

Corpus-based Thematic Classification and Analysis of Persian Prefixed and Compound Verbs

Mohammadhadi Fallahi

Department of Computational Linguistics, Islamic world science and technology monitoring and citation institute, Shiraz, Iran. (Corresponding author)

fallahi@isc.ac.ir

[ORCID: 4355-6601-0002-0000](https://orcid.org/4355-6601-0002-0000)

Razieh Amiri

Payam e noor university

a.amiri3798@gmail.com

[ORCID: 5814-2757-0000-0009](https://orcid.org/5814-2757-0000-0009)

Introduction

Verbs occupy a central position in both traditional and contemporary linguistic theories and are widely regarded as the core of sentence structure. Persian verbs have traditionally been classified from different perspectives, including morphological, syntactic, semantic, and modal characteristics. Among these approaches, the subjective classification of verbs offers a corpus-based perspective that focuses on actual language use. Rather than relying solely on the formal properties of verbs, this approach examines the frequency of verb usage across different subjective domains, including political, economic, social, sports, scientific, cultural, health and medical, and legal texts. This perspective makes it possible to distinguish highly frequent general-purpose verbs from domain-specific verbs and to identify the most frequently used verbs within each subjective domain.

The present study aims to provide a subjective classification of Persian prefixed and compound verbs based on their frequency of occurrence across different subjective domains. Such a classification can be used in several applied fields, particularly in teaching Persian to non-native speakers, translation studies, and computational linguistics. It provides teachers and learners with a practical resource for identifying the most frequently used prefixed and compound verbs, assists translators in selecting domain-specific verbs, and supplies corpus-based lexical information that can support natural language processing applications. Accordingly, this study investigates the frequency-based distribution of Persian prefixed and compound verbs across different subjective domains, identifies the most frequently used verbs in each domain, distinguishes general-purpose verbs from domain-specific verbs, and examines their usage patterns.

Method

This study focused on Persian prefixed and compound verbs extracted from four Persian corpora: the Bijankhan Corpus, the Updating Persian Corpus, the Hamshahri Corpus, and the Persika Corpus. Candidate verbs were first identified using the Persian Verb Syntactic Valency Lexicon and then conjugated into all major inflectional forms using a Persian verb conjugation tool to enable comprehensive retrieval from these corpora.

To ensure consistency across the corpora, the original subjective categories were consolidated into eight unified subjective domains through consultation with specialists in linguistics, natural language processing, and information science. These domains comprised political, economic, social, sports, scientific, cultural, health and medical, and legal texts. All extracted verbs were subsequently assigned to the appropriate subjective domain according to the classification of their source texts.

Several existing Persian natural language processing tools were evaluated for automatic verb extraction. However, none of these tools provided sufficient accuracy for reliably identifying the full range of Persian prefixed and compound verb forms. Consequently, a Python-based system employing regular expressions was developed to automatically extract all inflectional forms of the target verbs. Due to the computational complexity of processing large-scale corpora, the entire Bijankhan corpus was analyzed, whereas random samples of equivalent size were selected from the remaining corpora. Finally, the frequency of each extracted verb was calculated within each subjective domain, and the verbs were classified according to their frequency of occurrence.

Results

A comparison between the 50 most frequent Persian prefixed and compound verbs in the combined corpora and those in each subjective domain revealed that the social, political, and scientific domains contained the fewest domain-specific verbs, whereas the remaining subjective domains exhibited domain-specific high-frequency verbs.

From a total of 301,004 sampled sentences across all corpora, 260,777 prefixed and compound verbs were extracted. The 50 most frequent verbs accounted for 119,212 occurrences (45.71%) of all extracted verbs. No domain-specific verbs were identified among the 50 most frequent verbs in the social, political, and scientific domains, whereas some verbs were identified as domain-specific in more than one subjective domain.

The analysis also revealed differences in the diversity of prefixed and compound verbs across subjective domains. The social, scientific, and political domains exhibited the broadest range of different verbs, while the legal and health domains showed the smallest inventories. In addition, a considerable number of prefixed and compound verbs occurred only once, with 494 single-occurrence verbs identified across all corpora.

Discussion and Conclusion

The findings demonstrate that, while the social, political, and scientific domains are dominated by general-purpose high-frequency verbs, the remaining subjective domains exhibit a greater presence of domain-specific verbs. They also show that the social, scientific, and political domains employ a broader range of prefixed and compound verbs than the other subjective domains.

