی فعل آینده منفی دوم شخص مفرد): ن + خواه + ی + خورد = نخواهی خورد	
	اگر $H_3 = H_X$ (یعنی شناسه مضارع سوم شخص مفرد) آنگاه $V_{GX} = V_G$ (یعنی فعل آینده منفی سوم شخص مفرد): ن + خواه + د + خورد = نخواهد خورد	
	اگر $H_4 = H_X$ (یعنی شناسه مضارع اول شخص جمع) آنگاه $V_{GX} = V_G$ (یعنی فعل آینده منفی اول شخص جمع): ن + خواه + یم + خورد = نخواهیم خورد	
	اگر $H_5 = H_X$ (یعنی شناسه مضارع دوم شخص جمع) آنگاه $V_{GX} = V_G$ (یعنی فعل آینده منفی دوم شخص جمع): ن + خواه + ید + خورد = نخواهید خورد	
	اگر $H_6 = H_X$ (یعنی شناسه مضارع سوم شخص جمع) آنگاه $V_{GX} = V_G$ (یعنی فعل آینده منفی سوم شخص جمع): ن + خواه + ند + خورد = نخواهند خورد	

Abstract

Key words:

•



**Iranian Research Institute for Information Science and
Technology(irandoc)**

**Assimilated formal rules of Farsi grammatical phrases
(for language processing)**

Seyed Mehdi Samaei

2016