


	عنوان پروژه: فاز اول طرح جامع پیکره زبان فارسی با موضوع فاز اول مطالعاتی ایجاد پیکره متنی زبان فارسی		 شورای عالی اطلاع رسانی
	عنوان زیرپروژه: بررسی راهکارهای ایجاد ماشین تشخیص هوشمند افعال مرکب در زبان فارسی		
	تاریخ: 1388/03/31	ویرایش: 1/0	



عنوان زیرپروژه:

بررسی راهکارهای ایجاد ماشین تشخیص هوشمند افعال مرکب در زبان فارسی



	عنوان پروژه:		 شورای عالی اطلاع رسانی
	فاز اول طرح جامع پیکره زبان فارسی با موضوع فاز اول مطالعاتی ایجاد پیکره متنی زبان فارسی		
	عنوان زیرپروژه:		
	تاریخ: 1388/03/31	ویرایش: 1/0	کد زیرپروژه: پیکرمتن فارس - 2 - ب

فهرست مطالب



شماره صفحه	عنوان
5	1. مقدمه
7	1-1 تعریف مساله
9	2. پیش زمینه
9	1-2 ریخت‌شناسی (morphology)
10	1-1-2 ریخت‌شناسی حالت-متناهی
11	2-1-2 ریخت‌شناسی مبتنی بر الگوریتم porter
11	3-1-2 روش‌های مکمل
12	2-2 برچسب‌گذاری
13	3-2 هم‌شکلی
14	4-2 سطوح بالاتر پردازش
15	3. مبانی نظری زبان‌ها و اتومات‌ها
16	1-3 مجموعه
17	2-3 عبارات منظم
17	3-3 گرامرها
18	4-3 اتومات‌ها
19	5-3 مقایسه‌ی ابزارهای بیان فعل
20	4. مبانی فعل در ادبیات فارسی
20	1-4 اقسام فعل
21	1-1-4 فعل‌های ساده
22	2-1-4 فعل‌های پیشوندی
22	3-1-4 فعل‌های مرکب
22	4-1-4 فعل‌های پیشوندی مرکب

	عنوان پروژه:		 شورای عالی اطلاع رسانی
	فاز اول طرح جامع پیکره زبان فارسی با موضوع فاز اول مطالعاتی ایجاد پیکره متنی زبان فارسی		
	عنوان زیرپروژه:		
	تاریخ: 1388/03/31	ویرایش: 1/0	کد زیرپروژه: پیکرمتن فارس - 2 - ب
بررسی راهکارهای ایجاد ماشین تشخیص هوشمند افعال مرکب در زبان فارسی			

شماره صفحه	عنوان
22	5-1-4 عبارت‌های فعلی
23	6-1-4 فعل‌های لازم یک شخصه
23	2-4 صرف فعل
24	3-4 فعل‌های خاص
24	1-3-4 فعل ربطی (اسنادی)
24	2-3-4 فعل‌های شبه‌کمکی
25	3-3-4 فعل غیر شخصی
25	4-3-4 فعل وصفی
25	5-3-4 حذف فعل به قرینه
26	5. الگوی صرف افعال در زبان فارسی
28	1-5 فعل‌های ساده
28	1-1-5 ماضی ساده
30	2-1-5 ماضی استمراری
33	3-1-5 ماضی بعید
35	4-1-5 ماضی مستمر
37	5-1-5 ماضی ساده نقلی
39	6-1-5 ماضی استمراری نقلی
41	7-1-5 ماضی بعید نقلی
43	8-1-5 ماضی مستمر نقلی
45	9-1-5 ماضی التزامی
47	10-1-5 مضارع اخباری
49	11-1-5 مضارع مستمر
50	12-1-5 مضارع التزامی
53	13-1-5 آینده
55	14-1-5 امر
57	15-1-5 خلاصه الگوهای افعال ساده

	عنوان پروژه:		 شورای عالی اطلاع رسانی
	فاز اول طرح جامع پیکره زبان فارسی با موضوع فاز اول مطالعاتی ایجاد پیکره متنی زبان فارسی		
	عنوان زیر پروژه:		
	تاریخ: 1388/03/31	ویرایش: 1/0	کد زیر پروژه: پیکرمتن فارس - 2 - ب
بررسی راهکارهای ایجاد ماشین تشخیص هوشمند افعال مرکب در زبان فارسی			

شماره صفحه	عنوان
58.....	2-5 فعل‌های پیشوندی.....
60.....	3-5 فعل‌های مرکب.....
62.....	4-5 فعل‌های پیشوندی مرکب.....
65.....	5-5 عبارت‌های فعلی.....
65.....	6-5 فعل‌های لازم یک شخصه.....
66.....	7-5 فعل‌های خاص.....
66.....	1-7-5 فعل ربطی (اسنادی).....
68.....	2-7-5 فعل‌های شبه‌کمکی.....
68.....	3-7-5 فعل غیر شخصی.....
68.....	4-7-5 فعل وصفی.....
69.....	5-7-5 حذف فعل به قرینه.....
70.....	8-5 ویژگی‌های فعل مرکب فارسی.....
71.....	6. روش سلسله مراتبی در تشخیص فعل.....
72.....	1-6 طراحی ساختار داده.....
73.....	2-6 الگوریتم.....
74.....	3-6 پایگاه داده الگو.....
75.....	7. مراجع.....

	عنوان پروژه:		 شورای عالی اطلاع رسانی
	فاز اول طرح جامع پیکره زبان فارسی با موضوع فاز اول مطالعاتی ایجاد پیکره متنی زبان فارسی		
	عنوان زیر پروژه:		
تاریخ: 1388/03/31	ویرایش: 1/0	کد زیر پروژه: پیک‌متن‌فارس - 2 - ب	
		بررسی راهکارهای ایجاد ماشین تشخیص هوشمند افعال مرکب در زبان فارسی	



1. مقدمه

در نرم‌افزارهای پردازش زبان طبیعی مانند: غلطیاب نحوی، مدلسازی معنایی، استخراج اطلاعات، و ترجمه ماشینی، یکی از مراحل اولیه و تاثیرگذار، تشخیص «اجزای کلام» یا (Part of Speech (POS در جمله است. در تعیین اجزای کلام، تشخیص فعل مهمترین و تاثیرگذارترین بخش انجام کار است. همچنین در میان نقش‌های دستوری، فعل بیشترین حالت صرفی را در مقایسه با سایرین به خود می‌گیرد. فارغ از دنیای نرم‌افزار و در حوزه ادبیات فارسی، مساله «تشخیص فعل در جمله» دارای پیچ و خم‌ها و نکات ظریفی است که گاهی بین ادب‌دانان اختلاف نظر در تشخیص فعل به وجود می‌آورد. در حوزه نرم‌افزار مساله طور دیگری مطرح می‌شود. در اینجا تمرکز بر روی صورت ظاهری یا مدل صرف افعال است، که این صورت ظاهری، در نتیجه کار در حوزه ادبیات استخراج و ارایه می‌شود؛ آنگاه مبانی نظری و پیچیدگی‌های تکنیکی آن، در حوزه نرم‌افزار مطالعه و بررسی می‌شود؛ لذا در این تحقیق ما به پیچیدگی‌های دستوری نخواهیم پرداخت.

البته در حالت کلی نمی‌توان تشخیص فعل در جمله را بدون توجه به سایر کلمات در جمله انجام داد؛ خصوصاً در مورد افعال مرکب؛ چرا که ابهامات مختلفی ممکن است در اثر مشابهت ظاهری یا عوامل دیگر بوجود آیند که رفع ابهام آن بستگی به نقش سایر کلمات دارد. با این وجود - مانند بسیاری از کارهای قبلی - ما در اینجا «مساله تشخیص فعل» و بطور خاص «مساله تشخیص فعل مرکب» را بطور مستقل مورد مطالعه قرار می‌دهیم. ارزش این کار از این جهت است که در بسیاری از موارد می‌توان فعل را بطور مستقل تشخیص داد و در مواردی که نمی‌توان مستقل عمل کرد، می‌توان با احتمال قابل قبولی فعل را تشخیص داد و نتیجه را برای پردازش‌های سطح بالاتر ارایه داد.

در این تحقیق، روش کار ما برای تشخیص «افعال»، و به طور خاص «افعال مرکب»، بر ریخت‌شناسی حالت-متناهی (finite-state morphology) استوار است که در آن، قواعد زبان با استفاده از عبارات منظم تعریف می‌شوند. در این تحقیق کار ما بر سه محور استوار است:

- کارهای انجام شده در زمینه ریخت‌شناسی و برچسب‌گذاری افعال فارسی را مرور می‌کنیم.
- الگوی جامع صرف (ریخت‌شناسی) افعال در زبان فارسی استخراج و با استفاده از عبارات‌های منظم ارایه می‌کنیم.
- یک روش سلسله‌مراتبی برای تشخیص افعال بر اساس ماشین‌های حالت-متناهی ارایه می‌کنیم.

	عنوان پروژه:		 شورای عالی اطلاع رسانی
	عنوان زیر پروژه:		
	تاریخ: 1388/03/31	ویرایش: 1/0	

عنوان پروژه: فاز اول طرح جامع پیکره زبان فارسی با موضوع فاز اول مطالعاتی ایجاد پیکره متنی زبان فارسی

عنوان زیر پروژه:

بررسی راهکارهای ایجاد ماشین تشخیص هوشمند افعال مرکب در زبان فارسی

تاریخ: 1388/03/31

ویرایش: 1/0



کد زیر پروژه: پیک متن فارس - 2 - ب

در وهله اول ما کارهای انجام شده در حوزه ریخت‌شناسی (morphology) و برچسب‌گذاری (tagging) متون فارسی را بررسی کرده و نقاط قوت و ضعف هر کدام را بیان می‌کنیم. اگرچه تمرکز ما به طور خاص بر روی تشخیص افعال مرکب است، اما از دیدگاه روش‌شناسی فرایند تشخیص افعال اشتراکات زیادی با سایر انواع کلمات دارد که می‌تواند تحت یک دیدگاه جامع بررسی شوند.

بخش عمده دیگر کار ما شامل ارزیابی الگوی جامع صرف افعال فارسی می‌شود. البته ما مدعی نیستیم که تمام ساختارهای فعلی در این مقاله ارزیابی داده می‌شوند اما این تلاش انجام شده است و بخش قابل توجهی از ساختارهای زنده و متداول در زبان فارسی ارزیابی شده‌اند. کارهای زیادی در زمینه‌ی تشخیص افعال در زبان فارسی انجام شده‌اند اما هیچ کدام از آنان در بیان الگوهای صرف فعل فارسی جامعیت ندارند؛ یعنی صرفاً به بیان روش پرداخته‌اند و بصورت موردی برخی الگوها را ارزیابی کرده‌اند. در [1] سعی شده است تا مدل یکپارچه‌ای برای ریخت‌شناسی کلمات، از جمله افعال، در زبان فارسی ارزیابی شوند که به هیچ وجه جامع و کامل نیست.

همچنین در این تحقیق ما روشی برای طراحی و پیاده‌سازی کارآمد و بهینه تشخیص فعل از جنبه‌های ذخیره‌سازی، بازیابی و پردازشی، ارزیابی می‌دهیم. این روش براساس مدل ماشین حالت-متناهی است که الگوهای صرف افعال توسط عبارات منظم تعریف می‌شوند و در یک ساختار سلسله‌مراتبی سازمان‌دهی می‌شوند. این سازماندهی از نظر کارایی بسیار بهتر از سایر سیستم‌های تشخیص الگو - که در آنها هیچ ارتباطی بین الگوها نیست - عمل می‌کند.

در ادامه، ابتدا در بخش 1-1 به بیان حوزه مساله می‌پردازیم و حد و مرز کار را مشخص می‌کنیم. در بخش 0 به مرور کارهای انجام شده در زمینه تشخیص فعل یا بطور کلی پردازش ریخت‌شناسی و برچسب‌گذاری در زبان فارسی می‌پردازیم. در بخش 0 پیش‌نیازهای نظری در حوزه نظریه زبان‌ها و ماشین‌ها را ارزیابی می‌دهیم و از دید کلی جایگاه مساله‌ی تشخیص فعل را در میان زبان‌ها طبقه‌بندی می‌کنیم و ابزارهای لازم برای ارزیابی الگو را دسته‌بندی و مقایسه می‌کنیم. در بخش 4 به بیان پیش‌نیازهای لازم در حوزه ادبیات فارسی می‌پردازیم. توضیحات بخش **Error! Reference source not found.** به صورت خلاصه ارزیابی شده‌اند و برای تسلط بیشتر به مبانی ادبی، خواننده باید به کتب تخصصی این زمینه مراجعه کند [2, 3]. در بخش 5 که مهمترین بخش این تحقیق است، سعی شده است تا تمام الگوهای ساختاری صرف فعل در زبان فارسی استخراج و ارزیابی شوند. در این بخش، الگوی صرف برای تمام ساختارهای فعل فارسی بر اساس [3]، و برای تمام زمان‌های صرفی استخراج و ارزیابی

	عنوان پروژه:		 شورای عالی اطلاع رسانی
	فاز اول طرح جامع پیکره زبان فارسی با موضوع فاز اول مطالعاتی ایجاد پیکره متنی زبان فارسی		
	عنوان زیر پروژه:		
	تاریخ: 1388/03/31	ویرایش: 1/0	کد زیر پروژه: پیکرمتن فارسی - 2 - ب

شده است. در بخش 00 ما روش خود را برای تشخیص فعل ارایه و بسط داده‌ایم و مسایل مربوط به طراحی و پیاده‌سازی آن را بررسی کرده‌ایم.

1-1 تعریف مساله



از دیدگاه ادبیات فارسی فعل در 6 ساختار شناسایی شده است [3]. این ساختارها عبارتند از: فعل‌های ساده، فعل‌های پیشوندی، فعل‌های مرکب، فعل‌های پیشوندی مرکب، عبارت‌های فعلی، و فعل‌های لازم یک شخصه. ما تمام ساختارهای دیگر، بجز افعال ساده را به عنوان فعل مرکب در نظر می‌گیریم و تعریف خود را از افعال ساده و مرکب را به صورت زیر ارایه می‌دهیم:

- افعال ساده شامل یک «ریشه فعل» هستند که با کمک افعال کمکی و پیشوندها و پسوندها صرف می‌شود.
 - افعال مرکب شامل یک «ریشه فعل» و یک «بخش ثابت فعل» هستند که صرف نمی‌شود و همیشه ثابت است. این بخش ثابت ممکن است شامل یک یا چند کلمه دیگر باشد. ممکن است بین بخش‌های فعل فاصله باشد یا ترتیب بخش‌ها عوض شود.
 - در واقع اقسام فعلی مرکب، از نظر ظاهری، یک بخش صرف شونده دارند و یک بخش ثابت. پیش‌نیاز تعریف الگو برای افعال مرکب، تعریف الگو برای افعال ثابت است.
- «در واقع در این مقاله، ما الگوی صرف برای تمام ساختارهای فعل فارسی ارایه داده و روش جامعی برای تشخیص فعل (مرکب) ارایه می‌دهیم.»



برخی از مسایلی که در تشخیص فعل با آنها سروکار نداریم به صورت زیرند:

- غلط املائی
- مشکلات ناشی رسم‌الخط فارسی؛ مانند وجود فاصله یا نیم فاصله، و یونیکدهای مختلف.
- فرض ما بر اینست که آنها در یک مرحله پیش‌پردازی حل می‌شوند. همچنین در این مقاله به مورد:
- بیان محاوره‌ای کلمات

نمی‌پردازیم. در زبان محاوره‌ای، افعال بگونه‌ای تغییر می‌یابد که با نسخه نوشتاری آن بسیار متفاوت است؛ مثلاً «باشد» گفته می‌شود «باشه» یا «بروم» گفته می‌شود «برم». ما در اینجا فقط «بیان رسمی و

	عنوان پروژه: فاز اول طرح جامع پیکره زبان فارسی با موضوع فاز اول مطالعاتی ایجاد پیکره متنی زبان فارسی		 شورای عالی اطلاع رسانی
	عنوان زیرپروژه: بررسی راهکارهای ایجاد ماشین تشخیص هوشمند افعال مرکب در زبان فارسی		
	تاریخ: 1388/03/31	ویرایش: 1/0	

کتابی افعال» را مورد بررسی قرار می‌دهیم. تشخیص و شناسایی «بیان محاوره‌ای» خود مقوله‌ی دیگری است که نیاز به مطالعه جداگانه دارد.

	عنوان پروژه: فاز اول طرح جامع پیکره زبان فارسی با موضوع فاز اول مطالعاتی ایجاد پیکره متنی زبان فارسی		 شورای عالی اطلاع رسانی
	عنوان زیر پروژه: بررسی راهکارهای ایجاد ماشین تشخیص هوشمند افعال مرکب در زبان فارسی		
	تاریخ: 1388/03/31	ویرایش: 1/0	

2. پیش زمینه



مساله‌ی تشخیص افعال و بطور کلی تشخیص اجزای کلمه در جمله، تحت عناوین «ریخت‌شناسی» و «برچسب‌گذاری» در کارهای مختلف مورد مطالعه قرار گرفته است. این دو زمینه، دو دیدگاه متفاوت در برخورد با مساله هستند که می‌توانند به حل مساله نایل شوند. از نظر مفهومی برچسب‌گذاری قبل از ریخت‌شناسی باید انجام شود؛ یعنی ابتدا باید کارکرد کلمه در جمله مشخص شود سپس ریخت‌شناسی آن صورت گیرد. اما این دیدگاه بطور کلی صحیح نیست و ریخت‌شناسی کلمات می‌تواند به تعیین برچسب کلمه کمک کند. در واقع این یک ارتباط دوسویه است. لذا ما به هر دو مقوله می‌پردازیم.

2-1 ریخت‌شناسی (morphology)

ریخت‌شناسی بخشی از علم پردازش زبان طبیعی است که به ساختارهای کلمات و ریشه‌یابی واژگان می‌پردازد. به عمل بیرون آوردن ریشه اصلی یک واژه؛ ریشه‌یابی (stemming) گویند. در واقع ریخت‌شناسی به علم شناختن اجزای معنی‌دار از یک واژه گویند که آن واژه را می‌سازد؛ به این اجزای معنی‌دار تکواژ (morpheme) گویند.

دو نوع ریخت‌شناسی وجود دارد. نوع اول آن را که بیشتر در زبان‌های لاتین مانند انگلیسی مرسوم است، ریخت‌شناسی الحاقی (concatinative) گویند. گونه دیگر را ریخت‌شناسی غیرالحاقی می‌گویند. یکی از انواع ریخت‌شناسی غیرالحاقی، ریخت‌شناسی الگویی (templatic) و یا ریشه و الگویی (root and pattern) است؛ که در زبان‌هایی مانند عربی، عبری و فارسی این نوع از ریخت‌شناسی کاربرد بسیاری دارد. در ریخت‌شناسی الحاقی تمام اشکال یک واژه در یک کلمه نمود پیدا می‌کند در حالی‌که در ریخت‌شناسی غیرالحاقی، ممکن است که شکل یک واژه در سه مکان جدا از هم در یک جمله پیدا شوند.

در ریخت‌شناسی، واژه‌ها به دو طریق بسط می‌یابند: تصریف (inflection) و اشتقاق (derivation). در تصریف، از ترکیب یک واژه با اجزای دستوری دیگر، واژه‌ای جدید در همان نوع و رده واژه قبلی ایجاد می‌گردد. به عنوان مثال علامت جمع «ها» در فارسی که با اضافه کردنش به هر اسمی یک اسم جدید به وجود می‌آید؛ مثلاً واژه «کتاب» با اضافه شدن «ها» به «کتاب‌ها» تبدیل می‌شود که در این صورت، هم کتاب از نوع دستوری اسم است و هم کتاب‌ها. روش دوم، روش اشتقاق است. در اشتقاق با افزودن یک جز دستوری به یک واژه، یک واژه در رده جدیدی به وجود می‌آید. به عنوان مثال اگر تکواژ «-ش» را به

	عنوان پروژه:		 شورای عالی اطلاع رسانی
	فاز اول طرح جامع پیکره زبان فارسی با موضوع فاز اول مطالعاتی ایجاد پیکره متنی زبان فارسی		
	عنوان زیر پروژه:		
	تاریخ: 1388/03/31	ویرایش: 1/0	کد زیر پروژه: پیکرمتن فارسی - 2 - ب

واژه مصدری «کن» اضافه کنیم، واژه کنش به وجود می‌آید که واژه جدید دیگر از نوع مصدر نیست و یک اسم است. در افعال تولید کلمات از نوع تصریفی است.



2-1-1 ریخت‌شناسی حالت-متناهی

ریخت‌شناسی بر اساس مدل حالت-متناهی، روش متداولی است که در [4, 5] نمونه‌ای از آن را می‌توان دید. اساس کار آنها بر یک مدل زبان جهانی است که در [6] ارائه شده است. تعریف الگو بر اساس عبارات منظم انجام می‌شود و پیاده‌سازی آن بر اساس مدل ماشین حالت-متناهی است.