The frequency-based subjective classification developed in this study provides a corpus-based resource for identifying both general-purpose and domain-specific Persian prefixed and compound verbs. This classification can facilitate the teaching of Persian as a foreign language by enabling learners to focus on the most frequent verbs at each stage of language learning. It can also support translation studies by helping translators identify more appropriate verb equivalents for different subjective contexts.

Abstract

This study classified and examined Persian prefixed and compound verbs based on their frequency of use across eight thematic domains: social, economic, health, legal, political, scientific, cultural, and sports, using the Hamshahri, Persika, Bijankhan, and Updated Persian Text Corpus. The study was conducted by first extracting prefixed and compound verbs from the Syntactic Valency Lexicon of Persian Verbs (Rasouli and others, 2012). Subsequently, an eight-category classification of prefixed and compound verbs was developed by the researchers in consultation with specialists in linguistics, natural language processing, and information and knowledge science. This classification includes the political, economic, social, sports, scientific, cultural, health, and legal categories. Before verb extraction, the topic labels assigned to the corpora were consolidated into these eight categories. Next, verbs were extracted from the corpora. All verbs were conjugated using verb conjugation software, and their conjugated forms were extracted from the corpora. The frequency of each verb was calculated in each thematic domain. Finally, a thematic list of Persian prefixed and compound verbs was prepared according to the frequency of their use in each domain. The findings showed that the social, political, and scientific domains did not include domain-specific verbs among the most frequent verbs, whereas the economic, health, legal, cultural, and sports domains contained a greater number of domain-specific verbs. Furthermore, the social, scientific, and political domains showed relatively greater breadth and diversity in the use of Persian prefixed and compound verbs than the other thematic domains.

Keywords: prefixed and compound verbs, thematic classification, language corpus

طبقه‌بندی موضوعی پیکره بنیاد و بررسی افعال پیشوندی و مرکب زبان فارسی

محمد‌هادی فلاحی، استادیار گروه زبان‌شناسی رایانشی، موسسه استنادی و پایش علم و فن آوری جهان اسلام، شیراز، ایران. (نویسنده مسئول)

ایمیل: fallahi@isc.ac.ir

شماره ارکید: ۴۳۵۵-۶۶۰۱-۰۰۰۲-۰۰۰۰

راضیه امیری، دانش آموخته کارشناسی ارشد زبان شناسی، دانشگاه پیام نور شیراز، شیراز، ایران.

ایمیل: a.amiri3798@gmail.com

شماره ارکید: ۵۸۱۴-۲۷۵۷-۰۰۰۰-۰۰۰۹

چکیده

این پژوهش افعال پیشوندی و مرکب زبان فارسی را بر مبنای بسامد کاربرد در حوزه‌های هشت‌گانه موضوعی (اجتماعی، اقتصادی، بهداشت، حقوقی، سیاسی، علمی، فرهنگی و ورزشی) با استفاده از پیکره‌های زبانی همشهری، پرسیکا، بیجن‌خان و به‌روزشونده، طبقه‌بندی و بررسی کرده است. پژوهش حاضر بدین صورت انجام پذیرفت که نخست، افعال پیشوندی و مرکب از سامانه‌ی ظرفیت نحوی افعال زبان فارسی (رسولی و همکاران، ۱۳۹۱) استخراج شد و سپس، طبقه‌بندی هشت‌گانه‌ای برای افعال پیشوندی و مرکب توسط پژوهشگران با مشورت تعدادی از متخصصان زبان‌شناسی، پردازش زبان طبیعی و علم اطلاعات و دانش‌شناسی ارائه گردید. این طبقه‌بندی شامل طبقه‌های سیاسی، اقتصادی، اجتماعی، ورزشی، علمی، فرهنگی، بهداشت و حقوقی است. پیش از استخراج افعال، طبقه‌های برجسب‌خورده در پیکره‌ها ذیل هشت طبقه تجمیع شد. در مرحله‌ی بعدی، استخراج افعال از پیکره‌ها صورت گرفت. سپس، کلیه افعال با استفاده از نرم‌افزار صرف فعل، صرف و صیغه‌های آنها از پیکره‌های مذکور استخراج شد. در مرحله‌ی بعدی، بسامد کاربرد هر فعل در هر یک از حوزه‌های موضوعی محاسبه شد. در پایان، فهرست موضوعی افعال پیشوندی و مرکب بر اساس میزان کاربرد آنها در هر حوزه تهیه گردید. یافته‌ها نشان داد که حوزه‌های اجتماعی، سیاسی و علمی در میان افعال پربسامد فاقد افعال اختصاصی بودند، در حالی که حوزه‌های اقتصادی، بهداشت، حقوقی، فرهنگی و ورزشی از شمار بیشتری از افعال اختصاصی برخوردار بودند. همچنین، حوزه‌های اجتماعی، علمی و سیاسی در مقایسه با سایر حوزه‌های موضوعی از گستره و تنوع بیشتری در کاربرد افعال پیشوندی و مرکب برخوردارند.