یکی از روش‌های پردازش ریخت‌شناسی زبان، استفاده از نرم‌افزارهای غیروابسته به زبان است [6]. این نرم‌افزارها در واقع ویژگی‌های ریخت‌شناسی زبان جهانی را در خود جای داده‌اند. اساس کار این نرم‌افزارها بر ریخت‌شناسی حالت-متناهی استوار است. برای استفاده از این گونه نرم‌افزارها در هر زبان خاص، کافی است که آنها را برای آن زبان تنظیم کرد؛ یعنی ویژگی‌های مربوط به آن زبان را در محیط نرم‌افزار تعریف کرد. یکی از این نرم‌افزارها، Xerox است که در [6] معرفی شده است. در [4] با استفاده از این نرم‌افزار برخی از الگوهای فعلی واسمی در زبان فارسی پیاده‌سازی شده‌اند. در این نرم‌افزار امکان تعریف قوانین تصریف (یا اشتقاق) و ترکیب وجود دارد. اشکال عمده اینگونه نرم‌افزارها در این است که به دلیل جهانی بودن مدل آنها، کارایی آنها پایین است و تعریف الگو در آنها مشکل است.

طراحی پردازشگر ریخت‌شناسی حالت-متناهی را می‌توان به دو بخش مجزا تقسیم کرد: بخش مربوط به «طرح زبانی» و بخش مربوط به «طرح رایانه‌ای» [4]. منظور از طرح زبانی، ارائه توصیف نظری جامع و کامل و مانع از ریخت‌شناسی افعال در زبان فارسی است. ارائه توصیف جامع از ریخت‌شناسی در زبان فارسی به گونه‌ای که قابل کاربرد در برنامه‌های رایانه‌ای باشد، نخستین گام جهت طراحی برنامه‌های کاربردی است. این توصیف می‌بایست تمام صورت‌های تصریفی فعل در زبان فارسی را ارائه دهد. بخش دوم، طرح رایانه‌ای است. در این بخش نیازهای سخت‌افزاری و نرم‌افزاری پردازشگر تعریف می‌شود، طرح زبانی پیاده‌سازی شده و ویژگی‌ها و ساختار داخلی فایل‌های برنامه تشریح می‌گردد.

الگوریتم ماشین حالت-متناهی (یا DFA) راه حل سرراستی برای پیاده‌سازی ریخت‌شناسی است که در [7] نیز مورد استفاده قرار گرفته است.

	عنوان پروژه:		 شورای عالی اطلاع رسانی
	فاز اول طرح جامع پیکره زبان فارسی با موضوع فاز اول مطالعاتی ایجاد پیکره متنی زبان فارسی		
	عنوان زیرپروژه:		
	تاریخ: 1388/03/31	ویرایش: 1/0	کد زیرپروژه: پیکرمتن فارس - 2 - ب

نقطه قوت روش حالت-متناهی قابلیت تعریف الگو برای ریخت‌شناسی غیرالحاقی است. در این روش امکان تعریف الگو برای افعال چند تکه وجود دارد. روشی که در این تحقیق ارایه شده و بر آن تاکید می‌شود نیز بر رهیافت ریخت‌شناسی حالت-متناهی استوار است.

2-1-2 ریخت‌شناسی مبتنی بر الگوریتم porter



دسته‌ای دیگر از کارها، شامل الگوریتم‌های ریشه‌یابی هستند که بر اساس قوانین ریخت‌شناسی زبان مربوطه کار می‌کنند [8-10]. در این الگوریتم‌ها، برنامه از درون ساختاری تصمیم‌گیرنده مانند یک فلوجارت عبور کرده و با افزودن و کاستن وندها با رعایت قواعد املائی و دستوری، سعی در یافتن ریشه کلمات یا بطور خاص افعال دارد. این کارها عمدتاً مشابه الگوریتم porter هستند که برای زبان انگلیسی طراحی شده است [11, 12]. مشکل این دسته از الگوریتم‌ها برای کلمات جدا از هم است. با توجه به اینکه در فارسی مرز دقیق کلمات مشخص نیست، برای کلمات چندپاره این روش‌ها خوب عمل نمی‌کنند. در [8] یک سیستم ساده مبتنی بر قانون برای افعال فارسی ارایه شده است.

2-1-3 روش‌های مکمل

روش‌های خودکار استخراج الگو یا تشخیص اجزای کلمات بدون استفاده از پایگاه داده، می‌توانند به عنوان روش مکمل در کنار روش‌های ریخت‌شناسی حالت-متناهی مورد استفاده قرار گیرند. در دو جنبه می‌توان از این روش‌ها کمک گرفت:

- استخراج الگو
- غنی سازی پایگاه داده

برخی کارها سعی در استخراج خودکار ریخت‌شناسی کلمات از متن داشته‌اند [13]. در این دسته از کارها قوانین ریخت‌شناسی از پردازش حجم زیادی از متون استخراج می‌شود. این کارها نمیتوانند ذاتاً دقیق باشند. از طرفی به دلیل تغییر و تبدیل حروف در اتصالات وندها استخراج قوانین ریخت‌شناسی با خطای زیادی همراه خواهد بود. اما به هر حال می‌توانند به توسعه دهنده نرم‌افزار کمک کنند تا تمام الگوهای تصریف را شامل کند.

	عنوان پروژه:		 شورای عالی اطلاع رسانی
	فاز اول طرح جامع پیکره زبان فارسی با موضوع فاز اول مطالعاتی ایجاد پیکره متنی زبان فارسی		
	عنوان زیر پروژه:		
	تاریخ: 1388/03/31	ویرایش: 1/0	کد زیر پروژه: پیکرمتن فارسی - 2 - ب
بررسی راهکارهای ایجاد ماشین تشخیص هوشمند افعال مرکب در زبان فارسی			

دسته دیگر از کارها، می‌توانند به غنی‌سازی پایگاه داده کمک کنند. هر اندازه که برای ساختن پایگاه داده غنی و جامع تلاش کنیم، باز هم ممکن است موارد بسیاری از قلم بیافتد. از آنجا که انعطاف‌پذیری زیادی در زبان فارسی برای ساخت افعال یا اصطلاحات جدید وجود دارد، روش‌های تشخیصی که مبتنی بر پایگاه داده نیستند می‌توانند برای تکمیل پایگاه داده مفید باشند.



در این تحقیق، ریخت‌شناسی افعال بر اساس مطالعات در حوزه ادبیات فارسی صورت گرفته و در شناسایی الگوها تماما عامل انسانی دخیل است. همچنین پایگاه داده بصورت دستی پر می‌شود.

2-2 برچسب‌گذاری

در برچسب‌گذاری، به کلمات تشکیل دهنده‌ی متن یک برچسب متناظر می‌شود که نشان‌دهنده نقش کلمات در جمله است. از طرف دیگر در تشخیص فعل، ما نیازمند جداسازی وندها و ریشه فعل و در نتیجه، تعیین زمان و شخص فعل هستیم. در [14] استاندارد خاصی برای برچسب‌گذاری استفاده شده و 12 مقوله اصلی برای زبان فارسی استفاده شده است. هر کدام از این مقولات اصلی، شامل «ویژگی‌های توصیه شده» و «ویژگی‌های خاص زبان است». یکی از مقولات اصلی فعل است که برای آن ویژگی‌های: موقعیت، نوع، حالت، شخص، شمار، و زمان دستوری تعیین شده است. در حالت ایده آل، برچسب‌گذاری فعل در واقع منجر به تشخیص فعل و زمان و شخص آن است.

رویکردهای مختلفی در برچسب‌گذاری وجود دارند که می‌توان آنها را در دو دسته کلی طبقه‌بندی کرد: «رهیافت آماری» و «رهیافت مبتنی بر قانون». در رهیافت آماری از پیکره‌های برچسب‌گذاری شده (annotated corpora) بهره می‌جویند و در رهیافت‌های غیر آماری و مبتنی بر قانون که بر مبنای یادگیری ماشینی و دانش بشری استوار هستند. نمونه‌ای از روش‌های آماری در [15, 16] گزارش شده‌اند.

عمده‌ترین اشکال روش‌های آماری، نبودن منابع متنی کافی و مناسب برای پردازش است که باعث می‌شود دقت این روش‌ها در حد مطلوب نباشد. از طرف دیگر با توجه به مشکلات رسم‌الخط در زبان فارسی تشخیص مرز کلمات مشکل است. در [17] برخی چالش‌هایی که در توسعه یک برچسب‌گذار فارسی وجود دارد بیان شده‌اند.

	عنوان پروژه:		 شورای عالی اطلاع رسانی
	فاز اول طرح جامع پیکره زبان فارسی با موضوع فاز اول مطالعاتی ایجاد پیکره متنی زبان فارسی		
	عنوان زیر پروژه:		
	تاریخ: 1388/03/31	ویرایش: 1/0	کد زیر پروژه: پیکرمتن فارسی - 2 - ب
بررسی راهکارهای ایجاد ماشین تشخیص هوشمند افعال مرکب در زبان فارسی			

درصد بالایی از کلمات از نظر برچسب ابهام دارند، زیرا کلمات در متون مختلف برچسب‌های متفاوت دارند. بنابراین در برچسب‌گذاری، عمل ابهام زدایی از برچسب‌ها در متون مختلف انجام می‌شود. همچنین روش‌های برچسب‌گذاری عمدتاً در مورد کلمات جدا از هم با مشکل مواجه می‌شوند.

2-3 هم‌شکلی



یکی دیگر از مسائلی که در زمینه تشخیص فعل مطرح است وجود ابهام یا مشابهت ظاهری کلمات است. هم‌شکلی (Homograph) یعنی تشابه ظاهری دو کلمه که نقش دستوری و معنایی متفاوتی دارند. مساله هم‌شکلی، در برچسب‌گذاری و ریخت‌شناسی باعث بروز مشکل می‌شود.

هم‌شکلی در زبان فارسی به سبب نبودن اعراب بیشتر اتفاق می‌افتد. بطور خاص در مورد افعال ممکن است در موارد زیر تشابه اتفاق بیافتد:

- حالت‌های صرفی مختلف یک فعل؛ مانند: می‌ترسید (ماضی ساده و مضارع اخباری).
- حالت‌های مختلف صرفی دو فعل؛ مانند: نشستند (نشستن و نشستن).
- تشابه کلمه با فعل؛ مانند: بزن (فعل امر، صفت).

روش‌های مختلفی برای رفع ابهام در زبان‌های دیگر پیشنهاد شده‌اند. اساس این روش‌ها بر کارهای آماری است. در [9, 18] روش‌های متفاوتی برای رفع ابهام ارائه شده است.

در این تحقیق ما به مقوله رفع ابهام نمی‌پردازیم، بلکه آنرا به مراحل پردازشی بعد ماکول می‌کنیم. ما تمام کلمات و ترکیباتی که امکان فعل بودن را دارند شناسایی کرده و به پردازش سطح بالا انتقال می‌دهیم (حتی اگر این ترکیبات با هم اشتراک داشته باشند). در مرحله بعدی با استفاده از روش‌های آماری و پیکره‌های متنی، این رفع ابهام انجام می‌پذیرد.



	عنوان پروژه: فاز اول طرح جامع پیکره زبان فارسی با موضوع فاز اول مطالعاتی ایجاد پیکره متنی زبان فارسی		 شورای عالی اطلاع رسانی
	عنوان زیر پروژه: بررسی راهکارهای ایجاد ماشین تشخیص هوشمند افعال مرکب در زبان فارسی		
	تاریخ: 1388/03/31	ویرایش: 1/0	

2-4 سطوح بالاتر پردازش

هدف از تشخیص فعل، ارایه آن به سطوح پردازشی بالاتر است. در سطوح بالاتر می‌توان جمله را از نظر نحوی و معنایی تحلیل کرد. نحو (syntax) که به ارتباط کلمات به همدیگر و مباحث دستوری آن‌ها در گروه‌ها و جملات می‌پردازد. معناشناسی (semantics) که به ارتباطات معنایی کلمات می‌پردازد. برای آشنایی با چگونگی پردازش نحوی و معنایی جملات می‌توانید به [19] مراجعه کنید که برای پردازش زبان انگلیسی می‌باشد.

در زبان فارسی به دلیل مشخص نبودن مرز کلمه، مرز بین پردازش واژه و پردازش نحوی بطور دقیق مشخص نیست. مثلاً کلماتی وجود دارند که در واقع یک کلمه‌اند ولی گاهی جدا نوشته می‌شوند مانند: «فرمان برداری»، و برعکس گاهی یک کلمه معادل چند کلمه است؛ مانند «بکار» به جای «به کار». این مشکل در مورد افعال و بطور خاص افعال مرکب بسیار بروز می‌کند. اگر پردازشگر واژه موفق به تشخیص بخش‌های چندپاره شود، عملکرد پردازش سطح بالاتر بهتر می‌شود، اما اگر نتوان بخش‌های فعل را در پردازش واژه‌ای تشخیص داد این به عهده پردازش سطح بالا است. به عنوان مثال، در مورد فعل‌های چندپاره، بسیاری از کارهای انجام شده‌ی قبلی، این کار را به پردازش‌گر نحوی می‌سپارند تا ارتباط معنایی اجزای فعل را تشخیص دهند. معدودی از کارها تا حد کمی به این مساله در سطح ریخت‌شناسی پرداخته‌اند [4, 5].

در این گزارش سعی ما بر یافتن ارتباطات اجزای فعل مرکب در سطح ریخت‌شناسی است.

	عنوان پروژه: فاز اول طرح جامع پیکره زبان فارسی با موضوع فاز اول مطالعاتی ایجاد پیکره متنی زبان فارسی		 شورای عالی اطلاع رسانی
	عنوان زیرپروژه: بررسی راهکارهای ایجاد ماشین تشخیص هوشمند افعال مرکب در زبان فارسی		
	تاریخ: 1388/03/31	ویرایش: 1/0	

3. مبانی نظری زبان‌ها و اتومات‌ها

در این بخش به بیان مبانی نظری لازم برای ارایه الگوی تصریف و پیاده‌سازی رایانه‌ای الگوریتم تشخیص فعل که در بخش 0 ارایه شده می‌پردازیم.



در علوم نظری کامپیوتر (نظریه زبان‌ها و اتومات‌ها)، یک «زبان» (language) شامل مجموعه‌ای از کلمات (word) است؛ هر کلمه شامل یک رشته از نمادها یا حروف (letters) است. یک زبان را می‌توان به روش‌های مختلفی نشان داد [20]. ابزارهای بیان زبان به صورت زیرند:

- بیان توسط مجموعه: (1) نمایش مستقیم اعضا (2) نمایش توسط نمادها و عملوندهای ریاضی و رشته‌ای
- بیان توسط عبارت‌های منظم (Regular Expression)
- بیان توسط گرامرها (Grammar)
- بیان توسط اتومات‌ها (automata): انواع اتومات‌ها مانند: DFA, NFA, DPDA, NPDA، و تورینگ ماشین و ...

ما به «مساله تشخیص فعل»، به صورت یک «مساله تصمیم‌گیری» (Decision Problem) برای تعلق رشته به زبان نگاه می‌کنیم. لذا فعل را در قالب یک زبان تعریف می‌کنیم. سپس نوع این زبان را بدست می‌آوریم، و با توجه به نوع زبان، ابزار مناسب برای بیان زبان را به کار می‌بندیم. در اینجا فعلاً با پیچیدگی‌های پیاده‌سازی هر کدام از این ابزارها کار نداریم و فقط از جنبه نظری می‌خواهیم روش بیان فعل را فرمول‌بندی کنیم.

زبان‌ها در گروه‌ها یا مرتبه‌های مختلفی طبقه‌بندی می‌شوند [20]. برخی از این طبقه‌ها به قرار زیر هستند:

- زبان‌های بدون ستاره (Star-Free)
- زبان‌های منظم (Regular)
- زبان‌های مستقل از متن معین (Deterministic Context Free)
- زبان‌های مستقل از متن (Context Free)

	عنوان پروژه: فاز اول طرح جامع پیکره زبان فارسی با موضوع فاز اول مطالعاتی ایجاد پیکره متنی زبان فارسی		 شورای عالی اطلاع رسانی
	عنوان زیرپروژه: بررسی راهکارهای ایجاد ماشین تشخیص هوشمند افعال مرکب در زبان فارسی		
	تاریخ: 1388/03/31	ویرایش: 1/0	

- زبان‌های وابسته به متن (Context Sensitive)
- زبان‌های بازگشتی (Recursive)

در میان انواع ذکر شده در فوق رابطه زیرمجموعه‌گی برقرار است. یعنی زبان‌های بالایی زیرمجموعه زبان‌های پایینی هستند. هر کدام از ابزارهای بیان زبان، محدودیت‌ها و ویژگیهای خاصی دارند و استفاده از آنها بستگی به نوع کاربرد دارد. همچنین هر کدام از انواع زبان توسط ابزارهای خاصی قابل بیان هستند. هرچه نوع زبان محدودتر باشد ابزارهای بیان آن ساده‌ترند. در اینجا وقتی می‌گوییم یک زبان در طبقه زبان‌های مستقل از متن قرار دارد یعنی محدودترین طبقه‌ای که در آن جای می‌گیرد مستقل از متن است. طبعاً بدیهی است که اگر زبانی مستقل از متن باشد، بازگشتی نیز هست. در ادامه به ابزارهای بیان زبان می‌پردازیم.



3-1 مجموعه

بیان بصورت مجموعه در دو صورت (1) نمایش مستقیم اعضا و (2) نمایش توسط عملوندها و علائم ریاضی و رشته‌ای، بیان می‌شود. در مورد زبان‌های محدود (finite) می‌توان تمام اعضا را در یک مجموعه نشان داد. در مورد زبان‌های نامحدود (infinite) چند عضو اول را نمایش داده و الگوی تکرار اعضای ابتدایی می‌تواند اعضای بعدی را معین کند. در حالت دوم از عملوندها و علائم ریاضی برای بیان زبان استفاده می‌کنیم. استفاده از مجموعه برای بیان زبان دارای محدودیت‌های زیادی است. با این وجود می‌توان تمام زبان‌های محدود را توسط مجموعه بیان کنیم.

استفاده از روش مدل‌سازی مجموعه‌ای در پیاده‌سازی رایج است. به عنوان مثال برخی از نرم‌افزارهای غلطیاب املائی فارسی از این روش برای کار خود استفاده کرده‌اند [21]. به علت پیچیدگی الگوی کلمه‌سازی در زبان فارسی، استفاده از تمام حالت‌های صرفی در داخل مجموعه می‌تواند کار پیاده‌سازی را راحت‌تر بکند.

به عنوان مثال با استفاده از مجموعه، صیغه‌های فعل ماضی ساده از مصدر «گذشت» را به صورت زیر بیان می‌کنیم:

$$A = \{\text{گذشتم، گذشتی، گذشت، گذشتیم، گذشتید، گذشتند}\}$$

	عنوان پروژه: فاز اول طرح جامع پیکره زبان فارسی با موضوع فاز اول مطالعاتی ایجاد پیکره متنی زبان فارسی		 شورای عالی اطلاع رسانی
	عنوان زیرپروژه: بررسی راهکارهای ایجاد ماشین تشخیص هوشمند افعال مرکب در زبان فارسی		
	تاریخ: 1388/03/31	ویرایش: 1/0	

3-2 عبارات منظم

فقط زبان‌های منظم قابل بیان توسط عبارتهای منظم هستند. در واقع هر زبان منظمی را می‌توان با عبارات منظم بیان نمود و هر زبانی که بتوان آنرا با عبارت منظم بیان کرد یک زبان منظم است. عبارت منظم شامل نمادها و عملگرهای خاص خود است [20]. عبارات منظم در نمایش الگوی زبان بسیار خوانا هستند و نگارش آنها از دید کاربر انسانی بسیار راحت است.

اغلب الگوهای صرف افعال، که با آن سر و کار داریم، قابل بیان توسط عبارت منظم می‌باشند. در واقع هسته اصلی نمایش الگوی افعال در پیاده‌سازی ما توسط عبارات منظم صورت خواهد گرفت. به عنوان مثال با استفاده از عبارت منظم، الگوی صرف فعل ماضی استمراری نقلی از مصدر «گفتن» به صورت زیر است:

می‌گفت(ه) (ایم|اید|اند|ام|ای | است)؟

از علامت « » برای بیان فاصله و از «>» برای بیان نیم‌فاصله استفاده کردیم (الگو را از راست به چپ بخوانید). برای تشخیص سایر مصدرها می‌توان الگوی فوق را به صورت زیر تعمیم داد:



می‌گفت(خورد|زد|گرفت)ه (ایم|اید|اند|ام|ای | است)؟

در بالا فقط 4 بن به عنوان نمونه قرار داده شده‌اند. می‌توان تمامی بن‌های ماضی را در الگوی فوق قرار داد. با توجه به اینکه تعداد بن‌های ماضی فارسی محدود است، می‌توان آن را پیاده‌سازی کرد.

3-3 گرامرها

هر زبانی را میتوان توسط گرامرها بیان کرد. گرامرها خود نیز به انواع مختلف طبقه بندی می‌شوند که این طبقه‌بندی بر اساس شکل گرامر و همچنین توانایی گرامر برای بیان نوع زبان است. برای بیان گرامر نمادگذاری خاصی وجود دارد [20].

به عنوان مثال برای بیان الگوی صرف «ماضی استمراری نقلی» توسط گرامر به صورت زیر عمل می‌کنیم:

	عنوان پروژه:		 شورای عالی اطلاع رسانی
	فاز اول طرح جامع پیکره زبان فارسی با موضوع فاز اول مطالعاتی ایجاد پیکره متنی زبان فارسی		
	عنوان زیر پروژه:		
	تاریخ: 1388/03/31	ویرایش: 1/0	کد زیر پروژه: پیکرمتن فارس - 2 - ب
بررسی راهکارهای ایجاد ماشین تشخیص هوشمند افعال مرکب در زبان فارسی			

S → BME

B → می

E → ایم | اید | اند | ام | ای | است

M → RT

R → گفت | خورد | زد | گرفت

T → ه



برای خوانایی بیشتر، نمادهای فاصله و نیم‌فاصله نشان داده نشده‌اند. در قانون تولید پنجم، R، بن‌های موجود در پایگاه داده تولید می‌شوند.

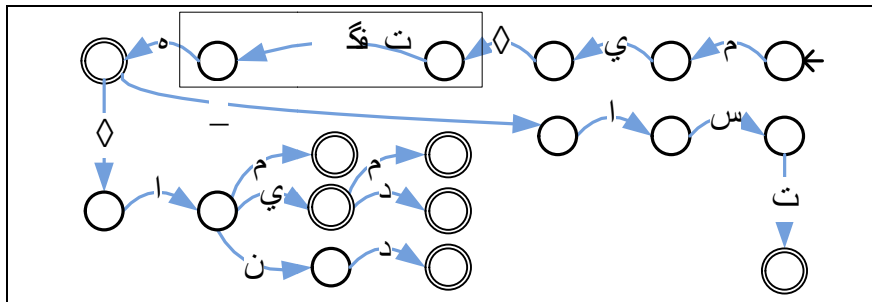
3-4 اتومات‌ها

اتومات‌ها انواع و اقسام مختلفی دارند که هر کدام از آن‌ها نوع خاصی از زبان را می‌توانند بیان کنند [20]. برخی از انواع اتومات‌ها عبارتند از:

- Deterministic Finite Automata (DFA)
- Non-Deterministic Finite Automata (NFA)
- Deterministic Pushdown Automata (DPDA)
- Non-deterministic Pushdown Automata (NPDA)
- Turing Machine (TM)

اتومات‌ها نزدیکترین ابزار به زبان‌های پیاده‌سازی هستند. اتومات‌های محدود، DFA و NFA، فقط قادر به بیان زبان‌های منظم هستند. اتومات‌های پشته‌ای، DPDA و NPDA، به ترتیب قادر به بیان زبان‌های «مستقل از متن قطعی» و «مستقل از متن» هستند. تورینگ ماشین‌ها، TM، قادر به بیان هر زبانی هستند. بطور نظری هر الگوریتم کامپیوتری توسط تورینگ ماشین قابل پیاده‌سازی است.

	عنوان پروژه: فاز اول طرح جامع پیکره زبان فارسی با موضوع فاز اول مطالعاتی ایجاد پیکره متنی زبان فارسی		 شورای عالی اطلاع رسانی
	عنوان زیر پروژه: بررسی راهکارهای ایجاد ماشین تشخیص هوشمند افعال مرکب در زبان فارسی		
	تاریخ: 1388/03/31	ویرایش: 1/0	



شکل 1 - الگوی صرف فعل ماضی استمراری نقلی از مصدر «گفتن» با استفاده از DFA



به عنوان مثال نمایش الگوی صرف فعل ماضی استمراری نقلی از مصدر «گفتن» با استفاده از DFA در شکل 1 نشان داده شده است.

در شکل 1 از نمادهای «_» و «>» به ترتیب برای بیان فاصله و نیم فاصله استفاده شده است. دایره‌های دوخطی به معنی وضعیت پایانی هستند. منطقه‌ای که توسط مستطیل مشخص شده نشان‌گر بن فعل است و در واقع پایگاه داده بن‌ها در این قسمت قرار می‌گیرد.

3-5 مقایسه‌ی ابزارهای بیان فعل

می‌توان افعال را توسط هر کدام از ابزارهای گفته‌شده در بخش‌های قبلی بیان کرد. در انتخاب ابزار مناسب برای بیان فعل باید به موارد زیر توجه کرد: (1) نوع زبان، (2) سادگی بیان برای انسان، (3) سادگی پیاده‌سازی برای ماشین.

بدیهی است که ابزار بیان باید بتواند نوع زبان را پوشش دهد. به عنوان مثال نمی‌توان زبان مستقل از متن را با استفاده از عبارت منظم توصیف کرد. برای دانستن محدوده ابزار مناسب برای بیان فعل باید نوع زبان آنرا شناسایی کنیم. الگوی تصریف افعال در زبان فارسی در طبقه زبان‌های «منظم» قرار دارند. علی‌رغم تمام پیچیدگی‌های تصریف افعال مرکب، تمام الگوها قابلیت بیان با عبارت منظم را دارند. البته این نتیجه‌گیری به خاطر محدود بودن افعال نیست - در تئوری هر مجموعه محدود منظم است. حتی با فرض بی‌نهایت بودن افعال الگوهای تصریف آنها منظم است.

	عنوان پروژه:		 شورای عالی اطلاع رسانی
	فاز اول طرح جامع پیکره زبان فارسی با موضوع فاز اول مطالعاتی ایجاد پیکره متنی زبان فارسی		
	عنوان زیر پروژه:		
تاریخ: 1388/03/31	ویرایش: 1/0	کد زیر پروژه: پیکرمتن فارس - 2 - ب	
بررسی راهکارهای ایجاد ماشین تشخیص هوشمند افعال مرکب در زبان فارسی			

با توجه به منظم بودن الگوی تصرف افعال، می توان آنها را توسط عبارات منظم، گرامرها و تمام اتوماتها بیان کرد. هر عبارت منظم را می توان توسط الگوریتمی به گرامر معادل آن تبدیل کرد. همچنین هر عبارت منظم را می توان توسط الگوریتمی به یک NFA معادل و هر NFA را به DFA تبدیل کرد.

عامل دیگر در انتخاب ابزار بیان، سادگی نگارش و فهم آن توسط کاربر انسانی است. در میان ابزارهای بیان، عبارت منظم ساده ترین و راحت ترین ابزار برای کاربر انسانی هستند. همچنین ساده ترین ابزارها برای پیاده سازی رایانه ای، NFA و DFA هستند. در عمل به علت کارایی بالا DFA بر NFA ترجیح دارد اما مصرف حافظه آن زیاد است.

با توجه به تعداد بن ها در زبان فارسی می توان یک پیاده سازی از NFA را برای الگوی تصریف افعال در درون حافظه داشت. با توجه به تعداد زیاد الگوهای تصریف در فارسی، ما در این تحقیق روشی سلسله مراتبی برای بررسی الگوها در بخش 0 ارائه می دهیم.



4. مبانی فعل در ادبیات فارسی

در هر ساختی از فعل، بخشی یا جزیی است که مفهوم کار یا حالت یا وجود یا اسناد از آن معلوم می شود. این جز را «بن» گویند. تمام صیغه های فعل از «بن فعل» و «قاعده صرف» آن بطور کاملاً قاعده مندی قابل استخراج هستند. هر فعلی دو بن دارد «بن ماضی» و «بن مضارع». اگرچه قواعدی برای تبدیل بن ماضی به بن مضارع وجود دارد، اما استثنائاتی نیز وجود دارد. در [3] افعال ساده فارسی در 8 گروه طبقه بندی شده اند که هر طبقه قوانین مخصوص خود را دارد. ما بن های افعال (ماضی و مضارع) را صرف نظر از قوانین تبدیل آنها به یکدیگر، در پایگاه داده نگه می داریم.

4-1 اقسام فعل

فعل را در زبان فارسی، از جهت ساختمان، به شش گروه می توان تقسیم کرد [3]:

- فعل های ساده

	عنوان پروژه: فاز اول طرح جامع پیکره زبان فارسی با موضوع فاز اول مطالعاتی ایجاد پیکره متنی زبان فارسی		 شورای عالی اطلاع رسانی
	عنوان زیر پروژه: بررسی راهکارهای ایجاد ماشین تشخیص هوشمند افعال مرکب در زبان فارسی		
	تاریخ: 1388/03/31	ویرایش: 1/0	

- فعل‌های پیشوندی
- فعل‌های مرکب
- فعل‌های پیشوندی مرکب
- عبارتهای فعلی
- فعل‌های لازم یک شخصه



هر کدام از این ساختارها در 14 زمان و 6 شخص صرف می‌شوند (البته برخی صیغه‌ها برای زمان‌ها و اشخاص وجود ندارند). ساختار هر کدام از این افعال در ادامه بطور مختصر (برای مراجعه سریع) بیان می‌شود. برای توضیحات بیشتر به [3] مراجعه کنید.

4-1-1-1 فعل‌های ساده

فعل‌های ساده فعل‌هایی هستند که بن مضارع آنها یک تکواژ است؛ مانند: رفتن، گفتن. این افعال در چند دسته قرار می‌گیرند:

- افعال عادی: گفتن، رفتن
- افعال جعلی: طلبیدن، رقصیدن
- افعال دارای چند صورت: زادن و زاییدن، شنیدن و شنفتن و شنودن
- افعال ساخته شده از مصدر با افزودن «اندن» یا «انیدن»: گنجیدن و گنجاندن، خوردن و خوراندن

از نظر قواعد صرف فعل، دسته‌های فوق هیچ فرقی با هم ندارند.

	عنوان پروژه: فاز اول طرح جامع پیکره زبان فارسی با موضوع فاز اول مطالعاتی ایجاد پیکره متنی زبان فارسی		 شورای عالی اطلاع رسانی
	عنوان زیر پروژه: بررسی راهکارهای ایجاد ماشین تشخیص هوشمند افعال مرکب در زبان فارسی		
	تاریخ: 1388/03/31	ویرایش: 1/0	

4-1-2 فعل های پیشوندی

فعل های پیشوندی فعل هایی اند که از یک پیشوند و یک فعل ساده ساخته شده اند. مهمترین پیشوندهایی که فعل پیشوندی می سازند عبارتند از: بر، در، فرو، فرا، باز، وا، ور؛ مانند: برداشتن، درافتادن، فرو رفتن، بازداشتن.

4-1-3 فعل های مرکب



فعل های مرکب فعل هایی هستند که از یک کلمه، که آنرا فعلیاری می نامند، با یک فعل ساده که همکرد نامیده می شود، ساخته می شوند و مجموعا معنی واحدی را می رسانند؛ مانند: آرایش کردن، تاسف خوردن. تشخیص فعل مرکب دارای پیچیدگی های دستوری است که با آنها سروکار نداریم، بلکه این افعال در پایگاه داده ذخیره می شوند.

4-1-4 فعل های پیشوندی مرکب

فعل های پیشوندی گاهی با کلمه ای ترکیب می شوند و معنی واحدی را بیان می کنند. معنی مزبور نسبت به معنی لغوی کلمه های سازنده غالبا مجازی است؛ مانند: دم در کشیدن، سر در آوردن، تن در دادن.

4-1-5 عبارتهای فعلی

عبارت فعلی به دسته ای از کلمات اطلاق می شود که از مجموع آنها معنی واحدی حاصل می شود. عبارتهای فعلی بیش از دو کلمه هستند که نخستین کلمه ترکیب، حروف اضافه است و مجموع عبارت نیز معمولا معنی مجازی دارد؛ مانند: از پای درآمدن، بر پا کردن.

	عنوان پروژه:		 شورای عالی اطلاع رسانی
	فاز اول طرح جامع پیکره زبان فارسی با موضوع فاز اول مطالعاتی ایجاد پیکره متنی زبان فارسی		
	عنوان زیر پروژه:		
	تاریخ: 1388/03/31	ویرایش: 1/0	کد زیر پروژه: پیکرمتن فارس - 2 - ب
بررسی راهکارهای ایجاد ماشین تشخیص هوشمند افعال مرکب در زبان فارسی			

4-1-6 فعل های لازم یک شخصه

مراد از این اصطلاح، فعل هایی است که به صورت لازم و فقط با ساخت سوم شخص مفرد به کار می روند و به جای شناسه، ضمیر پیوسته مفعولی و اضافی، شخص و زمان فعل را نشان می دهد. مانند: خوشم آمد. «م» ضمیر پیوسته به خوش، به جای شناسه، شخص فعل را معین می کند و «آمد» در همه اشکال فعل ماضی ساده به همان شکل باقی می ماند: خوشم آمد، خوشت آمد، خوشش آمد، ...



4-2 صرف فعل

فعل در زبان فارسی در 14 زمان و در 6 صیغه شخصی صرف می شود. در هر زمان، با توجه به «مجهول یا معلوم» بودن فعل و همچنین «منفی یا مثبت» بودن صیغه صرفی و همچنین «طرز بیان»، حالت های مختلفی ایجاد می شود. در صرف فعل، پیشوندها یا پسوندهایی به ریشه فعل افزوده می گردند. در برخی موارد هنگام افزایش پسوند یا پیشوند، حروف ریشه قلب به سایر حروف می گردند و یا حذف می شوند.

صرف فعل در تمام ساختارهای فعلی یکسان است و لذا یافتن الگوی صرف فعل حائز اهمیت است. در افعال مرکب فقط یک بخش فعل صرف می شود و بخش های دیگر ثابت می ماند.

طبق [2] افعال در زبان فارسی در 14 زمان زیر صرف می شوند.

- ماضی ساده: گفتم، گفتی، گفت، ...
- ماضی استمراری: می گفتم، می گفتی، می گفت، ...
- ماضی بعید: گفته بودم، گفته بودی، گفته بود، ...
- ماضی مستمر: داشتم می گفتم، داشتی می گفتی، داشت می گفت، ...
- ماضی ساده نقلی: گفته ام، گفته ای، گفته است، ...
- ماضی استمراری نقلی: می گفته ام، می گفته ای، می گفته است، ...
- ماضی بعید نقلی: گفته بوده ام، گفته بوده ای، گفته بوده است، ...
- ماضی مستمر نقلی: داشته ام می گفته ام، داشته ای می گفته ای، داشته است می گفته است، ...
- ماضی التزامی: گفته باشم، گفته باشی، گفته باشد، ...

	عنوان پروژه: فاز اول طرح جامع پیکره زبان فارسی با موضوع فاز اول مطالعاتی ایجاد پیکره متنی زبان فارسی		 شورای عالی اطلاع رسانی
	عنوان زیر پروژه: بررسی راهکارهای ایجاد ماشین تشخیص هوشمند افعال مرکب در زبان فارسی		
	تاریخ: 1388/03/31	ویرایش: 1/0	

- مضارع اخباری: می گویم، می گویی، می گوید، ...
- مضارع مستمر: دارم می گویم، داری می گویی، دارد می گوید، ...
- مضارع التزامی: بگویم، بگویی، بگوید، ...
- آینده: خواهم گفت، خواهی گفت، خواهد گفت، ...
- امر: بگو، بگوید

برخی از این زمان‌ها یا صیغه‌های خاص ممکن است امروزه مورد استفاده نباشند. یا مثلاً برخی از صیغه‌های زمانی و شخصی برای برخی افعال خاص استفاده نشوند. برای جزییات بیشتر به [2, 3] مراجعه کنید.

4-3 فعل‌های خاص



4-3-1 فعل ربطی (اسنادی)

افعال ربطی افعالی هستند که معنی کاملی ندارند و فقط برای اثبات یا نفی نسبت به کار می‌روند و معنای آنها با آوردن صفت یا کلمه‌ای دیگر کامل می‌شود [3]؛ مانند: هوا روشن است. معروفترین فعل‌های ربطی یا اسنادی عبارتند از: استیدن، بودن، شدن. همچنین افعال «گشتن» و «گردیدن» اگر به معنی «شدن» باشند، فعل ربطی به شمار می‌روند.

این افعال علاوه بر معنی اسنادی به عنوان فعل کمکی برای صرف بن‌ها نیز در متن ظاهر می‌شوند که نباید اشتباه شوند. الگوی صرف این افعال در معنی اسنادی محدود است و به طور جداگانه بررسی می‌شود.

4-3-2 فعل‌های شبه کمکی

فعل‌هایی مانند «بایستن»، «توانستن»، «شایستن»، «خواستن»، «شدن» که گاهی در جمله به پاره‌ای از فعل‌ها و بن ماضی فعل‌ها نوعی کمک معنایی و کاربردی می‌دهند، فعل شبه کمکی نامیده می‌شوند و

	عنوان پروژه:		 شورای عالی اطلاع رسانی
	فاز اول طرح جامع پیکره زبان فارسی با موضوع فاز اول مطالعاتی ایجاد پیکره متنی زبان فارسی		
	عنوان زیر پروژه:		
	تاریخ: 1388/03/31	ویرایش: 1/0	کد زیر پروژه: پیکرمتن فارس - 2 - ب

فعل همراه آن‌ها را فعل پیرو می‌گویند؛ مثلاً در جمله‌ی «سیما نتوانست به مسافرت برود»، «نتوانست» فعل شبه‌کمکی، و «برود»، فعل پیرو است.

فرق بین افعال کمکی و شبه‌کمکی در این است که افعال کمکی ارتباط تنگاتنگی با معنی فعل دارند و برای صرف فعل در زمان خاص به کار می‌روند. اما این ارتباط در افعال شبه‌کمکی وجود ندارد.

4-3-3 فعل غیر شخصی



فعل‌های شبه‌کمکی «توانستن»، «بایستن» و «شایستن» گاهی فعلی می‌سازند که بر شخص معینی دلالت نمی‌کنند؛ مانند «نتوان رفت»، «نباید گفت»، «نشاید رفت». این فعل‌ها را فعل‌های غیرشخصی می‌گویند.

4-3-4 فعل وصفی

فعل وصفی آن است که فعل را به صورت صفت مفعولی (= بن ماضی + ه) بیاورند. در این صورت، فعل دیگری در آخر جمله می‌آورند. ارزش فعل اخیر (از جهت شخص، شمار، زمان، و وجه) ارزش فعل وصفی را مشخص می‌کند؛ مانند «احمد به خانه رفته نهار خورد».

4-3-5 حذف فعل به قرینه

گاهی اوقات در افعال مرکب بخشی از فعل به قرینه لفظی یا معنوی حذف می‌شود. به عنوان مثال در جمله‌ی: «وزیر دفاع آمریکا روز گذشته با امیر کویت دیدار [کرد] و درباره ... مذاکره کرد»، بخشی از فعل «دیدار کردن» به قرینه لفظی حذف شده است.

	عنوان پروژه:		 شورای عالی اطلاع رسانی
	فاز اول طرح جامع پیکره زبان فارسی با موضوع فاز اول مطالعاتی ایجاد پیکره متنی زبان فارسی		
	عنوان زیر پروژه:		
	تاریخ: 1388/03/31	ویرایش: 1/0	کد زیر پروژه: پیکرمتن: فارس - 2 - ب



5. الگوی صرف افعال در زبان فارسی

در این بخش قوانین صرف فعل در زبان فارسی برای تمام ساختارهای فعل بطور کامل فرمول‌بندی شده و در قلب عبارات منظم بیان می‌شود. همانطور که در بخش قبل گفته شد 6 ساختار فعل وجود دارد. همه این ساختارها بجز «عبارت‌های فعلی» دارای الگوی یکسان برای تمام افعال متعلق به آن ساختار هستند. در مورد عبارت‌های فعلی - که ساختار کلی قانونمندی ندارند - بطور جداگانه روشی برای بیان و تشخیص آنها در بخش 7 ارائه می‌شود. در واقع در این بخش برای 5 ساختار فعل و 14 زمان و 2 حالت معلوم و مجهول (مجموعاً 140 حالت)، الگوی تصریف فعل ارائه می‌دهیم.

ابتدا الگوی صرف افعال ساده را ارائه می‌دهیم. الگوی صرف افعال ساده به عنوان پایه‌ای برای سایر ساختارهای فعل هستند که در آنها نیز بروز می‌کنند. لذا بیان الگوی افعال ساده به صورت مفصل به همراه توضیحات صورت می‌گیرد، اما در مورد ساختارهای دیگر، به بیان خلاصه الگوها اکتفا خواهیم کرد. نحوه بیان الگوها به این صورت است که ابتدا تمام الگوهای صرفی مربوط به یک زمان - مثلاً ماضی ساده - را استخراج کرده و به همراه مثال آنها ارائه می‌دهیم. سپس تمام این الگوها را تجمیع کرده و در قالب یک الگوی تصریف جامع، ارائه می‌دهیم. به ترتیب به آنها «زیرالگو» و «الگوی جامع» می‌گوییم. به هر کدام از زیرالگوها یک کد داده شده است. هر کدام از این کدها شامل بخش‌های زیر است:

- زمان اصلی: گذشته (G)، حال (H)، آینده و امر (A)
- زمان فرعی: یک شماره است که مشخص کننده زمان است.
- معلوم یا مجهول: I برای معلوم و T برای مجهول
- منفی یا مثبت: با P و N که به ترتیب بر مثبت و منفی دلالت می‌کنند، مشخص می‌شود.
- حالت بیان: با شماره مشخص می‌شود



علاوه بر زیرالگوهایی که برای هر بخش ارائه کردیم ممکن است زیرالگوهایی هم وجود داشته باشند. ما پر استفاده‌ترین زیرالگوها را به شمار آوردیم و زیرالگوهای کم استفاده را لیست نکردیم. لذا ممکن است مثال‌ها و زیرالگوها تمام حالت‌های صرفی را پوشش ندهند اما این جامعیت در الگوی جامع رعایت می‌شود. الگوی جامع شامل حالت‌های غلط نیز هست؛ این کار به دو دلیل صورت گرفته است؛ اول اینکه

	عنوان پروژه:		 شورای عالی اطلاع رسانی
	فاز اول طرح جامع پیکره زبان فارسی با موضوع فاز اول مطالعاتی ایجاد پیکره متنی زبان فارسی		
	عنوان زیر پروژه:		
	تاریخ: 1388/03/31	ویرایش: 1/0	کد زیر پروژه: پیکرمتن فارس - 2 - ب
بررسی راهکارهای ایجاد ماشین تشخیص هوشمند افعال مرکب در زبان فارسی			

بیان الگو راحتتر می‌شود و دوم اینکه در مواردی که بخواهیم الگوهای غلط را شناسایی و به کاربر اعلام کنیم این کار مفید واقع می‌شود.