کلیدواژه‌ها: افعال پیشوندی، افعال مرکب، طبقه‌بندی موضوعی، پیکره‌ی زبانی

۱. مقدمه

نگاهی اجمالی به رویکردهای سنتی و جدید در دستور زبان، ما را به این واقعیت می‌رساند که فعل، جایگاه ممتازی در کلیه نظریه‌های دستوری دارد. به عبارتی، در اغلب رویکردهای دستوری، فعل هسته‌ی جمله به شمار می‌رود. افعال بر اساس ساختمان، معنی، وجه و ... طبقه‌بندی شده‌اند. تقسیم‌بندی افعال از نظر ساخت‌واژی به سه دسته‌ی ساده، مرکب و پیشوندی، همچنین از نظر نحوی به دو دسته‌ی لازم و متعدی و همچنین تقسیم به وجوه اخباری، امری، پرسشی و التزامی از جمله تقسیم‌بندی‌های افعال به شمار می‌رود که هر تقسیم‌بندی از یک منظر انجام شده است. طبقه‌بندی موضوعی افعال شیوه‌ای نوین در طبقه‌بندی و شناخت کاربرد افعال است که با بهره‌گیری از پیکره‌های زبانی صورت می‌گیرد. در این شیوه میزان کاربرد هر فعل در هر طبقه‌ی موضوعی (سیاسی، اقتصادی، اجتماعی، ورزشی، علمی، فرهنگی، بهداشت و پزشکی و حقوقی) بررسی می‌شود و فهرست این افعال با توجه به بسامد آن‌ها ارائه می‌شود. به زبان ساده، این طبقه‌بندی نشان می‌دهد که در هر طبقه‌ی موضوعی چه افعالی بیشترین کاربرد را دارند و چه افعالی عمومی و چه افعالی تخصصی محسوب می‌شوند.

هدف اصلی این پژوهش، طبقه‌بندی موضوعی افعال پیشوندی و مرکب زبان فارسی است. فهرست طبقه‌بندی شده‌ی افعال می‌تواند در حوزه‌های آموزش زبان فارسی به غیرفارسی‌زبانان، ترجمه و زبان‌شناسی رایانشی مورد استفاده قرار گیرد. در حوزه‌ی آموزش زبان فارسی به غیرفارسی‌زبانان، این فهرست موضوعی افعال می‌تواند به معلم و زبان‌آموز، عمومی‌ترین و کاربردی‌ترین افعال پیشوندی و مرکب را که یادگیری آنها ضروری است، ارائه دهد. از سوی دیگر، به مترجمان، فهرست افعال تخصصی هر رشته یا رشته‌های نزدیک به هم را ارائه می‌دهد که در امر ترجمه می‌تواند مفید باشد. در هر دو حوزه‌ی آموزش زبان و ترجمه، این فهرست موضوعی می‌تواند افراد را از یادگیری فوری حجم بزرگی از افعال بی‌نیاز سازد.

این پژوهش در پی پاسخ‌گویی به پرسش‌های زیر است:

- در هر کدام از طبقات موضوعی، کدام افعال به ترتیب بیشترین میزان کاربرد را دارند؟
- افعال عمومی (افعالی که بالاترین بسامد را در همه‌ی طبقات موضوعی دارند)، کدامند؟
- افعال تخصصی هر طبقه‌ی موضوعی (افعالی که تنها در هر یک از طبقات موضوعی، بسامد بالاتری دارند) کدامند؟