علایم اختصاری که در بیان عبارات منظم در این بخش استفاده شده‌اند، در زیر آمده‌اند.

- [G]: بن ماضی
- [AG]: بن‌های ماضی که با الف شروع می‌شوند و «آ» به «ا» تبدیل شده است
- [BG]: بن ماضی که با الف شروع نمی‌شود
- [H]: بن مضارع
- [HA]: بن مضارعی که به الف ختم می‌شود
- [AH]: بن مضارعی که با الف شروع می‌شود
- [HB]: بن مضارعی که به الف ختم نمی‌شود
- [BH]: بن مضارعی که با الف شروع نمی‌شود
- [AHA]: بن مضارعی که با الف شروع و با الف خاتمه می‌یابد
- [AHB]: بن مضارعی که ابتدای آن الف است و انتهای آن الف نیست
- [BHA]: بن مضارعی که ابتدای آن الف نیست و انتهای آن الف است
- [BHB]: بن مضارعی که ابتدا و انتهای آن الف نیست
- [PFGS]: (م|ای|یم|ید|ند)؟
- [PFGN]: (یم|ید|ند|م|ای| است)؟
- [PFH]: (یم|ید|ند|م|ای|د)
- [Z]: (م|ات|ش|امان|تان|شان)؟
- [IMG]: بن‌های ماضی مجهول‌ساز: (شد|گشت|گردید)
- [IMH]: بن‌های مضارع مجهول‌ساز: (شو|گرد)
- {D}: بیانگر فاصله بین بخش‌های فعل

	عنوان پروژه:		 شورای عالی اطلاع رسانی
	فاز اول طرح جامع پیکره زبان فارسی با موضوع فاز اول مطالعاتی ایجاد پیکره متنی زبان فارسی		
	عنوان زیر پروژه:		
تاریخ: 1388/03/31	ویرایش: 1/0	کد زیر پروژه: پیکرمتن فارس - 2 - ب	

در مورد فعل‌های کمکی مجهول ساز گاهی اوقات فعل‌های «گردیدن» و «گشتن» به جای «شدن» استفاده می‌شوند، مثلاً «گفته می‌گردد» به جای «گفته می‌شود». در موارد بسیار نادر فعل رفتن به جای شدن استفاده می‌شود، مثل «گفته رفت» به جای «گفته شد».

توجه کنید که الگوهای تصریفی مجهول فقط برای افعال متعدی کاربرد دارند. در ادامه از مصدر «گفتن» برای بیان مثال‌ها استفاده کردیم که یک فعل متعدی است. در مورد افعال لازم مانند «رفتن» صیغه‌های مجهول صرف نمی‌شوند.

برای منفی سازی در فارسی می‌توان هم از «ن» استفاده کرد و در موارد کمتر از «م»، همچنین گاهی برای تاکید در جملات مثبت از «ب» در ابتدای فعل استفاده می‌شود. با در نظر گرفتن این حالت‌ها ممکن است تعداد حالت‌های صرفی زیاد شود؛ که ما همه‌ی این حالت‌ها را در مثال‌ها و زیرالگوها لیست نکردیم و البته در الگوی جامع آنها را در نظر گرفته‌ایم. در زیرالگوها ما الگوهای متداول را لیست کرده‌ایم.

5-1 فعل‌های ساده



5-1-1 ماضی ساده

5-1-1-1 معلوم

مثال

کد	اول شخص	دوم شخص	سوم شخص	اول شخص	دوم شخص	سوم شخص
	مفرد	مفرد	مفرد	جمع	جمع	جمع
G1IP1	گفتم	گفتی	گفت	گفتیم	گفتید	گفتند
G1IP2	بگفتم	بگفتی	بگفت	بگفتیم	بگفتید	بگفتند
G1IN1	نگفتم	نگفتی	نگفت	نگفتیم	نگفتید	نگفتند

جدول 1- صرف فعل گفتن در ماضی ساده معلوم

	عنوان پروژه:		 شورای عالی اطلاع رسانی
	فاز اول طرح جامع پیکره زبان فارسی با موضوع فاز اول مطالعاتی ایجاد پیکره متنی زبان فارسی		
	عنوان زیر پروژه:		
	تاریخ: 1388/03/31	ویرایش: 1/0	کد زیر پروژه: پیکرمتن فارسی - 2 - ب

الگو

کد	عبارت منظم	توضیحات
G1IP1	[G](م ای ایم اید اند)؟	ماضی ساده معلوم مثبت یکم
G1IP2	(ب) ی ([AG] [BG]) (م ای ایم اید اند)؟	ماضی ساده معلوم مثبت دوم
G1IN1	(ن ام) ی ([AG] [BG]) (م ای ایم اید اند)؟	ماضی ساده معلوم منفی یکم
G1I	(ب ان ام) ی ([AG] [BG]) ([G] [PFGS])	ماضی ساده معلوم

جدول 2 - الگوی ماضی ساده معلوم

توضیحات



در G1IP2 و G1IN1 وقتی بن ماضی با الف شروع شود، «ی» به اول بن افزوده می شود و «آ» قلب به «ا» می شود. مثلاً از مصدر «آویخت» می گوئیم «بیاویختم».

5-1-1-2 مجهول

مثال

کد	اول شخص	دوم شخص	سوم شخص	اول شخص	دوم شخص	سوم شخص
	مفرد	مفرد	مفرد	جمع	جمع	جمع
G1TP1	گفته شدم	گفته شدی	گفته شد	گفته شدیم	گفته شدید	گفته شدند
G1TP2	بگفته شدم	بگفته شدی	بگفته شد	بگفته شدیم	بگفته شدید	بگفته شدند
G1TP3	گفته بشدم	گفته بشدی	گفته بشد	گفته بشدیم	گفته بشدید	گفته بشدند
G1TN1	گفته نشدم	گفته نشدی	گفته نشد	گفته نشدیم	گفته نشدید	گفته نشدند
G1TN2	نگفته شدم	نگفته شدی	نگفته شد	نگفته شدیم	نگفته شدید	نگفته شدند

جدول 3 - صرف فعل گفتن در ماضی ساده مجهول

	عنوان پروژه: فاز اول طرح جامع پیکره زبان فارسی با موضوع فاز اول مطالعاتی ایجاد پیکره متنی زبان فارسی		 شورای عالی اطلاع رسانی
	عنوان زیر پروژه: بررسی راهکارهای ایجاد ماشین تشخیص هوشمند افعال مرکب در زبان فارسی		
	تاریخ: 1388/03/31	ویرایش: 1/0	

الگو

توضیحات	عبارت منظم	کد
ماضی ساده مجهول مثبت یکم	شد(یم ید ندام ی)؟ [G]	G1TP1
ماضی ساده مجهول مثبت دوم	شد(یم ید ندام ی)؟ [BG][AG]	G1TP2
ماضی ساده مجهول مثبت سوم	بشد(یم ید ندام ی)؟ [G]	G1TP3
ماضی ساده مجهول منفی یکم	نشد(یم ید ندام ی)؟ [G]	G1TN1
ماضی ساده مجهول منفی دوم	نشد(یم ید ندام ی)؟ [BG][AG]	G1TN2
ماضی ساده مجهول	نشد(یم ید ندام ی)؟ [G][AG][BG] (ب ان) [PFGS][IMG]	G1T

جدول 4 - الگوی ماضی ساده مجهول

توضیحات

بین جزء مجهول ساز، «شدن»، و جزء صرفی «بن فعل» فاصله نمی‌افتد (یعنی کلمه‌ی دیگری قرار نمی‌گیرد).



5-1-2 ماضی استمراری

5-1-2-1 معلوم

مثال

کد	اول شخص	دوم شخص	سوم شخص	اول شخص	دوم شخص	سوم شخص
	مفرد	مفرد	مفرد	جمع	جمع	جمع
G2IP1	می‌گفتم	می‌گفتی	می‌گفت	می‌گفتیم	می‌گفتید	می‌گفتند
G2IN1	نمی‌گفتم	نمی‌گفتی	نمی‌گفت	نمی‌گفتیم	نمی‌گفتید	نمی‌گفتند

جدول 5 - صرف فعل گفتن در ماضی استمراری معلوم

	عنوان پروژه: فاز اول طرح جامع پیکره زبان فارسی با موضوع فاز اول مطالعاتی ایجاد پیکره متنی زبان فارسی		 شورای عالی اطلاع رسانی
	عنوان زیر پروژه: بررسی راهکارهای ایجاد ماشین تشخیص هوشمند افعال مرکب در زبان فارسی		
	کد زیر پروژه: پیکرمتن فارسی - 2 - ب	ویرایش: 1/0	

الگو

کد	عبارت منظم	توضیحات
G2IP1	می [G] (یم ید اندام ای)؟	ماضی استمراری معلوم مثبت یکم
G2IN1	نمی [G] (یم ید اندام ای)؟	ماضی استمراری معلوم مثبت یکم
G2I	(ن) می [G] [PFGS]	ماضی استمراری معلوم



جدول 6 - الگوی ماضی استمراری معلوم

توضیحات

جزء «می» به همراه «نیم فاصله» می آید. در مواقعی که مصدر با «الف» شروع می شود، تغییری در آن حاصل نمی شود. در این الگو گویش های قدیم زیر لحاظ نشده اند:

کد	اول شخص مفرد	دوم شخص مفرد	سوم شخص مفرد	اول شخص جمع	دوم شخص جمع	سوم شخص جمع
-	می بگفتم	می بگفتی	می بگفت	می بگفتیم	می بگفتید	می بگفتند
-	همی گفتم	همی گفتی	همی گفت	همی گفتیم	همی گفتید	همی گفتند
-	همی بگفتم	همی بگفتی	همی بگفت	همی بگفتیم	همی بگفتید	همی بگفتند
-	می نگفتم	می نگفتی	می نگفت	می نگفتیم	می نگفتید	می نگفتند
-	همی نگفتم	همی نگفتی	همی نگفت	همی نگفتیم	همی نگفتید	همی نگفتند
-	همی نگفتم	همی نگفتی	همی نگفت	همی نگفتیم	همی نگفتید	همی نگفتند

جدول 7 - گویش های قدیم ماضی استمراری معلوم

	عنوان پروژه: فاز اول طرح جامع پیکره زبان فارسی با موضوع فاز اول مطالعاتی ایجاد پیکره متنی زبان فارسی		 شورای عالی اطلاع رسانی
	عنوان زیر پروژه: بررسی راهکارهای ایجاد ماشین تشخیص هوشمند افعال مرکب در زبان فارسی		
	تاریخ: 1388/03/31	ویرایش: 1/0	

5-2-1-2 مجهول

مثال

کد	اول شخص	دوم شخص	سوم شخص	اول شخص	دوم شخص	سوم شخص
	مفرد	مفرد	مفرد	جمع	جمع	جمع
G2TP1	گفته می‌شدم	گفته می‌شدی	گفته می‌شد	گفته می‌شدیم	گفته می‌شدید	گفته می‌شدند
G2TP2	بگفته می‌شدم	بگفته می‌شدی	بگفته می‌شد	بگفته می‌شدیم	بگفته می‌شدید	بگفته می‌شدند
G2TN1	گفته نمی‌شدم	گفته نمی‌شدی	گفته نمی‌شد	گفته نمی‌شدیم	گفته نمی‌شدید	گفته نمی‌شدند
G2TN2	نگفته می‌شدم	نگفته می‌شدی	نگفته می‌شد	نگفته می‌شدیم	نگفته می‌شدید	نگفته می‌شدند

جدول 8 - صرف فعل گفتن در ماضی استمراری مجهول

الگو



کد	عبارت منظم	توضیحات
G2TP1	می‌گفتند (میدانند/ای)؟	ماضی استمراری مجهول مثبت یکم
G2TP2	ب؟ [G] می‌گفتند (میدانند/ای)؟	ماضی استمراری مجهول مثبت دوم
G2TN1	[G] نمی‌گفتند (میدانند/ای)؟	ماضی استمراری مجهول منفی یکم
G2TN2	ن [G] می‌گفتند (میدانند/ای)؟	ماضی استمراری مجهول منفی دوم
G2T	(ب/ن)؟ [G] می‌گفتند [PFGS][IMG]	ماضی استمراری مجهول

جدول 9 - الگوی ماضی استمراری مجهول

توضیحات

در این الگو گویش‌های قدیم زیر لحاظ نشده‌اند:

کد	اول شخص	دوم شخص	سوم شخص	اول شخص	دوم شخص	سوم شخص
	مفرد	مفرد	مفرد	جمع	جمع	جمع
-	گفته همی‌شدم	گفته همی‌شدی	گفته همی‌شد	گفته همی‌شدیم	گفته همی‌شدید	گفته همی‌شدند

	عنوان پروژه: فاز اول طرح جامع پیکره زبان فارسی با موضوع فاز اول مطالعاتی ایجاد پیکره متنی زبان فارسی		 شورای عالی اطلاع رسانی
	عنوان زیر پروژه: بررسی راهکارهای ایجاد ماشین تشخیص هوشمند افعال مرکب در زبان فارسی		
	تاریخ: 1388/03/31	ویرایش: 1/0	

همی گفته می شدند	همی گفته می شدید	همی گفته می شدیم	همی گفته می شد	همی گفته می شدی	همی گفته می شدم	-
بگفته همی شدند	بگفته همی شدید	بگفته همی شدیم	بگفته همی شد	بگفته همی شدی	بگفته همی شدم	-
گفته همی نشدند	گفته همی نشدید	گفته همی نشدیم	گفته همی نشد	گفته همی نشدی	گفته همی نشدم	-
همی گفته نمی شدند	همی گفته نمی شدید	همی گفته نمی شدیم	همی گفته نمی شد	همی گفته نمی شدی	همی گفته نمی شدم	-

جدول 10 - گویش های قدیم ماضی استمراری مجهول

3-1-5 ماضی بعید

1-3-1-5 معلوم

مثال



کد	اول شخص	دوم شخص	سوم شخص	اول شخص	دوم شخص	سوم شخص
	مفرد	مفرد	مفرد	جمع	جمع	جمع
G3IP1	گفته بودم	گفته بودی	گفته بود	گفته بودیم	گفته بودید	گفته بودند
G3IP2	بگفته بودم	بگفته بودی	بگفته بود	بگفته بودیم	بگفته بودید	بگفته بودند
G3IN1	نگفته بودم	نگفته بودی	نگفته بود	نگفته بودیم	نگفته بودید	نگفته بودند

جدول 11 - صرف فعل گفتن در ماضی بعید معلوم

الگو

کد	عبارت منظم	توضیحات
G3IP1	[G] ه بود(یم ید ند ام ای)؟	ماضی بعید معلوم مثبت یکم
G3IP2	[G] ب ه بود(یم ید ند ام ای)؟	ماضی بعید معلوم مثبت دوم
G3IN1	[G] ن ه بود(یم ید ند ام ای)؟	ماضی بعید معلوم منفی یکم
G3I	[G] ه بود(ان)؟ [PFGS]	ماضی بعید معلوم

جدول 12 - الگوی ماضی بعید معلوم

	عنوان پروژه: فاز اول طرح جامع پیکره زبان فارسی با موضوع فاز اول مطالعاتی ایجاد پیکره متنی زبان فارسی		 شورای عالی اطلاع رسانی
	عنوان زیر پروژه: بررسی راهکارهای ایجاد ماشین تشخیص هوشمند افعال مرکب در زبان فارسی		
	تاریخ: 1388/03/31	ویرایش: 1/0	

توضیحات

در ماضی بعید معلوم منفی چسباندن «ن» به ابتدای «بود»، مانند: «گفته نبودم»، غلط است. برخی الگوهای صرفی قدیمی در الگوی فوق لحاظ نشده است. بین اجزای فعل کلمه دیگری قرار نمی‌گیرد.

5-1-3-2 مجهول

مثال



کد	اول شخص	دوم شخص	سوم شخص	اول شخص	دوم شخص	سوم شخص
	مفرد	مفرد	مفرد	جمع	جمع	جمع
G3TP1	گفته شده بودم	گفته شده بودی	گفته شده بود	گفته شده بودیم	گفته شده بودید	گفته شده بودند
G3TP2	بگفته شده بودم	بگفته شده بودی	بگفته شده بود	بگفته شده بودیم	بگفته شده بودید	بگفته شده بودند
G3TN1	گفته نشده بودم	گفته نشده بودی	گفته نشده بود	گفته نشده بودیم	گفته نشده بودید	گفته نشده بودند
G3TN2	نگفته شده بودم	نگفته شده بودی	نگفته شده بود	نگفته شده بودیم	نگفته شده بودید	نگفته شده بودند

جدول 13 - صرف فعل گفتن در ماضی بعید مجهول

الگو

کد	عبارت منظم	توضیحات
G3TP1	[G]ه شده بود(یم ید ندام ی)؟	ماضی بعید مجهول مثبت یکم
G3TP2	ب[G]ه شده بود(یم ید ندام ی)؟	ماضی بعید مجهول مثبت دوم
G3TN1	[G]ه نشده بود(یم ید ندام ی)؟	ماضی بعید مجهول منفی یکم
G3TN2	ن[G]ه شده بود(یم ید ندام ی)؟	ماضی بعید مجهول منفی دوم
G3T	(ب ن ام)؟[G]ه (ب ن)؟[IMG]ه بود[PFGS]	ماضی بعید مجهول

جدول 14 - الگوی ماضی بعید مجهول

	عنوان پروژه:		 شورای عالی اطلاع رسانی
	فاز اول طرح جامع پیکره زبان فارسی با موضوع فاز اول مطالعاتی ایجاد پیکره متنی زبان فارسی		
	عنوان زیر پروژه:		
	تاریخ: 1388/03/31	ویرایش: 1/0	کد زیر پروژه: پیکرمتن فارسی - 2 - ب

توضیحات

چسباندن «ن» نفی به ابتدای «بود» غلط است.

چسباندن «ب» به ابتدای «شد» مجاز است و در الگوی کلی لحاظ شده است.

بین اجزای فعل کلمه دیگری قرار نمی‌گیرد.

5-1-4 ماضی مستمر

5-1-4-1 معلوم

مثال



کد	اول شخص	دوم شخص	سوم شخص	اول شخص	دوم شخص	سوم شخص
	مفرد	مفرد	مفرد	جمع	جمع	جمع
G4TP1	داشتم می‌گفتم	داشتی می‌گفتی	داشت می‌گفت	داشتیم می‌گفتیم	داشتید می‌گفتید	داشتند می‌گفتند
G4TN1	داشتم نمی‌گفتم	داشتی نمی‌گفتی	داشت نمی‌گفت	داشتیم نمی‌گفتیم	داشتید نمی‌گفتید	داشتند نمی‌گفتند

جدول 15 - صرف فعل گفتن در ماضی مستمر معلوم

الگو

G4TP1	داشت (یم ید اند ام ای)؟ {D} می {G} [یم ید اند ام ای]؟	ماضی مستمر معلوم مثبت یکم
G4TN1	داشت (یم ید اند ام ای)؟ {D} نمی {G} [یم ید اند ام ای]؟	ماضی مستمر معلوم منفی یکم
G4T	داشت [PFGS] {D} {ن}؟ می {G} [PFGS]	ماضی مستمر معلوم

جدول 16 - الگوی ماضی مستمر معلوم

	عنوان پروژه:		 شورای عالی اطلاع رسانی
	فاز اول طرح جامع پیکره زبان فارسی با موضوع فاز اول مطالعاتی ایجاد پیکره متنی زبان فارسی		
	عنوان زیر پروژه:		
تاریخ: 1388/03/31		ویرایش: 1/0	کد زیر پروژه: پیکرمتن فارس - 2 - ب

توضیحات

بین اجزای فعل فاصله می‌افتد؛ مانند: «داشتم سیب را می‌خوردم». در الگوی فوق D معرف این فاصله است. پسوند شخص باید در دو جزء فعل یکسان باشد.

5-1-4-2 مجهول

مثال

کد	اول شخص	دوم شخص	سوم شخص	اول شخص	دوم شخص	سوم شخص
G5IP1	داشتم گفته	داستی گفته	داشت گفته	داشتیم گفته	داشتید گفته	داشتند گفته
G5IN1	می‌شدم	می‌شدی	می‌شد	می‌شدیم	می‌شدید	می‌شدند
G5IP1	داشتم گفته	داستی گفته	داشت گفته	داشتیم گفته	داشتید گفته	داشتند گفته
G5IN1	نمی‌شدم	نمی‌شدی	نمی‌شد	نمی‌شدیم	نمی‌شدید	نمی‌شدند

جدول 17 - صرف فعل گفتن در ماضی مستمر مجهول

الگو



G5IP1	داشت (یم ید ند ام ای)؟ [G]{D}	می‌شد (یم ید ند ام ای)؟	ماضی مستمر مجهول مثبت یکم
G5IN1	داشت (یم ید ند ام ای)؟ [G]{D}	نمی‌شد (یم ید ند ام ای)؟	ماضی مستمر مجهول منفی یکم
G5I	داشت [G]{D}[PFGS] (ن)؟ می [PFGS][IMG]		ماضی مستمر مجهول

جدول 18 - الگوی ماضی مستمر مجهول

توضیحات

چسباندن «ن» نفی به ابتدای بن فعل غلط است.

بین اجزای فعل کلمات دیگری قرار می‌گیرند. در الگو این فاصله با «D» نشان داده شده است.

	عنوان پروژه:		 شورای عالی اطلاع رسانی
	فاز اول طرح جامع پیکره زبان فارسی با موضوع فاز اول مطالعاتی ایجاد پیکره متنی زبان فارسی		
	عنوان زیر پروژه:		
تاریخ: 1388/03/31	ویرایش: 1/0	کد زیر پروژه: پیکرمتن فارس - 2 - ب	

پسوند شخص باید در دو جزء فعل یکسان باشد.

5-1-5 ماضی ساده نقلی

1-5-1-5 معلوم

مثال

کد	اول شخص	دوم شخص	سوم شخص	اول شخص	دوم شخص	سوم شخص
G5IP1	گفته‌ام	گفته‌ای	گفته است	گفته‌ایم	گفته‌اید	گفته‌اند
G5IP2	بگفته‌ام	بگفته‌ای	بگفته است	بگفته‌ایم	بگفته‌اید	بگفته‌اند
G5IN1	نگفته‌ام	نگفته‌ای	نگفته است	نگفته‌ایم	نگفته‌اید	نگفته‌اند

جدول 19 - صرف فعل گفتن در ماضی ساده نقلی مجهول

الگو



G5IP1	[G]ه(آیم آید آند آم ای است)؟	ماضی ساده نقلی معلوم مثبت یکم
G5IP2	ب[G]ه(آیم آید آند آم ای است)؟	ماضی ساده نقلی معلوم مثبت دوم
G5IN1	ن[G]ه(آیم آید آند آم ای است)؟	ماضی ساده نقلی معلوم منفی یکم
G5I	ب(ان)؟[G]ه[PFGN]	ماضی ساده نقلی معلوم

جدول 20 - الگوی ماضی ساده نقلی مجهول

توضیحات

در سوم شخص مفرد «است» می‌تواند حذف شود؛ مثلاً: «گفته» به جای «گفته است».

حالت‌های صرفی قدیمی در این الگو لحاظ نشده است.

	عنوان پروژه: فاز اول طرح جامع پیکره زبان فارسی با موضوع فاز اول مطالعاتی ایجاد پیکره متنی زبان فارسی		 شورای عالی اطلاع رسانی
	عنوان زیر پروژه: بررسی راهکارهای ایجاد ماشین تشخیص هوشمند افعال مرکب در زبان فارسی		
	تاریخ: 1388/03/31	ویرایش: 1/0	

5-1-5-2 مجهول

مثال

کد	اول شخص	دوم شخص	سوم شخص	اول شخص	دوم شخص	سوم شخص
	مفرد	مفرد	مفرد	جمع	جمع	جمع
G5TP1	گفته شده‌ام	گفته شده‌ای	گفته شده است	گفته شده‌ایم	گفته شده‌اید	گفته شده‌اند
G5TP2	بگفته شده‌ام	بگفته شده‌ای	بگفته شده است	بگفته شده‌ایم	بگفته شده‌اید	بگفته شده‌اند
G5TN1	گفته نشده‌ام	گفته نشده‌ای	گفته نشده است	گفته نشده‌ایم	گفته نشده‌اید	گفته نشده‌اند
G5TN2	نگفته شده‌ام	نگفته شده‌ای	نگفته شده است	نگفته شده‌ایم	نگفته شده‌اید	نگفته شده‌اند

جدول 21 - صرف فعل گفتن در ماضی ساده نقلی مجهول

الگوی



G5TP1	[G]ه	شده (گفته‌اید گفته‌اید گفته‌اند گفته‌ام گفته‌ای است)؟	ماضی ساده نقلی مجهول مثبت یکم
G5TP2	ب [G]ه	شده (بگفته‌اید بگفته‌اید بگفته‌اند بگفته‌ام بگفته‌ای است)؟	ماضی ساده نقلی مجهول مثبت دوم
G5TN1	[G]ه	نشده (گفته‌اید گفته‌اید گفته‌اند گفته‌ام گفته‌ای است)؟	ماضی ساده نقلی مجهول منفی یکم
G5TN2	ن [G]ه	شده (نگفته‌اید نگفته‌اید نگفته‌اند نگفته‌ام نگفته‌ای است)؟	ماضی ساده نقلی مجهول منفی دوم
G5T	(ب ن)؟ [G]ه	(ب ن)؟ [IMG]ه [PFGN]	ماضی ساده نقلی مجهول

جدول 22 - الگوی ماضی ساده نقلی مجهول

توضیحات

در سوم شخص مفرد «است» می‌تواند حذف شود.

چسباندن «ب» به ابتدای «شد» مجاز است که در الگوی کلی لحاظ شده است.

	عنوان پروژه:		 شورای عالی اطلاع رسانی
	فاز اول طرح جامع پیکره زبان فارسی با موضوع فاز اول مطالعاتی ایجاد پیکره متنی زبان فارسی		
	عنوان زیر پروژه:		
تاریخ: 1388/03/31	ویرایش: 1/0	کد زیر پروژه: پیکرمتن فارسی - 2 - ب	
6-1-5 ماضی استمراری نقلی			

1-6-1-5 معلوم

مثال

کد	اول شخص	دوم شخص	سوم شخص	اول شخص	دوم شخص	سوم شخص
G6IP1	مفرد	مفرد	مفرد	جمع	جمع	جمع
G6IN1	نمی گفته‌ام	نمی گفته‌ای	نمی گفته است	نمی گفته‌ایم	نمی گفته‌اید	نمی گفته‌اند

جدول 23 - صرف فعل گفتن در ماضی استمراری نقلی معلوم

الگو

G6IP1	می‌گویم/می‌گویی/می‌گویند ام/ای است؟	ماضی استمراری نقلی معلوم مثبت یکم
G6IN1	نمی‌گویم/نمی‌گویی/نمی‌گویند ام/ای است؟	ماضی استمراری نقلی معلوم منفی یکم
G6I	می‌گویم/می‌گویی/می‌گویند (ن)؟	ماضی استمراری نقلی معلوم

جدول 24 - الگوی ماضی استمراری نقلی معلوم



توضیحات

حالت‌های صرفی قدیمی با پیشوند «همی» در این الگو لحاظ نشده است.

در سوم شخص مفرد «است» می‌تواند حذف شود.

2-6-1-5 مجهول

مثال

	عنوان پروژه:		 شورای عالی اطلاع رسانی
	فاز اول طرح جامع پیکره زبان فارسی با موضوع فاز اول مطالعاتی ایجاد پیکره متنی زبان فارسی		
	عنوان زیر پروژه:		
تاریخ: 1388/03/31		ویرایش: 1/0	کد زیر پروژه: پیکرمتن فارسی - 2 - ب

کد	اول شخص	دوم شخص	سوم شخص	اول شخص	دوم شخص	سوم شخص
	مفرد	مفرد	مفرد	جمع	جمع	جمع
G6TP1	گفته می‌شده‌ام	گفته می‌شده‌ای	گفته می‌شده‌است	گفته می‌شده‌ایم	گفته می‌شده‌اید	گفته می‌شده‌اند
G6TP2	بگفته می‌شده‌ام	بگفته می‌شده‌ای	بگفته می‌شده‌است	بگفته می‌شده‌ایم	بگفته می‌شده‌اید	بگفته می‌شده‌اند
G6TN1	گفته نمی‌شده‌ام	گفته نمی‌شده‌ای	گفته نمی‌شده‌است	گفته نمی‌شده‌ایم	گفته نمی‌شده‌اید	گفته نمی‌شده‌اند
G6TN2	نگفته می‌شده‌ام	نگفته می‌شده‌ای	نگفته می‌شده‌است	نگفته می‌شده‌ایم	نگفته می‌شده‌اید	نگفته می‌شده‌اند

جدول 25 - صرف فعل گفتن در ماضی استمراری نقلی مجهول

الگو

G6TP1	[G] ه می‌شده (میدانم میدانی است)؟	ماضی استمراری نقلی مجهول مثبت یکم
G6TP2	ب [G] ه می‌شده (میدانم میدانی است)؟	ماضی استمراری نقلی مجهول مثبت دوم
G6TN1	[G] ه نمی‌شده (میدانم میدانی است)؟	ماضی استمراری نقلی مجهول منفی یکم
G6TN2	ن [G] ه می‌شده (میدانم میدانی است)؟	ماضی استمراری نقلی مجهول منفی دوم
G6T	ب (ن)؟ [G] ه [IMG] [PFGN]	ماضی استمراری نقلی مجهول



جدول 26 - الگوی ماضی استمراری نقلی مجهول

توضیحات

حالت‌های صرفی قدیمی با پیشوند «همی» در این الگو لحاظ نشده است.

در سوم شخص مفرد «است» می‌تواند حذف شود.

بین اجزای فعل کلمه دیگری قرار نمی‌گیرد.

	عنوان پروژه: فاز اول طرح جامع پیکره زبان فارسی با موضوع فاز اول مطالعاتی ایجاد پیکره متنی زبان فارسی		 شورای عالی اطلاع رسانی
	عنوان زیر پروژه: بررسی راهکارهای ایجاد ماشین تشخیص هوشمند افعال مرکب در زبان فارسی		
	تاریخ: 1388/03/31	ویرایش: 1/0	

5-1-7 ماضی بعید نقلی

5-1-7-1 معلوم

مثال

کد	اول شخص	دوم شخص	سوم شخص	اول شخص	دوم شخص	سوم شخص
	مفرد	مفرد	مفرد	جمع	جمع	جمع
G7IP1	گفته بوده‌ام	گفته بوده‌ای	گفته بوده است	گفته بوده‌ایم	گفته بوده‌اید	گفته بوده‌اند
G7IP2	بگفته بوده‌ام	بگفته بوده‌ای	بگفته بوده است	بگفته بوده‌ایم	بگفته بوده‌اید	بگفته بوده‌اند
G7IN1	نگفته بوده‌ام	نگفته بوده‌ای	نگفته بوده است	نگفته بوده‌ایم	نگفته بوده‌اید	نگفته بوده‌اند

جدول 27 - صرف فعل گفتن در ماضی بعید نقلی معلوم

الگو

G7IP1	[G]ه بوده (آید/آیدم/آیدم ای ای است)؟	ماضی بعید نقلی معلوم مثبت یکم
G7IP2	ب [G]ه بوده (آیدم/آیدم ای ای است)؟	ماضی بعید نقلی معلوم مثبت دوم
G7IN1	ن [G]ه بوده (آیدم/آیدم ای ای است)؟	ماضی بعید نقلی معلوم منفی یکم
G7I	(ب ن)؟ [G]ه بوده [PFGN]	ماضی بعید نقلی معلوم

جدول 28 - الگوی ماضی بعید نقلی معلوم



توضیحات

اتصال «ن» به «بود» برای منفی کردن، نادرست است.

بین اجزای فعل کلمه دیگری قرار نمی‌گیرد.

5-1-7-2 مجهول

مثال

	عنوان پروژه: فاز اول طرح جامع پیکره زبان فارسی با موضوع فاز اول مطالعاتی ایجاد پیکره متنی زبان فارسی		 شورای عالی اطلاع رسانی
	عنوان زیر پروژه: بررسی راهکارهای ایجاد ماشین تشخیص هوشمند افعال مرکب در زبان فارسی		
	تاریخ: 1388/03/31	ویرایش: 1/0	

کد	اول شخص	دوم شخص	سوم شخص	اول شخص	دوم شخص	سوم شخص
G7TP1	مفرد	مفرد	مفرد	جمع	جمع	جمع
G7TP2	مفرد	مفرد	مفرد	جمع	جمع	جمع
G7TP3	مفرد	مفرد	مفرد	جمع	جمع	جمع
G7TN1	مفرد	مفرد	مفرد	جمع	جمع	جمع
G7TN2	مفرد	مفرد	مفرد	جمع	جمع	جمع



جدول 29 - صرف فعل گفتن در ماضی بعید نقلی مجهول

الگو

G7TP1	[G]ه شده بوده (آیم آید آند آم آی است)؟	ماضی بعید نقلی مجهول مثبت یکم
G7TP2	[G]ب شده بوده (آیم آید آند آم آی است)؟	ماضی بعید نقلی مجهول مثبت دوم
G7TP3	[G]ه بشده بوده (آیم آید آند آم آی است)؟	ماضی بعید نقلی مجهول مثبت سوم
G7TN1	[G]ه نشده بوده (آیم آید آند آم آی است)؟	ماضی بعید نقلی مجهول منفی یکم
G7TN2	[G]ن شده بوده (آیم آید آند آم آی است)؟	ماضی بعید نقلی مجهول منفی دوم
G7T	[G]ه (ب ن)؟ [G]ه (ب ن)؟ [IMG]ه بوده [PFGN]	ماضی بعید نقلی مجهول

جدول 30 - الگوی ماضی بعید نقلی مجهول

توضیحات

	عنوان پروژه:		 شورای عالی اطلاع رسانی
	فاز اول طرح جامع پیکره زبان فارسی با موضوع فاز اول مطالعاتی ایجاد پیکره متنی زبان فارسی		
	عنوان زیر پروژه:		
	تاریخ: 1388/03/31	ویرایش: 1/0	کد زیر پروژه: پیکرمتن فارس - 2 - ب

بین اجزای فعل کلمه دیگری قرار نمی‌گیرد.
چسباندن «ن» به ابتدای «بود» برای منفی سازی غلط است.
در سوم شخص مفرد «است» می‌تواند حذف شود.

5-1-8 ماضی مستمر نقلی

5-1-8-1 معلوم

مثال



کد	اول شخص	دوم شخص	سوم شخص	اول شخص	دوم شخص	سوم شخص
	مفرد	مفرد	مفرد	جمع	جمع	جمع
G8IP1	داشته‌ام	داشته‌ای	داشته‌است	داشته‌ایم	داشته‌اید	داشته‌اند
	می‌گفتم	می‌گفته‌ای	می‌گفته‌است	می‌گفته‌ایم	می‌گفته‌اید	می‌گفته‌اند
G8IN1	داشته‌ام	داشته‌ای	داشته‌است	داشته‌ایم	داشته‌اید	داشته‌اند
	نمی‌گفتم	نمی‌گفته‌ای	نمی‌گفته‌است	نمی‌گفته‌ایم	نمی‌گفته‌اید	نمی‌گفته‌اند

جدول 31 - صرف فعل گفتن در ماضی مستمر نقلی معلوم

الگو

ماضی مستمر نقلی معلوم مثبت یکم	داشته (ایم اید اند ام ای است)؟ {D} می‌{G}ه [ایم اید اند ام ای است]؟	G8IP1
ماضی مستمر نقلی معلوم منفی یکم	داشته (ایم اید اند ام ای است)؟ {D} نمی‌{G}ه [ایم اید اند ام ای است]؟	G8IN1
ماضی مستمر نقلی معلوم	داشته {D} (ن) می‌{G}ه [PFGN]	G8I

جدول 32 - الگوی ماضی مستمر نقلی معلوم

	عنوان پروژه:		 شورای عالی اطلاع رسانی
	فاز اول طرح جامع پیکره زبان فارسی با موضوع فاز اول مطالعاتی ایجاد پیکره متنی زبان فارسی		
	عنوان زیر پروژه:		
	تاریخ: 1388/03/31	ویرایش: 1/0	کد زیر پروژه: پیکرمتن فارسی - 2 - ب

توضیحات

بین اجزای فعل کلمات دیگری قرار می‌گیرد.
 در سوم شخص مفرد «است» می‌تواند حذف شود.
 پسوند ضمیر در دو جزء فعل باید یکسان باشد. برای سادگی الگو را بصورت فوق نوشتیم.
 این زمان خیلی کم مورد استفاده قرار می‌گیرد.

5-1-9 ماضی التزامی

5-1-9-1 معلوم



مثال

کد	اول شخص	دوم شخص	سوم شخص	اول شخص	دوم شخص	سوم شخص
	مفرد	مفرد	مفرد	جمع	جمع	جمع
G9IP1	گفته باشم	گفته باشی	گفته باشد	گفته باشیم	گفته باشید	گفته باشند
G9IP2	بگفته باشم	بگفته باشی	بگفته باشد	بگفته باشیم	بگفته باشید	بگفته باشند
G9IN1	نگفته باشم	نگفته باشی	نگفته باشد	نگفته باشیم	نگفته باشید	نگفته باشند
G9IN2	گفته نباشم	گفته نباشی	گفته نباشد	گفته نباشیم	گفته نباشید	گفته نباشند

جدول 35 - صرف فعل گفتن در ماضی التزامی معلوم

الگو

ماضی التزامی معلوم مثبت یکم	G9IP1 [G]ه باش (یم ید ندام ی)؟
ماضی التزامی معلوم مثبت دوم	G9IP2 [G]ب باش (یم ید ندام ی)؟
ماضی التزامی معلوم منفی یکم	G9IN1 [G]ن باش (یم ید ندام ی)؟

	عنوان پروژه:		 شورای عالی اطلاع رسانی
	فاز اول طرح جامع پیکره زبان فارسی با موضوع فاز اول مطالعاتی ایجاد پیکره متنی زبان فارسی		
	عنوان زیر پروژه:		
	تاریخ: 1388/03/31	ویرایش: 1/0	کد زیر پروژه: پیکرمتن فارس - 2 - ب

ماضی التزامی معلوم منفی دوم	G9IN2 [G] نباش (یم ید ام ای)؟
ماضی التزامی معلوم	G9I [G] باش (ن)؟ [PFGS]

جدول 36 - الگوی ماضی التزامی معلوم

توضیحات

بین اجزای فعل کلمه دیگری قرار نمی‌گیرد.

5-1-9-2 مجهول



مثال

کد	اول شخص	دوم شخص	سوم شخص	اول شخص	دوم شخص	سوم شخص
	مفرد	مفرد	مفرد	جمع	جمع	جمع
G9TP1	گفته شده باشم	گفته شده باشی	گفته شده باشد	گفته شده باشیم	گفته شده باشید	گفته شده باشند
G9TP2	بگفته شده باشم	بگفته شده باشی	بگفته شده باشد	بگفته شده باشیم	بگفته شده باشید	بگفته شده باشند
G9TP3	گفته بشده باشم	گفته بشده باشی	گفته بشده باشد	گفته بشده باشیم	گفته بشده باشید	گفته بشده باشند
G9TN1	نگفته شده باشم	نگفته شده باشی	نگفته شده باشد	نگفته شده باشیم	نگفته شده باشید	نگفته شده باشند
G9TN2	گفته نشده باشم	گفته نشده باشی	گفته نشده باشد	گفته نشده باشیم	گفته نشده باشید	گفته نشده باشند

جدول 37 - صرف فعل گفتن در ماضی التزامی مجهول

الگو

ماضی التزامی مجهول مثبت یکم	G9TP1 [G] شده باش (یم ید ام ای)؟
ماضی التزامی مجهول مثبت دوم	G9TP2 [G] ب شده باش (یم ید ام ای)؟
ماضی التزامی مجهول مثبت سوم	G9TP3 [G] ه شده باش (یم ید ام ای)؟

	عنوان پروژه:		 شورای عالی اطلاع رسانی
	فاز اول طرح جامع پیکره زبان فارسی با موضوع فاز اول مطالعاتی ایجاد پیکره متنی زبان فارسی		
	عنوان زیر پروژه:		
	تاریخ: 1388/03/31	ویرایش: 1/0	کد زیر پروژه: پیکرمتن فارس - 2 - ب

ماضی التزامی مجهول منفی یکم	G9TN1	ن [G] ه شده باش (یم ید ند ام ای)؟
ماضی التزامی مجهول منفی دوم	G9TN2	[G] ه نشده باش (یم ید ند ام ای)؟
ماضی التزامی مجهول	G9T	(ب ان ام)؟ [G] ه (ب ان ام)؟ [IMG] ه باش [PFGS]

جدول 38 - الگوی ماضی التزامی مجهول

توضیحات

بین اجزای فعل کلمه دیگری قرار نمی گیرد.

«م» می تواند به جای «ن» برای منفی کردن استفاده شود که در الگوی کلی لحاظ شده است.

5-1-10 مضارع اخباری

5-1-10-1 معلوم



مثال

کد	اول شخص	دوم شخص	سوم شخص	اول شخص	دوم شخص	سوم شخص
	مفرد	مفرد	مفرد	جمع	جمع	جمع
H1IP1	می گویم	می گویی	می گوید	می گوئیم	می گوید	می گویند
H1IN1	نمی گویم	نمی گویی	نمی گوید	نمی گوئیم	نمی گوید	نمی گویند

جدول 39 - صرف فعل گفتن در مضارع اخباری معلوم

الگو

مضارع اخباری معلوم مثبت یکم	H1IP1	می [HA] ی [HB] ((یم ید ند ام ای اد))
مضارع اخباری معلوم منفی یکم	H1IN1	نمی [HA] ی [HB] ((یم ید ند ام ای اد))
مضارع اخباری معلوم	H1I	می [HA] ی [HB] [PFH] ((ن)؟)

	عنوان پروژه:		 شورای عالی اطلاع رسانی
	فاز اول طرح جامع پیکره زبان فارسی با موضوع فاز اول مطالعاتی ایجاد پیکره متنی زبان فارسی		
	عنوان زیر پروژه:		
	تاریخ: 1388/03/31	ویرایش: 1/0	کد زیر پروژه: پیکرمتن فارس - 2 - ب

جدول 40 - الگوی مضارع اخباری معلوم

توضیحات

بن‌هایی که به الف ختم می‌شوند، به انتهای آنها «ی» افزوده می‌گردد. الگوهای صرف قدیمی با «همی» در این الگو لحاظ نشده است.