۲. پیشینه پژوهش

طبقه‌بندی چیزهایی که انسان با آنها سروکار داشته است، اعم از اشیاء، پدیده‌ها و علوم گوناگون، از دیرباز مورد توجه انسان بوده است؛ زیرا برای نظم بخشیدن به ساختار ذهنی و فکری خویش به این طبقه‌بندی نیاز دارد. طبقه‌بندی در هر عملی ابزاری کارآمد برای شناخت و معرفت است (مصطفوی فرد، ۱۳۹۳). با توجه به این که طبقه‌بندی موضوعات می‌تواند بسیار کلی و یا بسیار جزئی باشد و معیار دقیقی برای تعیین مرز بین این موضوعات نیست (مثلاً مرز بین موضوعات هنری، فرهنگی، ادبی و اجتماعی)، در این پژوهش، هشت طبقه‌ی موضوعی که دربردارنده‌ی همه‌ی موضوعات جزئی‌تر باشند، مدنظر قرار گرفت که عبارتند از: سیاسی، اقتصادی، اجتماعی، ورزشی، علمی، فرهنگی، بهداشت و پزشکی و حقوقی. طبقه‌بندی افعال به دو دسته‌ی افعال عمومی و تخصصی، آموزش، یادگیری و ترجمه‌ی افعال و متون را تسهیل می‌کند. نمونه‌ی این فهرست‌بندی را در زبان انگلیسی در کتاب‌های ۱۱۰۰ واژه، ۵۰۴ واژه و کتاب‌هایی از این دست، می‌توان یافت که واژه‌های عمومی و پرکاربرد زبان انگلیسی را معرفی می‌کنند. تا کنون در پژوهش‌های متعددی به بررسی

شناسایی، طبقه‌بندی و تحلیل افعال پرداخته شده است. در ادامه، برخی از پژوهش‌های مرتبط در این زمینه مرور می‌شوند. پژوهش تسلیمی پور (۱۳۹۱) بر روی پردازش خودکار افعال مرکب زبان فارسی با ساخت (اسم+فعل) متمرکز شده است. در این پژوهش از کلاس‌های معنایی واژه‌ها استفاده شده است. همچنین، در تعیین گروه‌های معنایی واژه‌ها از مدل فضای برداری استفاده شده است و معنای واژه‌ها در فضای برداری نمایش داده شده است. عسکریان (۱۳۹۱) نیز به شناسایی خودکار افعال مرکب فارسی پرداخته است. وی به رایج‌ترین اشکال فعل مرکب فارسی پرداخته و علاوه بر معیارهای استاندارد همبستگی، از معیارهای پایستگی و ترجمه‌ای برای آموزش دسته‌بندی SVM برای شناسایی آنها استفاده کرده است. چاقری و شمس‌فرد (۱۳۹۲) برای شناسایی افعال ساده و مرکب زبان فارسی دو روش یکی مبتنی بر رویکرد بدون نظارت و دیگری رویکرد بانظارت را ارائه کرده‌اند. در هر دو رویکرد، نخست، افعال ساده و سبک جملات با استفاده از برجسب نحوی واژه‌ها و ریشه‌یابی مشخص شده‌اند و سپس به یافتن بقیه‌ی افعال پرداخته شده است. به ادعای نویسندگان این شیوه دقت بالایی در تشخیص افعال دارد. کریمی دوستان و چراغی (۱۳۹۲) به طبقه‌بندی افعال زبان فارسی بر اساس ساخت رویدادی و نمودی پرداخته‌اند. در این پژوهش، پیکره‌ای متشکل از ۶۰ فعل به پیروی از نظریه‌ی کرافت^۱ (۲۰۱۲) مورد بررسی قرار گرفته است و افعال به ۴ طبقه‌ی اصلی و ۱۰ طبقه‌ی فرعی تقسیم شده‌اند. شیبانی مقدم و غلامی (۱۳۹۳) به کاربرد افعال وجهی در پیکره‌ی دستی شش متن از قرن ششم هجری پرداخته‌اند و نتایج پژوهش آنها نشان داد که میزان استفاده از این نوع افعال در کنار دیگر انواع فعل، چشمگیر است.

گرابوفسکی^۲ و مینت^۳ (۱۹۹۵) افعال بی‌قاعده‌ی زبان انگلیسی را از پیکره‌های براون^۴ برای انگلیسی آمریکایی و لب^۵ برای انگلیسی بریتانیایی استخراج کردند و با توجه به میزان بسامد، آنها را طبقه‌بندی کردند. والد^۶ (۱۹۸۸) با استفاده از پایگاه دادگان وردنت^۷ و پیکره‌ی ملی انگلستان^۸ که شامل ۱۰۰ میلیون ترکیب واژگانی از انگلیسی بریتانیایی است که از متون گفتاری و نوشتاری به دست آمده است، با توجه به ویژگی‌های نحوی و معنایی افعال، آنها را طبقه‌بندی کرده است. همچنین شیوه‌های یافتن و حدس فعل نیز ارائه شده است. گال و همکاران^۹ (۲۰۰۴) با استفاده از پیکره‌ی انگلیسی آمریکایی، بسامد زیرطبقه‌ی ۲۸۱ فعل را که در حوزه‌ی روان‌شناسی زبان کاربرد گسترده‌ای دارند، مورد بررسی قرار دادند. این شیوه نقش مهمی در پردازش‌های زبانی دارد. وینس^{۱۰} و سیریک^{۱۱} (۲۰۱۰) با بهره‌گیری از متون روزنامه‌ای، اخبار اقتصادی و متون حقوقی به تهیه‌ی پیکره‌ای اقدام نمودند و ساخت گروه فعلی افعال سبک را در زبان مجارستانی بررسی کردند. استیونسون^{۱۲} و مارلو^{۱۳} (۲۰۱۰) با استفاده از شیوه‌ی یادگیری ماشین به طبقه‌بندی افعال زبان انگلیسی با توجه به معیارهای دستوری پرداختند. آنان افعال را در سه طبقه‌ی غیرکنایی^{۱۴} غیرمفعولی^۱ و مفعول انداز^۲ طبقه‌بندی کردند و

-
1. Craft
 2. Grabowski
 3. Mindt
 4. Brown
 5. Lob
 - 6 Walde.
 7. Wordnet
 8. British National Corpus
 ۹. Gahl
 10. Vincze
 11. Csirik
 12. Stevenson
 13. Marlo
 14. unergative

برای هر طبقه ۲۰ فعل را مورد بررسی قرار دادند. گراس^۳ و هیرشمن^۴ (۲۰۱۸) به جمع‌آوری ۳۹۰ فعل عملی و چگونگی انجام آنها و توصیف آنها از منظر ۱۲ شرکت‌کننده پرداختند. داده‌های این طرح شامل داده‌های صوتی، تصویری و حرکتی بود.

۳. روش پژوهش

جامعه‌ی آماری این پژوهش کلیه‌ی افعال پیشوندی و مرکب زبان فارسی است و نمونه‌های این پژوهش نیز افعال پیشوندی و مرکب زبان فارسی است که از پیکره‌های بیجن‌خان (۲۰۱۱)، به‌روزشونده (۱۳۹۷)، همشهری (۲۰۰۹) و پرسیکا (۲۰۱۲) استخراج شده است.

مراحل انجام این پژوهش به شرح ذیل است:

الف) استخراج افعال پیشوندی و مرکب از سامانه‌ی ظرفیت نحوی افعال زبان فارسی (رسولی و همکاران، ۱۳۹۱).
ب) تعیین طبقه‌های موضوعی متون: پیکره‌ی همشهری شامل ۱۶ طبقه‌ی موضوعی و پیکره‌ی پرسیکا شامل ۱۱ طبقه‌ی موضوعی است. پیکره‌ی به‌روزشونده‌ی زبان فارسی نیز شش طبقه‌ی موضوعی را بازمی‌شناساند. و پیکره‌ی بیجن‌خان نیز شامل ۹۲ طبقه‌ی موضوعی است.

در این پژوهش، طبقه‌بندی افعال با میزان کاربرد آنان در متون موضوعی مختلف انجام شد. دسته‌بندی متون از دهه‌ی ۱۹۶۰ میلادی آغاز شده و امروزه با پیشرفت علوم رایانه، نتایج چشمگیری داشته است. با توجه به این‌که مبنای طبقه‌بندی افعال، طبقه‌بندی متون است و طبقه‌بندی جامعی برای متون وجود ندارد، باید یک طبقه‌بندی جامع ارائه شود. بدین منظور، طبقه‌بندی ۸ گانه‌ای توسط پژوهشگران با مشورت متخصصان زبان‌شناسی، پردازش زبان طبیعی و علم اطلاعات و دانش‌شناسی ارائه شد. این طبقه‌بندی شامل طبقه‌های سیاسی، اقتصادی، اجتماعی، ورزشی، علمی، فرهنگی، بهداشت و پزشکی و حقوقی است که افعال به شرح ذیل از پیکره‌ها استخراج گردید:

سیاسی از پیکره‌های به‌روزشونده، همشهری، پرسیکا و بیجن‌خان

- ۱) اقتصادی از پیکره‌های به‌روزشونده، همشهری، پرسیکا و بیجن‌خان
- ۲) اجتماعی از پیکره‌های به‌روزشونده، همشهری، پرسیکا و بیجن‌خان
- ۳) ورزشی از پیکره‌های به‌روزشونده، همشهری، پرسیکا و بیجن‌خان
- ۴) علمی از پیکره‌های به‌روزشونده، پرسیکا و بیجن‌خان
- ۵) فرهنگی از پیکره‌های به‌روزشونده، پرسیکا و بیجن‌خان
- ۶) بهداشت و پزشکی از پیکره‌های پرسیکا و بیجن‌خان
- ۷) حقوقی از پیکره‌های پرسیکا و بیجن‌خان