5-1-10-2 مجهول

مثال

کد	اول شخص	دوم شخص	سوم شخص	اول شخص	دوم شخص	سوم شخص
	مفرد	مفرد	مفرد	جمع	جمع	جمع
H1TP1	گفته می‌شوم	گفته می‌شوی	گفته می‌شود	گفته می‌شویم	گفته می‌شوید	گفته می‌شوند
H1TN1	گفته نمی‌شوم	گفته نمی‌شوی	گفته نمی‌شود	گفته نمی‌شویم	گفته نمی‌شوید	گفته نمی‌شوند

جدول 41 - صرف فعل گفتن در مضارع اخباری مجهول



الگو

مضارع اخباری مجهول مثبت یکم	H1TP1 [G] ه می‌گو(یم ید ندام د)
مضارع اخباری مجهول منفی یکم	H1TN1 [G] ه نمی‌گو(یم ید ندام د)
مضارع اخباری مجهول	H1T [G] ه (ن)؟ می‌گو [IMH][PFH]

جدول 42 - الگوی مضارع اخباری مجهول

توضیحات

بین اجزای فعل فاصله نمی‌افتد.

	عنوان پروژه: فاز اول طرح جامع پیکره زبان فارسی با موضوع فاز اول مطالعاتی ایجاد پیکره متنی زبان فارسی		 شورای عالی اطلاع رسانی
	عنوان زیر پروژه: بررسی راهکارهای ایجاد ماشین تشخیص هوشمند افعال مرکب در زبان فارسی		
	تاریخ: 1388/03/31	ویرایش: 1/0	

5-1-11 مضارع مستمر

5-1-11-1 معلوم

مثال

کد	اول شخص	دوم شخص	سوم شخص	اول شخص	دوم شخص	سوم شخص
	مفرد	مفرد	مفرد	جمع	جمع	جمع
H2IP1	دارم می گویم	داری می گویی	دارد می گوید	داریم می گوئیم	دارید می گوئید	دارند می گویند
H2IN1	دارم نمی گویم	داری نمی گویی	دارد نمی گوید	داریم نمی گوئیم	دارید نمی گوئید	دارند نمی گویند

جدول 43 - صرف فعل گفتن در مضارع مستمر معلوم

الگو

H2IP1	دار (یم ید ند ام ای اد) {D} می {HA} ای {HB} (یم ید ند ام ای اد)	مضارع مستمر معلوم مثبت یکم
H2IN1	دار (یم ید ند ام ای اد) {D} نمی {HA} ای {HB} (یم ید ند ام ای اد)	مضارع مستمر معلوم منفی یکم
H2I	دار {D} {PFH} ای {HA} ای {HB} {PFH}	مضارع مستمر معلوم

جدول 44 - الگوی مضارع مستمر معلوم



توضیحات

پسوند شخص در دو جزء فعل باید یکسان باشد. برای سادگی الگو آنرا به صورت فوق نشان دادیم.

5-1-11-2 مجهول

مثال

کد	اول شخص	دوم شخص	سوم شخص	اول شخص	دوم شخص	سوم شخص
	مفرد	مفرد	مفرد	جمع	جمع	جمع

	عنوان پروژه:		 شورای عالی اطلاع رسانی
	فاز اول طرح جامع پیکره زبان فارسی با موضوع فاز اول مطالعاتی ایجاد پیکره متنی زبان فارسی		
	عنوان زیر پروژه:		
تاریخ: 1388/03/31	ویرایش: 1/0	کد زیر پروژه: پیکرمتن فارس - 2 - ب	

دارم گفته	داری گفته	دارد گفته	داریم گفته	دارید گفته	دارند گفته	H2TP1
می شوم	می شوی	می شود	می شویم	می شوید	می شوند	
دارم گفته	داری گفته	دارد گفته	داریم گفته	دارید گفته	دارند گفته	H2TN1
نمی شوم	نمی شوی	نمی شود	نمی شویم	نمی شوید	نمی شوند	

جدول 45 - صرف فعل گفتن در مضارع مستمر مجهول

الگو

مضارع مستمر مجهول مثبت یکم	دار(یم ید ند ام ای د) [G]{D} ه می (شو یم ید ند ام ای د)	H2TP1
مضارع مستمر مجهول منفی یکم	دار(یم ید ند ام ای د) [G]{D} ه نمی (شو یم ید ند ام ای د)	H2TN1
مضارع مستمر مجهول	دار [G]{D} [PFH] [IMH] (ن) می؟ [PFH] [IMH]	H2T

جدول 46 - الگوی ماضی ساده مجهول

توضیحات



بین اجزای فعل فاصله می افتد که توسط «D» مشخص شده است. پسوند شخص در دو جزء فعل باید یکسان باشد. برای سادگی الگو آنرا به صورت فوق نشان دادیم.

5-1-12 مضارع التزامی

5-1-12-1 معلوم

مثال

کد	اول شخص	دوم شخص	سوم شخص	اول شخص	دوم شخص	سوم شخص
	مفرد	مفرد	مفرد	جمع	جمع	جمع

	عنوان پروژه:		 شورای عالی اطلاع رسانی
	فاز اول طرح جامع پیکره زبان فارسی با موضوع فاز اول مطالعاتی ایجاد پیکره متنی زبان فارسی		
	عنوان زیر پروژه:		
تاریخ: 1388/03/31	ویرایش: 1/0	کد زیر پروژه: پیکرمتن فارسی - 2 - ب	

H3IP1	بگویم	بگویی	بگوید	بگوئیم	بگوئید	بگویند
H3IP2	گویم	گویی	گوید	گوئیم	گوئید	گویند
H3IN1	نگویم	نگویی	نگوید	نگوئیم	نگوئید	نگویند

جدول 47 - صرف فعل گفتن در مضارع التزامی معلوم

الگو

H3IP1	ب(ی) [AHA] ای ای [AHB] [BHA] ای [BHB] (یم اید اندام ای اد)	مضارع التزامی معلوم مثبت یکم
H3IP2	([BHA] ای [BHB] (یم اید اندام ای اد)	مضارع التزامی معلوم مثبت دوم
H3IN1	ن(ی) [AHA] ای ای [AHB] [BHA] ای [BHB] (یم اید اندام ای اد)	مضارع التزامی معلوم منفی یکم
H3I	(ب ان ام)؟ (ی) [AHA] ای ای [AHB] [BHA] ای [BHB] [PFH]	مضارع التزامی معلوم

جدول 48 - الگوی مضارع التزامی معلوم

توضیحات



«م» برای منفی سازی فعل به جای «ن» استفاده می شود که در الگوی کلی لحاظ شده است. در بنهایی که با الف شروع می شوند، در صورت افزودن «ب» به ابتدا، «ی» به ابتدای آنها اضافه می شود؛ مانند: بن «آزار» که می شود «بیآزار». در بنهایی که به الف ختم می شوند، «ی» به انتهای آنها افزوده می شود؛ مانند: بن «زدا» که می شود «بزدایم». افعالی که با الف شروع و خاتمه می یابند، در هر دو طرف «ی» می گیرند؛ مانند: بن «آسا» که می شود «بیاساید».

در بنهایی که با «آ» شروع می شوند وقتی «ب» افزوده می شود، «آ» به «ا» تبدیل می شود.

5-12-1-2 مجهول

مثال

کد	اول شخص	دوم شخص	سوم شخص	اول شخص	دوم شخص	سوم شخص
----	---------	---------	---------	---------	---------	---------

	عنوان پروژه:		 شورای عالی اطلاع رسانی
	فاز اول طرح جامع پیکره زبان فارسی با موضوع فاز اول مطالعاتی ایجاد پیکره متنی زبان فارسی		
	عنوان زیر پروژه:		
تاریخ: 1388/03/31	ویرایش: 1/0	کد زیر پروژه: پیک-متن-فارس - 2 - ب	
بررسی راهکارهای ایجاد ماشین تشخیص هوشمند افعال مرکب در زبان فارسی			

مفرد	مفرد	مفرد	جمع	جمع	جمع
H3TP1	گفته شوم	گفته شوی	گفته شود	گفته شویم	گفته شوند
H3TP2	گفته بشوم	گفته بشوی	گفته بشود	گفته بشویم	گفته بشوند
H3TP3	بگفته شوم	بگفته شوی	بگفته شود	بگفته شویم	بگفته شوند
H3TN1	گفته نشوم	گفته نشوی	گفته نشود	گفته نشویم	گفته نشوند
H3TN2	نگفته شوم	نگفته شوی	نگفته شود	نگفته شویم	نگفته شوند

جدول 49 - صرف فعل گفتن در مضارع التزامی مجهول



الگو

H3TP1	[G]ه شو(یم ید ند ام ای اد)	مضارع التزامی مجهول مثبت یکم
H3TP2	[G]ه بشو(یم ید ند ام ای اد)	مضارع التزامی مجهول مثبت دوم
H3TP3	[G]ب شو(یم ید ند ام ای اد)	مضارع التزامی مجهول مثبت سوم
H3TN1	[G]ه نشو(یم ید ند ام ای اد)	مضارع التزامی مجهول منفی یکم
H3TN2	[G]ن شو(یم ید ند ام ای اد)	مضارع التزامی مجهول منفی دوم
H3T	[G]ه (ب ان ام)؟ [IMH][PFH]	مضارع التزامی مجهول

جدول 50 - الگوی مضارع التزامی مجهول

توضیحات

بین اجزای فعل فاصله نمی‌افتد.

	عنوان پروژه:		 شورای عالی اطلاع رسانی
	عنوان زیر پروژه:		
	تاریخ: 1388/03/31	ویرایش: 1/0	
فاز اول طرح جامع پیکره زبان فارسی با موضوع فاز اول مطالعاتی ایجاد پیکره متنی زبان فارسی			
بررسی راهکارهای ایجاد ماشین تشخیص هوشمند افعال مرکب در زبان فارسی			

5-1-13 آینده

5-1-13-1 معلوم

مثال

کد	اول شخص	دوم شخص	سوم شخص	اول شخص	دوم شخص	سوم شخص
	مفرد	مفرد	مفرد	جمع	جمع	جمع
A1IP1	خواهم گفت	خواهی گفت	خواهد گفت	خواهیم گفت	خواهید گفت	خواهند گفت
A1IP2	خواهم بگفتم	خواهی بگفتم	خواهد بگفتم	خواهیم بگفتم	خواهید بگفتم	خواهند بگفتم
A1IP3	بخوام گفت	بخوای گفت	بخواید گفت	بخوایم گفت	بخواید گفت	بخوایند گفت
A1IN1	نخواهم گفت	نخواهی گفت	نخواهد گفت	نخواهیم گفت	نخواهید گفت	نخواهند گفت

جدول 51 - صرف فعل گفتن در آینده معلوم



الگو

A1IP1	خواه (یم ید ند ام ای د) [G]	آینده معلوم مثبت یکم
A1IP2	خواه (یم ید ند ام ای د) ب [G]	آینده معلوم مثبت دوم
A1IP3	بخواه (یم ید ند ام ای د) [G]	آینده معلوم مثبت سوم
A1IN1	نخواه (یم ید ند ام ای د) [G]	آینده معلوم منفی یکم
A1I	(ب ن ام)؟خواه [PFH] (ب ن ام)؟ [G]	آینده معلوم

جدول 52 - الگوی آینده معلوم

توضیحات

بین اجزای فعل فاصله نمی‌افتد. در صورت ایجاد فاصله، «خواستن» فعل کمکی حساب می‌شود.

	عنوان پروژه: فاز اول طرح جامع پیکره زبان فارسی با موضوع فاز اول مطالعاتی ایجاد پیکره متنی زبان فارسی		 شورای عالی اطلاع رسانی
	عنوان زیر پروژه: بررسی راهکارهای ایجاد ماشین تشخیص هوشمند افعال مرکب در زبان فارسی		
	تاریخ: 1388/03/31	ویرایش: 1/0	

5-13-1-2 مجهول

مثال

کد	اول شخص	دوم شخص	سوم شخص	اول شخص	دوم شخص	سوم شخص
	مفرد	مفرد	مفرد	جمع	جمع	جمع
A1TP1	گفته خواهم شد	گفته خواهی شد	گفته خواهد شد	گفته خواهیم شد	گفته خواهید شد	گفته خواهند شد
A1TP2	بگفته خواهم شد	بگفته خواهی شد	بگفته خواهد شد	بگفته خواهیم شد	بگفته خواهید شد	بگفته خواهند شد
A1TP3	گفته بخوام شد	گفته بخوای شد	گفته بخواد شد	گفته بخوایم شد	گفته بخواید شد	گفته خواهند شد
A1TN1	گفته نخواهم شد	گفته نخواهی شد	گفته نخواهد شد	گفته نخواهیم شد	گفته نخواهید شد	گفته نخواهند شد
A1TN2	نگفته خواهم شد	نگفته خواهی شد	نگفته خواهد شد	نگفته خواهیم شد	نگفته خواهید شد	نگفته خواهند شد

جدول 53 - صرف فعل گفتن در آینده مجهول



الگو

آینده مجهول مثبت یکم	A1TP1	[G] خواه (یم ید ندام ی د) شد
آینده مجهول مثبت دوم	A1TP2	ب[G] خواه (یم ید ندام ی د) شد
آینده مجهول مثبت سوم	A1TP3	[G] خواه (یم ید ندام ی د) شد
آینده مجهول منفی یکم	A1TN1	[G] نخواه (یم ید ندام ی د) شد
آینده مجهول منفی دوم	A1TN2	ن[G] خواه (یم ید ندام ی د) شد
آینده مجهول	A1T	(ب ان ام)؟[G] خواه (ب ان ام)؟[PFH] [IMG] آینده مجهول

جدول 54 - الگوی آینده مجهول

توضیحات

اتصال «ن» به ابتدای «شد» برای منفی سازی اشتباه است.
از «م» به جای «ن» می‌توان برای منفی کردن استفاده کرد.

	عنوان پروژه:		 شورای عالی اطلاع رسانی
	فاز اول طرح جامع پیکره زبان فارسی با موضوع فاز اول مطالعاتی ایجاد پیکره متنی زبان فارسی		
	عنوان زیر پروژه:		
تاریخ: 1388/03/31	ویرایش: 1/0	کد زیر پروژه: پیکرمتن فارس - 2 - ب	
بررسی راهکارهای ایجاد ماشین تشخیص هوشمند افعال مرکب در زبان فارسی			

بین اجزای فعل فاصله نمی‌افتد.

5-1-14 امر

5-1-14-1 معلوم

مثال

کد	اول شخص	دوم شخص	سوم شخص	اول شخص	دوم شخص	سوم شخص
	مفرد	مفرد	مفرد	جمع	جمع	جمع
A2IP1	-	بگو(ی)	-	-	بگوید	-
A2IN1	-	نگو(ی)	-	-	نگوید	-
A2IN1	-	مگو(ی)	-	-	مگوید	-

جدول 55 - صرف فعل گفتن امر معلوم



الگو

امر معلوم مثبت یکم	ب(ی)[AHA]؟ای[AHB][BHA]؟ی[BHB]	A2IP1
امر معلوم منفی یکم	([BHB]ی[AHB])	A2IN1
امر معلوم منفی یکم	(ن(ام)ی[AHA]؟ای[AHB][BHA]؟ی[BHB])	A2IN1
امر معلوم	(ب(ان(ام)؟ی[AHA]؟ای[AHB][BHA]؟ی[BHB])	A2I

جدول 56 - الگوی امر معلوم

توضیحات

در بن‌هایی که با الف شروع می‌شوند هنگام اتصال «ب، ن، م» به ابتدای آن، «ی» به اول بن افزوده می‌شود و «آ» به «ا» تبدیل می‌شود؛ مانند: «بیاویز» از بن «آویز».

	عنوان پروژه: فاز اول طرح جامع پیکره زبان فارسی با موضوع فاز اول مطالعاتی ایجاد پیکره متنی زبان فارسی		 شورای عالی اطلاع رسانی
	عنوان زیر پروژه: بررسی راهکارهای ایجاد ماشین تشخیص هوشمند افعال مرکب در زبان فارسی		
	تاریخ: 1388/03/31	ویرایش: 1/0	

در بن‌هایی که به الف ختم می‌شود، می‌توان «ی» به انتهای آن اضافه کرد (اختیاری است)؛ مانند: «بزدای» از بن «زدا». در بن‌هایی که به «ی» ختم می‌شوند می‌توان «ی» را حذف کرد (اختیاری است). مانند: «بگوی» و «بگو» از بن «گوی».

5-1-14-2 مجهول

مثال

کد	اول شخص	دوم شخص	سوم شخص	اول شخص	دوم شخص	سوم شخص
	مفرد	مفرد	مفرد	جمع	جمع	جمع
A2TP1	-	گفته بشو	-	-	گفته بشوید	-
A2TP2	-	گفته شو	-	-	گفته شوید	-
A2TN1	-	گفته نشو	-	-	گفته نشوید	-

جدول 57 - صرف فعل گفتن در امر مجهول



الگو

A2TP1	[G] ه [بشو(ید)]؟	امر مجهول مثبت یکم
A2TP2	[G] ه [شو(ید)]؟	امر مجهول مثبت دوم
A2TN1	[G] ه [نشو(ید)]؟	امر مجهول منفی یکم
A2T	[G] ه [ب(ان)ام(شو)ید)]؟	امر مجهول

جدول 58 - الگوی امر مجهول

توضیحات



می‌توان از «م» برای نهی استفاده کرد.

	عنوان پروژه: فاز اول طرح جامع پیکره زبان فارسی با موضوع فاز اول مطالعاتی ایجاد پیکره متنی زبان فارسی		 شورای عالی اطلاع رسانی
	عنوان زیر پروژه: بررسی راهکارهای ایجاد ماشین تشخیص هوشمند افعال مرکب در زبان فارسی		
	تاریخ: 1388/03/31	ویرایش: 1/0	

5-1-15 خلاصه الگوهای افعال ساده

در جدول زیر تمام الگوهای مربوط با افعال ساده را مشاهده می کنید.

ماضی ساده معلوم	G1I (ب ان ام)ی [AG][BG][G] [PFGS][Z]؟
ماضی ساده مجهول	G1T [G] (ب ان ام)ی [AG][BG] ه [PFGS][IMG] (ب ان)
ماضی استمراری معلوم	G2I (ن)می [G] [PFGS][Z]؟
ماضی استمراری مجهول	G2T (ب ان)ی [G] ه (ن)می [G] [PFGS][IMG]؟
ماضی بعید معلوم	G3I (ب ان)ی [G] ه بود [PFGS][Z]؟
ماضی بعید مجهول	G3T (ب ان ام)ی [G] ه (ب ان)ی [G] [PFGS] بود
ماضی مستمر معلوم	G4T داشت [PFGS][Z] [D] { (ن)می [G] [PFGS][Z]؟
ماضی مستمر مجهول	G5I داشت [PFGS][D] [G] ه (ن)می [G] شد [PFGS]
ماضی ساده نقلی معلوم	G5I (ب ان)ی [G] ه [PFGN][Z]؟
ماضی ساده نقلی مجهول	G5T (ب ان)ی [G] ه (ب ان)ی [G] [PFGN] ه [IMG]
ماضی استمراری نقلی معلوم	G6I (ن)می [G] [PFGN] ه [Z]؟
ماضی استمراری نقلی مجهول	G6T (ب ان)ی [G] ه (ن)می [G] [PFGN] ه [IMG]
ماضی بعید نقلی معلوم	G7I (ب ان)ی [G] ه بوده [PFGN][Z]؟
ماضی بعید نقلی مجهول	G7T (ب ان)ی [G] ه (ب ان)ی [G] [PFGN] ه [IMG] بوده
ماضی مستمر نقلی معلوم	G8I داشته [PFGN][Z] [D] { (ن)می [G] [PFGN][Z]؟
ماضی مستمر نقلی مجهول	G8T داشته [PFGN][D] [G] ه (ن)می [G] [PFGN] ه [IMG]
ماضی التزامی معلوم	G9I (ب ان)ی [G] ه (ن)باش [PFGS][Z]؟
ماضی التزامی مجهول	G9T (ب ان ام)ی [G] ه (ب ان ام)ی [G] [PFGS] ه [IMG] باش

	عنوان پروژه:		 شورای عالی اطلاع رسانی
	فاز اول طرح جامع پیکره زبان فارسی با موضوع فاز اول مطالعاتی ایجاد پیکره متنی زبان فارسی		
	عنوان زیر پروژه:		
	تاریخ: 1388/03/31	ویرایش: 1/0	کد زیر پروژه: پیکرمتن فارسی - 2 - ب



مضارع اخباری معلوم	H1I	(ن)می (HA)ی (HB)ی (Z)PFH
مضارع اخباری مجهول	H1T	[G] ه (ن)می (PFH)IMH
مضارع مستمر معلوم	H2I	دار (Z)PFH (D)ن می (HA)ی (HB)ی (Z)PFH
مضارع مستمر مجهول	H2T	دار (D)PFH (G) ه (ن)می (PFH)IMH
مضارع التزامی معلوم	H3I	(ب ان ام)ی (AHA)ی (AHB)ی (BHA)ی (BHB)ی (Z)PFH
مضارع التزامی مجهول	H3T	(ب ان ام)ی (G) ه (ب ان ام)ی (PFH)IMH
آینده معلوم	A1I	(ب ان ام) خواهد (PFH) (G)ی
آینده مجهول	A1T	(ب ان ام) خواهد (PFH) (G)ی (IMG)
امر معلوم	A2I	(ب ان ام)ی (AHA)ی (AHB)ی (BHA)ی (BHB)ی
امر مجهول	A2T	(ب ان ام) شو (ید)ی (G) ه

جدول 59 - الگوی افعال ساده



5-2 فعل های پیشوندی

در اینجا الگوی مربوط به افعال پیشوندی به طور خلاصه آورده شده‌اند. این افعال شامل یک بخش ثابت پیشوندی هستند که به «VPF» نشان داده شده است. بخش صرفی مانند افعال ثابت است.

ماضی ساده معلوم	G1I	(ب ان ام)ی (AG)ی (BG)ی (G)ی (Z)PFGS (Z)VPF
ماضی ساده مجهول	G1T	(ب ان) (IMG)PFGS (G)ی (ب ان) (AG)ی (BG)ی (Z)VPF
ماضی استمراری معلوم	G2I	(ب ان)می (Z)PFGS (G)ی (Z)VPF
ماضی استمراری مجهول	G2T	(ب ان)می (Z)PFGS (G)ی (ب ان) (Z)VPF (IMG)PFGS

	عنوان پروژه: فاز اول طرح جامع پیکره زبان فارسی با موضوع فاز اول مطالعاتی ایجاد پیکره متنی زبان فارسی		 شورای عالی اطلاع رسانی
	عنوان زیر پروژه: بررسی راهکارهای ایجاد ماشین تشخیص هوشمند افعال مرکب در زبان فارسی		
	تاریخ: 1388/03/31	ویرایش: 1/0	

ماضی بعید معلوم	؟[Z][VFP]؟(ب ان)؟[G]ه بود[PFGS][Z]	G3I
ماضی بعید مجهول	[VFP]؟(ب ان)؟[G]ه بود[IMG]؟[PFGS]	G3T
ماضی مستمر معلوم	داشت[PFGS][Z]؟{D}؟[Z][VFP]؟(ن)؟می؟[G]؟[PFGS][Z]	G4I
ماضی مستمر مجهول	داشت[PFGS]؟{D}؟[VFP]؟[G]ه (ن)؟می؟[IMG]؟[PFGS]	G4T
ماضی ساده نقلی معلوم	؟[Z][VFP]؟(ب ان)؟[G]ه[PFGN][Z]	G5I
ماضی ساده نقلی مجهول	[VFP]؟(ب ان)؟[G]ه (ب ان)؟[IMG]؟[PFGN]	G5T
ماضی استمراری نقلی معلوم	؟[Z][VFP]؟(ن)؟می؟[G]؟[PFGN][Z]	G6I
ماضی استمراری نقلی مجهول	[VFP]؟(ب ان)؟[G]ه (ن)؟می؟[IMG]؟[PFGN]	G6T
ماضی بعید نقلی معلوم	؟[Z][VFP]؟(ب ان)؟[G]ه بوده[PFGN][Z]	G7I
ماضی بعید نقلی مجهول	[VFP]؟(ب ان)؟[G]ه (ب ان)؟[IMG]؟[PFGN] بوده	G7T
ماضی مستمر نقلی معلوم	داشته[PFGN][Z]؟{D}؟[Z][VFP]؟(ن)؟می؟[G]؟[PFGN][Z]	G8I
ماضی مستمر نقلی مجهول	داشته[PFGN]؟{D}؟[VFP]؟[G]ه (ن)؟می؟[IMG]؟[PFGN]	G8T
ماضی التزامی معلوم	؟[Z][VFP]؟(ب ان)؟[G]ه (ن)؟باش[PFGS][Z]	G9I
ماضی التزامی مجهول	[VFP]؟(ب ان)؟[G]ه (ب ان)؟[IMG]؟[PFGS] باش	G9T
مضارع اخباری معلوم	؟[Z][VFP]؟(ن)؟می؟[HA]ی ؟[HB]؟[PFH][Z]	H1I
مضارع اخباری مجهول	[VFP]؟[G]ه (ن)؟می؟[IMH]؟[PFH]	H1T
مضارع مستمر معلوم	دار[PFGN][Z]؟{D}؟[Z][VFP]؟(ن)؟می؟[HA]ی ؟[HB]؟[PFH][Z]	H2I
مضارع مستمر مجهول	دار[PFGN]؟{D}؟[VFP]؟[G]ه (ن)؟می؟[IMH]؟[PFH]	H2T
مضارع التزامی معلوم	؟[Z][VFP]؟(ب ان)؟[G]ی ؟[AHA]ی ؟[AHB] ؟[BHA]ی ؟[BHB]؟[PFH][Z]	H3I
مضارع التزامی مجهول	[VFP]؟(ب ان)؟[G]ه (ب ان)؟[IMH]؟[PFH]	H3T
آینده معلوم	؟[Z][VFP]؟(ب ان)؟[G]خواه؟[PFH][Z] ؟[Z][VFP]؟(ب ان)؟[G]؟[Z]	A1I

	عنوان پروژه:		 شورای عالی اطلاع رسانی
	فاز اول طرح جامع پیکره زبان فارسی با موضوع فاز اول مطالعاتی ایجاد پیکره متنی زبان فارسی		
	عنوان زیر پروژه:		
	تاریخ: 1388/03/31	ویرایش: 1/0	کد زیر پروژه: پیکرمتن فارس - 2 - ب

آینده مجهول	[IMG] [PFH] خواه؟(ب ان ام) [G] ه [Z][VPF] (ب ان ام)؟ [A1T]
امر معلوم	[Z]([BHB])؟ [BHA] [AHB] ای؟ [AHA] ای؟ (ب ان ام)؟ [Z][VPF] [A2I]
امر مجهول	[G][VPF] ه (ب ان ام) شو(ید)؟ [A2T]

جدول 60 - الگوی فعل‌های پیشوندی



در A1I جزء غیر صرفی فعل در یکی از دو موقعیت مشخص شده قرار می‌گیرد.

3-5 فعل‌های مرکب



شامل یک بخش ثابت است که آنرا با «VC» نشان می‌دهیم. این بخش ثابت می‌تواند پسوند به خود بگیرد. این پسوندها می‌توانند ضمیر یا ادات جمع یا احتمالا موارد دیگر باشند که همگی آنها را با «X» نشان می‌دهیم.

ترتیب اجزا در فعل‌های مرکب ممکن است عوض شود. البته در گفتار معمول، جزء ثابت ابتدا می‌آید سپس جزء تصریفی، اما در متون ادبی امکان دارد اجزا جابه‌جا شوند.

ماضی ساده معلوم	$\varphi[X][VC] = R_1$	G1I
	$\varphi[Z][PFGS][G](ب ان ام) = R_2$	
	$R_2R_1 \mid R_1R_2 = R$	
ماضی استمراری معلوم	$\varphi[X][VC] = R_1$	G2I
	$\varphi[Z][PFGS][G] \diamond \varphi(ن) = R_2$	
	$R_2R_1 \mid R_1R_2 = R$	
ماضی بعید معلوم	$\varphi[X][VC] = R_1$	G3I
	$\varphi[Z][PFGS] \text{ بود} [G] \varphi(ب ان) = R_2$	

	عنوان پروژه: فاز اول طرح جامع پیکره زبان فارسی با موضوع فاز اول مطالعاتی ایجاد پیکره متنی زبان فارسی		 شورای عالی اطلاع رسانی
	عنوان زیر پروژه: بررسی راهکارهای ایجاد ماشین تشخیص هوشمند افعال مرکب در زبان فارسی		
	تاریخ: 1388/03/31	ویرایش: 1/0	

ماضی مستمر معلوم	$R_2R_1 \mid R_1R_2 = R$ $\{[Z][PFGS]\} \text{ داشت} = R_1$ $\{[X][VC]\} = R_2$ $\{[Z][PFGS][G]\} \text{ می} \{ن\} = R_3$	G4I
	$R_1R_3R_2 \mid R_1R_2R_3 = R$ $\{[X][VC]\} = R_1$	
ماضی ساده نقلی معلوم	$\{[Z][PFGN]\} \{ب\} \{ن\} = R_2$ $R_2R_1 \mid R_1R_2 = R$ $\{[X][VC]\} = R_1$	G5I
	$\{[Z][PFGN]\} \{ن\} \text{ می} = R_2$	G6I
ماضی استمراری نقلی معلوم	$R_2R_1 \mid R_1R_2 = R$ $\{[X][VC]\} = R_1$	
	$\{[Z][PFGN]\} \text{ بوده} \{ب\} \{ن\} = R_2$ $R_2R_1 \mid R_1R_2 = R$ $\{[Z][PFGN]\} \text{ داشته} = R_1$	G7I
ماضی بعید نقلی معلوم	$\{[X][VC]\} = R_2$ $\{[Z][PFGN]\} \{ن\} \text{ می} = R_3$	G8I
	$R_2R_1 \mid R_1R_2 = R$ $\{[X][VC]\} = R_1$	
ماضی التزامی معلوم	$\{[Z][PFGS]\} \{ن\} \text{ باش} \{ب\} \{ن\} = R_2$ $R_2R_1 \mid R_1R_2 = R$ $\{[X][VC]\} = R_1$	G9I
	$\{[Z][PFGN]\} \{ن\} \text{ می} \{ب\} \{ن\} = R_2$	H1I
مضارع اخباری معلوم	$R_2R_1 \mid R_1R_2 = R$	

	عنوان پروژه:		 شورای عالی اطلاع رسانی
	فاز اول طرح جامع پیکره زبان فارسی با موضوع فاز اول مطالعاتی ایجاد پیکره متنی زبان فارسی		
	عنوان زیر پروژه:		
	تاریخ: 1388/03/31	ویرایش: 1/0	کد زیر پروژه: پیکرمتن فارس - 2 - ب



مضارع مستمر معلوم	$\{[X][PFH]\} = R_1$	H2I
	$\{[Z][VC]\} = R_2$	
	$\{[Z][PFH]([HB])\} \text{ ای } \{[HA]\} = R_3$	
	$R_2R_1 \mid R_1R_2 = R$	
مضارع التزامی معلوم	$\{[X][VC]\} = R_1$	H3I
	$\{[Z][PFH]([BHB])\} \text{ ای } \{[BHA]\} \text{ ای } \{[AHB]\} \text{ ای } \{[AHA]\} = R_2$	
	$R_2R_1 \mid R_1R_2 = R$	
آینده معلوم	$\{[X]\{[VC]\} = R_1$	A1I
	$\{[Z][PFH]\} \text{ خواه } \{[VC]\} = R_2$	
	$\{[Z][G]\} \text{ ای } \{[VC]\} = R_3$	
	$R_2R_1R_3 \mid R_1R_2R_3 = R$	
امر معلوم	$\{[X][VC]\} = R_1$	A2I
	$\{[Z]([BHB])\} \text{ ای } \{[BHA]\} \text{ ای } \{[AHB]\} \text{ ای } \{[AHA]\} = R_2$	
	$R_2R_1 \mid R_1R_2 = R$	

جدول 61 - الگوی فعل‌های مرکب



گاهی اوقات در تشخیص فعل ابهام ایجاد می‌شود؛ یعنی گاهی بخش ثابت جزء فعل مرکب است و گاهی نیست. مثلاً «دوست داشتن» در جمله «من در مدرسه تعدادی دوست دارم» فعل نیست، بلکه «داشتن» فعل است. این رفع ابهام مربوط به مرحله بالاتر پردازش در پردازش نحوی است.

4-5 فعل‌های پیشوندی مرکب



فعل‌های پیشوندی مرکب هم دارای یک پیشوند هستند که به «VPF» نشان داده می‌شود و هم دارای کلمه ثابت که با «VC» نشان داده می‌شود.

	عنوان پروژه:		 شورای عالی اطلاع رسانی
	فاز اول طرح جامع پیکره زبان فارسی با موضوع فاز اول مطالعاتی ایجاد پیکره متنی زبان فارسی		
	عنوان زیر پروژه:		
	تاریخ: 1388/03/31	ویرایش: 1/0	کد زیر پروژه: پیکرمتن فارس - 2 - ب

ماضی ساده معلوم	$\varphi[X][VC] = R_1$	G1I
	$\varphi[Z][PFGS][G](ب\text{ان}م)[VPF] = R_2$	
	$R_2R_1 \mid R_1R_2 = R$	
ماضی استمراری معلوم	$\varphi[X][VC] = R_1$	G2I
	$\varphi[Z][PFGS][G]\diamond\varphi\text{می}(ن)[VPF] = R_2$	
	$R_2R_1 \mid R_1R_2 = R$	
ماضی بعید معلوم	$\varphi[X][VC] = R_1$	G3I
	$\varphi[Z][PFGS] \text{بوده} [G]\varphi(ب\text{ان})[VPF] = R_2$	
	$R_2R_1 \mid R_1R_2 = R$	
ماضی مستمر معلوم	$\varphi[Z][PFGS] \text{داشت} = R_1$	G4I
	$\varphi[X][VC] = R_2$	
	$\varphi[Z][PFGS][G]\diamond\varphi\text{می}(ن)[VPF] = R_3$	
	$R_1R_3R_2 \mid R_1R_2R_3 = R$	
ماضی ساده نقلی معلوم	$\varphi[X][VC] = R_1$	G5I
	$\varphi[Z][PFGN] \text{ه} [G]\varphi(ب\text{ان})[VPF] = R_2$	
	$R_2R_1 \mid R_1R_2 = R$	
ماضی استمراری نقلی معلوم	$\varphi[X][VC] = R_1$	G6I
	$\varphi[Z][PFGN] \text{ه} [G]\diamond\varphi\text{می}(ن)[VPF] = R_2$	
	$R_2R_1 \mid R_1R_2 = R$	
ماضی بعید نقلی معلوم	$\varphi[X][VC] = R_1$	G7I
	$\varphi[Z][PFGN] \text{بوده} [G]\varphi(ب\text{ان})[VPF] = R_2$	
	$R_2R_1 \mid R_1R_2 = R$	

	عنوان پروژه: فاز اول طرح جامع پیکره زبان فارسی با موضوع فاز اول مطالعاتی ایجاد پیکره متنی زبان فارسی		 شورای عالی اطلاع رسانی
	عنوان زیر پروژه: بررسی راهکارهای ایجاد ماشین تشخیص هوشمند افعال مرکب در زبان فارسی		
	تاریخ: 1388/03/31	ویرایش: 1/0	

ماضی مستمر نقلی معلوم	$\{[Z][PFGN]\}$ داشته = R_1	G8I
	$\{[X][VC]\} = R_2$	
	$\{[Z][PFGN] \text{ه} [G] \text{می} \text{ن} \text{ن} [VPF]\} = R_3$	
ماضی التزامی معلوم	$R_2 R_1 \mid R_1 R_2 = R$	G9I
	$\{[X][VC]\} = R_1$	
	$\{[Z][PFGS] \text{ه} [G] \text{ن} \text{باش} \text{ن} [VPF]\} = R_2$	
مضارع اخباری معلوم	$R_2 R_1 \mid R_1 R_2 = R$	H1I
	$\{[X][VC]\} = R_1$	
	$\{[Z][PFH] \text{ی} [HA] \text{ی} [HB] \text{ی} \text{می} \text{ن} [VPF]\} = R_2$	
مضارع مستمر معلوم	$R_2 R_1 \mid R_1 R_2 = R$	H2I
	$\{[X][VC]\} = R_1$	
	$\{[Z][VC]\} = R_2$	
مضارع التزامی معلوم	$\{[Z][PFH] \text{ی} [HA] \text{ی} [HB] \text{ی} \text{می} \text{ن} [VPF]\} = R_3$	H3I
	$R_2 R_1 \mid R_1 R_2 = R$	
	$\{[X][VC]\} = R_1$	
آینده معلوم	$\{[Z][PFH] \text{ی} [HA] \text{ی} [HB] \text{ی} \text{می} \text{ن} [VPF]\} = R_2$	A1I
	$\{[X][VC]\} = R_1$	
	$\{[Z][G] \text{ی} [VC] \text{ی} [VPF]\} = R_3$	
	$R_2 R_1 R_3 \mid R_1 R_2 R_3 = R$	

	عنوان پروژه:		 شورای عالی اطلاع رسانی
	فاز اول طرح جامع پیکره زبان فارسی با موضوع فاز اول مطالعاتی ایجاد پیکره متنی زبان فارسی		
	عنوان زیر پروژه:		
	تاریخ: 1388/03/31	ویرایش: 1/0	کد زیر پروژه: پیکرمتن فارسی - 2 - ب
بررسی راهکارهای ایجاد ماشین تشخیص هوشمند افعال مرکب در زبان فارسی			

امر معلوم	$\{X\}[VC] = R_1$ $\{Z\}([BHB])\{[BHA][AHB]\{[AHA]\{[VPF] = R_2 \text{ A2I}$ $R_2 R_1 \mid R_1 R_2 = R$
-----------	---------------------------------------------------------------------------------------------------------

جدول 62 - الگوی فعل‌های پیشوندی مرکب



5-5 عبارتهای فعلی

عبارتهای فعلی پیچیده‌ترین و بغرنج‌ترین ساختار فعلی در ادبیات فارسی هستند. این ساختار فعل به هیچ وجه دارای ساختار کلی و یکپارچه برای تمام افعال متعلق به آن نیست. این افعال دارای یک بخش ثابت و یک بخش صرف شدنی هستند. عبارت ثابت در این افعال، از کلمات مختلفی تشکیل شده‌اند که امکان افزودن پیشوند یا پسوند به هر کدام از آنها وجود دارد و همچنین امکان وقوع جابه‌جایی در ترتیب کلمات آن وجود دارد. لذا در اینجا فرمول کلی برای این افعال ارائه نمی‌دهیم چرا که هر کدام از این افعال دارای شکل و الگوی مجزایی هستند؛ بلکه در بخش 0 راهکاری برای بیان و جستجوی این افعال ارائه می‌دهیم.

5-6 فعل‌های لازم یک‌شخصه

فعل‌های لازم یک‌شخصه الگوی تصریف محدودی دارند و ترتیب بخش‌های فل در آنها ثابت است. اما ممکن است بین بخش‌های فعل فاصله بیفتد.

ماضی ساده معلوم	$\{Z\}([G])([BG][AG])\{D\}\{[Z][VC] \text{ G1I}$
ماضی استمراری معلوم	$\{Z\}[G]\{D\}\{[Z][VC] \text{ G2I}$
ماضی بعید معلوم	$\{Z\}\{D\}\{[Z][VC] \text{ G3I}$

	عنوان پروژه: فاز اول طرح جامع پیکره زبان فارسی با موضوع فاز اول مطالعاتی ایجاد پیکره متنی زبان فارسی		 شورای عالی اطلاع رسانی
	عنوان زیر پروژه: بررسی راهکارهای ایجاد ماشین تشخیص هوشمند افعال مرکب در زبان فارسی		
	تاریخ: 1388/03/31	ویرایش: 1/0	

ماضی مستمر معلوم	G4T	داشت [Z]؟ [Z][VC]{D}؟ [Z][VC]{D}؟ (ن) می [G] [Z]؟
ماضی ساده نقلی معلوم	G5I	[Z][VC]؟ [D]؟ (ب) ان [G] ه (است) [Z]؟
ماضی استمراری نقلی معلوم	G6I	[Z][VC]؟ [D]؟ (ن) می [G] ه (است) [Z]؟
ماضی بعید نقلی معلوم	G7I	[Z][VC]؟ [D]؟ (ب) ان [G] ه بوده (است) [Z]؟
ماضی مستمر نقلی معلوم	G8I	داشته (است) [Z]؟ [D]؟ [Z][VC]{D}؟ (ن) می [G] ه (است) [Z]؟
ماضی التزامی معلوم	G9I	[Z][VC]؟ [D]؟ (ب) ان [G] ه (ن) باشد [Z]؟
مضارع اخباری معلوم	H1I	[Z][VC]؟ [D]؟ (ن) می [HA] ای [HB] د [Z]؟
مضارع مستمر معلوم	H2I	دارد [Z][VC]{D}؟ [D]؟ (ن) می [HA] ای [HB] د [Z]؟
مضارع التزامی معلوم	H3I	[Z][VC]؟ [D]؟ (ب) ان ام [Y] ای [AHA] ای [AHB] ای [BHA] ای [BHB] د [Z]؟
آینده معلوم	A1I	[Z][VC]؟ [D]؟ (ب) ان ام؟ خواهد (ب) ان ام [Y] ای [AG] ای [BG] ای [G]؟
امر معلوم	A2I	-



جدول 63 - الگوی فعل‌های لازم یک شخصه

5-7 فعل‌های خاص



5-7-1 فعل ربطی (اسنادی)

معروفترین فعل‌های ربطی یا اسنادی عبارتند از: استیدن، بودن، شدن، باشیدن. همچنین افعال «گشتن» و «گردیدن» اگر به معنی «شدن» باشند، فعل ربطی به شمار می‌روند. این افعال در تمام زمان‌ها صرف نمی‌شوند.

ماضی ساده معلوم	G1I	(ن) بود [PFGS]
ماضی ساده مجهول	G1T	(ب) ان ام شد [PFGS]
ماضی استمراری معلوم	G2I	(ن) می بود [PFGS]

	عنوان پروژه:		 شورای عالی اطلاع رسانی
	فاز اول طرح جامع پیکره زبان فارسی با موضوع فاز اول مطالعاتی ایجاد پیکره متنی زبان فارسی		
	عنوان زیر پروژه:		
	تاریخ: 1388/03/31	ویرایش: 1/0	کد زیر پروژه: پیکرمتن فارسی - 2 - ب

ماضی استمراری مجهول	G2T	(ن)؟می [PFGS][IMG]
ماضی بعید معلوم	G3I	-
ماضی بعید مجهول	G3T	-
ماضی مستمر معلوم	G4T	-
ماضی مستمر مجهول	G5I	داشت [PFGS]{D} (ن)؟می شد [PFGS]
ماضی ساده نقلی معلوم	G5I	(ن)؟بوده [PFGN]
ماضی ساده نقلی مجهول	G5T	(ب ن)؟[IMG]ه [PFGN]
ماضی استمراری نقلی معلوم	G6I	(ن)؟می بوده [PFGN]
ماضی استمراری نقلی مجهول	G6T	(ن)؟می [IMG]ه [PFGN]
ماضی بعید نقلی معلوم	G7I	بوده [PFGN]
ماضی بعید نقلی مجهول	G7T	(ب ن)؟[IMG]ه بوده [PFGN]
ماضی مستمر نقلی معلوم	G8I	-
ماضی مستمر نقلی مجهول	G8T	داشته [PFGN]{D} (ن)؟می [IMG]ه [PFGN]
ماضی التزامی معلوم	G9I	(ن)؟باش [PFGS]
ماضی التزامی مجهول	G9T	(ب ن م)؟[IMG]ه باش [PFGS]
مضارع اخباری معلوم	H1I	(بی)؟ست (ا ه)ست [PFH]
مضارع اخباری مجهول	H1T	-
مضارع مستمر معلوم	H2I	-
مضارع مستمر مجهول	H2T	دار [PFH]{D} (ن)؟می [IMH] [PFH]
مضارع التزامی معلوم	H3I	(ن م)؟(باش) [PFH]
مضارع التزامی مجهول	H3T	(ب ن م)؟[IMH] [PFH]
آینده معلوم	A1I	(ب ن م)؟خواه [PFH] بود

	عنوان پروژه:		 شورای عالی اطلاع رسانی
	فاز اول طرح جامع پیکره زبان فارسی با موضوع فاز اول مطالعاتی ایجاد پیکره متنی زبان فارسی		
	عنوان زیر پروژه:		
	تاریخ: 1388/03/31	ویرایش: 1/0	کد زیر پروژه: پیکرمتن فارس - 2 - ب

آینده مجهول	A1T (ب ان ام)؟ خواه [PFH] [IMG]
امر معلوم	A2I باش
امر مجهول	A2T بشو

جدول 64 - الگوی فعل‌های ربطی

همچنین فعل «است» قابلیت چسبیدن به کلمه قبل از خود را دارد که باید بطور جداگانه پردازش شود.

5-7-2 فعل‌های شبه کمکی

در جملاتی مانند «میخواستم ... بگویم» یا «می توانستی ... بگویی» یا «بایست ... می گفت»، ما دو فعل با دو معنی مجزا استنباط می کنیم. در واقع برقراری ارتباط بین این دو فعل در تحلیل ساختار جمله صورت می گیرد.

5-7-3 فعل غیر شخصی

فعل‌های شبه کمکی «توانستن»، «بایستن» و «شایستن» گاهی فعلی می سازند که بر شخص معینی دلالت نمی کنند؛ مانند «توان رفت»، «نباید گفت»، «نشاید رفت». الگوی صرف این افعال به صورت زیر است:



ن؟ (توان | باید | شاید) [G]

5-7-4 فعل وصفی

الگوی تصریف این افعال به صورت زیر است:



(ب ان ام)؟ [G]ه

تشخیص دقیق این افعال در پردازش گره نحوی صورت می گیرد.

	عنوان پروژه: فاز اول طرح جامع پیکره زبان فارسی با موضوع فاز اول مطالعاتی ایجاد پیکره متنی زبان فارسی		 شورای عالی اطلاع رسانی
	عنوان زیرپروژه: بررسی راهکارهای ایجاد ماشین تشخیص هوشمند افعال مرکب در زبان فارسی		
	تاریخ: 1388/03/31	ویرایش: 1/0	

5-7-5 حذف فعل به قرینه

تشخیص این ساختار فعل باید در تحلیل ساختار جمله صورت بگیرد و ما در اینجا به آن نمی پردازیم.

	عنوان پروژه: فاز اول طرح جامع پیکره زبان فارسی با موضوع فاز اول مطالعاتی ایجاد پیکره متنی زبان فارسی		 شورای عالی اطلاع رسانی
	عنوان زیر پروژه: بررسی راهکارهای ایجاد ماشین تشخیص هوشمند افعال مرکب در زبان فارسی		
	تاریخ: 1388/03/31	ویرایش: 1/0	

5-8 ویژگی‌های فعل مرکب فارسی

با توجه به الگوهای تصریفی فعل که در بخش‌های قبل بیان شد، علاوه بر مشکلات تشخیص افعال ساده، افعال مرکب ویژگی‌هایی دارند که مسایلی را سر راه تشخیص آنها به وجود می‌آورد:



- فاصله بین اجزای فعل
- جابه‌جا شدن اجزای فعل
- اتصال فعل و سایر کلمات

این امکان وجود دارد که بین اجزای فعل مرکب فاصله ایجاد شود. به عنوان مثال فعل مرکب «گمان داشتن» در جمله «او گمان بد دارد». البته در افعال ساده نیز در برخی زمان‌ها امکان ایجاد فاصله بین اجزای فعل وجود دارد. مانند جمله «او دارد سبب می‌خورد». مکانیزم تشخیص برای افعال ساده و مرکب یکسان است. تقریباً تمام افعال مرکب قابلیت جدا شدن اجزا دارند.

همچنین در زبان فارسی جابه‌جا شدن اجزای فعل ممکن است اتفاق بیافتد - خصوصاً در متون ادبی. به عنوان مثال فعل «ستم کردن» در جمله «کردن ستم بسیار». بطور کلی فرض بر این می‌گیریم که تمام افعال مرکب قابلیت جابه‌جا شدن اجزا را دارند.

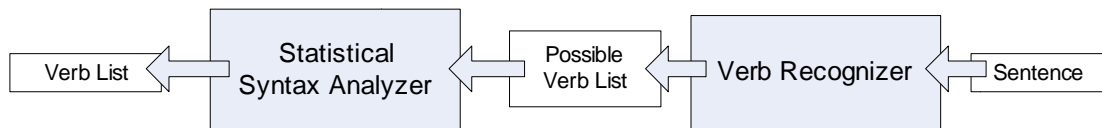
همچنین در زبان فارسی بسیار متداول است که فعل به ضمائر یا سایر کلمات می‌چسبد. به عنوان مثال «گفتمش» و «عالیست». در مورد ضمائر، همه‌ی آنها در الگوهای صرف لحاظ شده‌اند. در مورد فعل خاص «است» باید به طور جداگانه بررسی شود.

در بخش بعدی روش خود را برای تشخیص افعال در جمله ارائه می‌دهیم.

	عنوان پروژه:		 شورای عالی اطلاع رسانی
	فاز اول طرح جامع پیکره زبان فارسی با موضوع فاز اول مطالعاتی ایجاد پیکره متنی زبان فارسی		
	عنوان زیر پروژه:		
تاریخ: 1388/03/31	ویرایش: 1/0	کد زیر پروژه: پیکرمتن فارس - 2 - ب	

6. روش سلسله مراتبی در تشخیص فعل



در این بخش روش خود را برای جستجوی فعل در جمله بر اساس الگوی عبارت منظم توضیح می‌دهیم. با توجه به الگوهای تعریف شده در بخش 5، می‌بینیم که تعداد زیادی الگوی تصریفی وجود دارد و برای هر فعل باید یک ورودی در پایگاه داده تعریف شود که بن فعل و بخش‌های ثابت آن و طبقه‌ی ساختاری فعل را مشخص کند. لذا می‌بینیم که برای ارزیابی این الگوها در متن پردازش زیادی باید صرف شود. اما در واقع لازم نیست همه‌ی الگوهای عبارات منظم بررسی شوند؛ بلکه در اینجا ما یک روش سلسله مراتبی ارائه می‌دهیم که قادر است به میزان قابل توجهی از حجم پردازش بکاهد.



شکل 2 - مراحل پردازش جمله

بطور کلی رویکرد ما در مسأله‌ی تشخیص فعل، یک کار دو مرحله‌ایست. در مرحله‌ی اول - که موضوع این گزارش است - ما افعال را فقط بر اساس صورت ظاهری آنها تشخیص می‌دهیم. این کار باعث می‌شود تا هر شکلی از کلمه که با یکی از ساختارهای فعل مشابهت ظاهری داشته باشد به عنوان فعل تشخیص داده و معرفی شود. افعال ابتدایی تشخیص داده شده ممکن است کلمات اضافی دیگری را به عنوان فعل معرفی کند اما هرگز هیچ فعلی را از قلم نخواهد انداخت. همچنین ممکن است این افعال معرفی شده، دارای بخش‌های مشترک با یکدیگر باشند. در مرحله‌ی بعدی - که ما به آن پردازش سطح بالاتر می‌گوییم، سعی می‌شود تا محتمل‌ترین افعال را به عنوان نتیجه نهایی بیابد. اساس کار مرحله بعد می‌تواند بر پردازش نحوی جمله یا کارهای آماری باشد که ما به آن «پردازشگر نحوی آماری» می‌گوییم. در شکل 2 مراحل پردازش جمله به صورت کلی دیده می‌شود. لذا در روش ما رفع ابهام در مرحله‌ی بالاتر پردازشی صورت می‌گیرد.

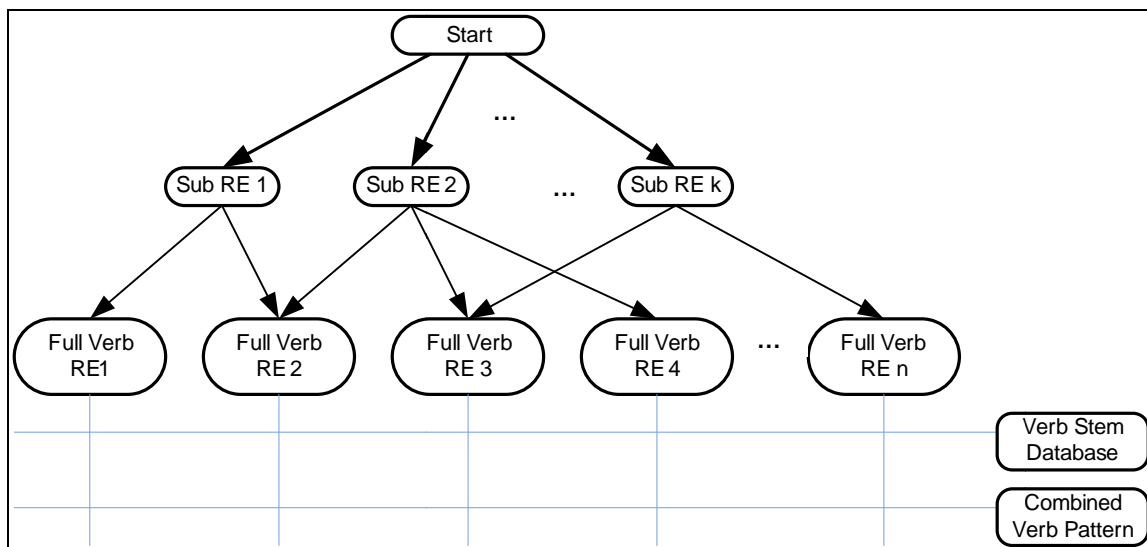
در بخش «تشخیص‌دهنده فعل»، ورودی یک جمله است و خروجی آن لیست تمام افعال ممکن. در این بخش است که وجود افعال در جمله با توجه به الگوی تصریف آنها و همچنین پایگاه داده صورت می‌گیرد. در پردازش جمله، ورودی حرف به حرف از درون ساختار اتومات ساخته شده عبور کرده و در

	عنوان پروژه: فاز اول طرح جامع پیکره زبان فارسی با موضوع فاز اول مطالعاتی ایجاد پیکره متنی زبان فارسی		 شورای عالی اطلاع رسانی
	عنوان زیر پروژه: بررسی راهکارهای ایجاد ماشین تشخیص هوشمند افعال مرکب در زبان فارسی		
	تاریخ: 1388/03/31	ویرایش: 1/0	

صورت تشخیص فعل آنرا به عنوان یک واژه فعلی به پردازشگر سطح بالاتر می فرستد. در واقع طراحی و پیاده سازی الگوریتم برای تشخیص افعال ساده و مرکب، در قالب یک الگوی یکپارچه، صورت گرفته.

6-1 طراحی ساختار داده



ما الگوهای منظم را در یک گراف که دارای یک ساختار سلسله مراتبی است، طوری سازماندهی می کنیم که کمترین پردازش ممکن برای بررسی عبارات منظم مصرف شود. ساختار داده طراحی شده برای تشخیص فعل بر اساس الگوی عبارات منظم در در شکل 3 دیده می شود.



شکل 3 - طراحی ساختار سلسله مراتبی الگوی فعل

در شکل 3 گرافی دیده می شود که شبیه یک گراف جریان داده است. در ابتدا نود شروع قرار دارد و در سطح پایین «زیرالگو»ها قرار می گیرند و در سطح بعدی، الگوهای کامل فعل قرار دارند.

به جای اینکه الگوی کامل تصریفی افعال را برای تمام ساختارها بررسی کنیم، ابتدا زیرالگویی از تمام الگوها استخراج می کنیم. این زیرالگو نشان دهنده اینست که اگر آن در جمله دیده شود احتمال اینکه الگوی اصلی در جمله باشد، وجود دارد. این زیرالگوها در میان الگوهای اصلی فعل ممکن است مشترک باشند؛ لذا بعد از بررسی زیرالگوها، ممکن است چند الگو در مرحله بعد نامزد بررسی شوند. همانطور که

	عنوان پروژه:		 شورای عالی اطلاع رسانی
	فاز اول طرح جامع پیکره زبان فارسی با موضوع فاز اول مطالعاتی ایجاد پیکره متنی زبان فارسی		
	عنوان زیر پروژه:		
	تاریخ: 1388/03/31	ویرایش: 1/0	کد زیر پروژه: پیکرمتن فارس - 2 - ب
بررسی راهکارهای ایجاد ماشین تشخیص هوشمند افعال مرکب در زبان فارسی			



در شکل نیز دیده می‌شود، از زیرالگوها به سمت الگوها فلش‌هایی وجود دارد که در واقع نشان دهنده رابطه زیرمجموعه بودن است.

زیرالگوها و الگوها به صورت الگوی بدون ریشه هستند، یعنی به جای داشتن تمام ریشه‌ها در داخل عبارت منظم، آنرا به صورت یک کاراکتر عمومی در داخل الگو نمایش می‌دهیم. پس از تطبیق الگو، بخش عمومی را با پایگاه داده مقایسه می‌کنیم. در شکل، پایگاه داده را در سمت راست با عنوان: «پایگاه داده ریشه فعل» و «الگوی افعال مرکب» مشخص کرده‌ایم. به عنوان مثال در مورد بخش ثابت در الگوی افعال مرکب، این امکان وجود دارد که هر کدام از کلمات جمله به عنوان همکرد فعل باشند که ای جستجو در پایگاه داده به صورت سریع انجام می‌شود؛ چون اندیس گذاری بر اساس ریشه فعل و کلمات ثابت فعل، صورت گرفته است.

6-2 الگوریتم

بطور کلی الگوریتم تشخیص به صورت زیر است:

1. ابتدا تمام زیرالگوها را بررسی کن
 2. الگوهای نامزد را بدست آور
 3. الگوهای نامزد را بررسی کن
 4. ریشه فعل و بخش‌های ثابت فعل را بدست آور
 5. ریشه فعل و بخش‌های ثابت را در پایگاه داده جستجو کن
 6. الگوهای یافت شده در پایگاه داده را به عنوان افعال احتمالی به مرحله بعد بده
- توجه کنید که الگوهای فعلی شامل حالت‌های غلط صرفی نیز هستند که این به مراحل بعدی پردازش، اعلام می‌شود.

	عنوان پروژه: فاز اول طرح جامع پیکره زبان فارسی با موضوع فاز اول مطالعاتی ایجاد پیکره متنی زبان فارسی		 شورای عالی اطلاع رسانی
	عنوان زیر پروژه: بررسی راهکارهای ایجاد ماشین تشخیص هوشمند افعال مرکب در زبان فارسی		
	تاریخ: 1388/03/31	ویرایش: 1/0	



3-6 پایگاه داده الگو

مساله‌ای که وجود دارد، چگونگی سازماندهی الگوی افعال در درون پایگاه داده است. همچنین کاربر چگونه باید لیست افعال خود را برای پایگاه داده تعریف کند.

ما یک زبان ساده برای کاربر ارایه می‌دهیم که بتواند افعال را تعریف کند. کاربر باید بخش‌های ثابت و تصریفی فعل و همچنین نوع ساختار فعل را مشخص کند. باقی کارها توسط الگوریتم اندیس گذاری در پایگاه داده انجام خواهد شد. بع عنوان مثال کاربر برای تعریف فعل مرکب «روشن کردن» به صورت زیر عمل می‌کند:



VC(روشن)VERB(کردن)، TYPE(فعل مرکب)

در پایگاه داده این ورودی با استفاده از کلمه بخش ثابت «روشن» و ریشه فعل «کردن»، اندیس گذاری می‌شود. لذا دسترسی یا جستجوی آن بطور سریع انجام می‌شود. همچنین برای سایر ساختارها، نحوه بیان قابل گسترش است.

	عنوان پروژه: فاز اول طرح جامع پیکره زبان فارسی با موضوع فاز اول مطالعاتی ایجاد پیکره متنی زبان فارسی		 شورای عالی اطلاع رسانی
	عنوان زیرپروژه: بررسی راهکارهای ایجاد ماشین تشخیص هوشمند افعال مرکب در زبان فارسی		
	تاریخ: 1388/03/31	ویرایش: 1/0	

7. مراجع

- [۱] K. Megerdooian, "Unification-Based Persian Morphology," in CICLing, 2000.
- [2] ت. و. کامیار، و غ. عمرانی، دستور زبان فارسی: سمت، 1384.
- [3] ح. انوری، و ح. ا. گیوی، دستور زبان فارسی: موسسه فرهنگی فاطمی، 1385.
- [4] ا. دفتری نژاد، "ساختواژه حالت-متناهی: روشی مناسب برای طراحی پردازشگر ساختواژی"، هفتمین همایش زبانشناسی ایران، 1386.
- [۵] K. Megerdooian, "Finite-State Morphological Analysis of Persian," in Workshop on Computational Approach to Arabic Script-Based Languages, 2004.
- [۶] K. Beesley, and L. Karttunen, *Finite State Morphology*: Stanford, CSLI Publications, 2003.
- [۷] K. Taghva, R. Beckley, and M. Sadeh, "A Stemming Algorithm for the Farsi Language," in Proceedings of the International Conference on Information Technology: Coding and Computing (ITCC'05) - Volume I - Volume 01, 2005.
- [۸] A. Mokhtaripour, and S. Jahanpour, "Introduction to a new Farsi stemmer," in Proceedings of the 15th ACM international conference on Information and knowledge management, Arlington, Virginia, USA, 2006.
- [۹] M. I. Mobarakeh, and B. Minaei-Bidgoli, "Verb Detection in Persian Corpus," *International Journal of Digital Content Technology and its Applications* vol. 3, pp. 58-65, 2009.
- [۱۰] R. Hessami-Fard, and G. Ghasem-Sani, "Stemmer Algorithm Design for Persian Language," in 11th International CSI Computer Conference (CSICC'2006), Tehran, Iran, 2006.
- [۱۱] M. F. Porter, "An algorithm for suffix stripping," *Program* 14(3), pp. 130-167, 1980.
- [۱۲] K. S. Jones, and P. Willett, *Readings in information retrieval*: Morgan Kaufmann Publishers Inc., 1997.
- [۱۳] M. Arabsorkhi, and M. Shamsfard, "Unsupervised Discovery of Persian Morphemes," in 11th Conference of the European Chapter of the Association for Computational Linguistics (EACL), 2006.

	عنوان پروژه: فاز اول طرح جامع پیکره زبان فارسی با موضوع فاز اول مطالعاتی ایجاد پیکره متنی زبان فارسی		 شورای عالی اطلاع رسانی
	عنوان زیرپروژه: بررسی راهکارهای ایجاد ماشین تشخیص هوشمند افعال مرکب در زبان فارسی		
	تاریخ: 1388/03/31	ویرایش: 1/0	

[14] ر. روح‌پرور، و م. بی‌جن‌خان، "به کارگیری یک نظام برچسب‌دهی برای تعبیر و تفسیر یک پیکره متنی زبان فارسی،" in هفتمین همایش زبان‌شناسی ایران، 1386.

[15] م. محسنی، و ب. م. بیدگلی، "مدل مارکوف مرتبه‌ی دو برای برچسب‌گذاری پیکره زبان فارسی،" در هفتمین همایش زبان‌شناسی فارسی، 1386.

[۱۶] M. Mohseni, H. Motalebi, B. Minaei-bidgoli *et al.*, "A Farsi part-of-speech tagger based on Markov model," in Proceedings of the 2008 ACM symposium on Applied computing, Fortaleza, Ceara, Brazil, 2008.

[۱۷] Megerdooian, and Karine, "Developing a Persian Part-of-Speech Tagger," in First Workshop on Persian Language and Computers, Tehran University, Iran, 2004.

[۱۸] R. Makki, and M. M. Homayounpour, "Word Sense Disambiguation of Farsi Homographs Using Thesaurus and Corpus," *Advances in Natural Language Processing, Springer Berlin / Heidelberg*, vol. 5221/2008, pp. 315-323, August 28, 2008.

[۱۹] T. Winograd, *Language As a Cognitive Process: Syntax*: Addison-Wesley, 1982.

[۲۰] J. E. Hopcroft, R. Motwani, and J. D. Ullman, *Introduction to Automata Theory, Languages, and Computation*, 2nd ed.: Addison Wesley, 2000.

[۲۱] K. Atkinson, "GNU Aspell," GNU, 2008.