پیش از استخراج افعال، طبقه‌های برچسب‌خورده در پیکره‌ها، ذیل هشت طبقه‌بندی بالا تجمیع گردید. مثلاً در پیکره‌ی همشهری، طبقه‌های موضوعی بورس و بانک، اقتصاد جهانی و اقتصادی ذیل طبقه‌ی موضوعی «اقتصادی» قرار گرفت و یا در پیکره‌ی بیجن‌خان طبقات موضوعی ادبی، ادبیات، تاریخ، تاریخی، داستانی و ... ذیل طبقه‌ی «فرهنگی» قرار گرفت و یا طبقات زبان‌شناسی، روان‌شناسی، زیست‌شناسی، ژنتیک و .. در زیر طبقه‌ی «علمی» قرار گرفت (جدول‌های ۱، ۲، ۳ و ۴).

1. unaccusative
2. object-drop
3. Gross
4. Hirshman

ج) استخراج افعال از پیکره‌ها: کلیه‌ی افعال در تمام صیغه‌ها با استفاده از نرم افزار صرف فعل، صرف شد (حال ساده، حال استمراری، حال التزامی، گذشته‌ی ساده، گذشته‌ی استمراری، گذشته‌ی بعید، گذشته‌ی التزامی، آینده، مجهول، امر) سپس کلیه‌ی صیغه‌های افعال از پیکره‌ها استخراج شد.

به منظور جست‌وجوی افعال، فرض بر این بود که نرم‌افزاری کاربردی و جامع برای این کار وجود داشته باشد. در گام نخست، نرم‌افزارهای ذیل مورد بررسی قرار گرفت:

- ۱- مجموعه ابزار متن‌باز «هضم» که روی زبان برنامه‌نویسی پایتون اجرا می‌شود و هم از روش پردازش قانون-محور و هم برای برچسب‌زنی صرفی و نحوی از مدل‌های یادگیری ماشین بهره می‌برد.
- ۲- آماده‌ساز استاندارد متن برای زبان فارسی که به تقطیع، تصحیح املائی، تحلیل واژشناختی و برچسب‌گذاری اجزای سخن می‌پردازد (شمس‌فرد و همکاران، ۲۰۱۰)
- ۳- پارسی‌پرداز که شامل مجموعه ابزاری است که از نرمال‌ساز، تقطیع‌گر، برچسب‌زن اجزای سخن، تحلیل‌گر ساخت‌واژی (شامل ریشه‌یاب واژگانی و بدون‌وند)، پیمایشگر وابستگی و درنهایت برچسب‌زن نقش معنایی، تشکیل شده است (سرابی و همکاران، ۲۰۱۳).
- ۴- ابزار پیش‌پردازش متون فارسی که به زبان جاوا تهیه شده و شامل بخش‌های مربوط به نرمال‌سازی، تقطیع، جمله‌یابی، ریشه‌یابی، تعیین ایست‌واژه‌ها، تشخیص عبارات اسمی و موجودیت‌های اسمی است (محسنی و همکاران، ۲۰۱۶).
- ۵- تحلیل‌گر فعل فارسی که به زبان سی‌شارپ نگاشته شده است و بر اساس یادگیری ماشین، به تشخیص افعال فارسی از جمله افعال چندکلمه‌ای و مرکب می‌پردازد (رسولی و همکاران، ۲۰۱۶).

پس از بررسی نرم‌افزارهای ذکر شده و آزمودن آنها، نتایج ذیل حاصل شد:

- بر اساس یافته‌ها، محققین در مجموعه هضم و در بخش ریشه‌یابی به ارائه معیار و نتایج ابزار خود نپرداخته‌اند اما از آنجا که بر اساس روش‌های قانون-محور بدان پرداخته‌اند، بررسی‌های آزمایشگاهی ما به نتایج رضایت‌بخشی نرسید.
- تحلیل‌گر فعل فارسی، دارای محدودیت تعداد واژه در جمله ورودی است و چون برخی جملات مورد نظر ما دارای تعداد واژه‌های بیشتری است، قابل استفاده نبود. ضمن آن‌که دقت آن در آزمایشات اولیه نیز برای تحقیق حاضر قابل اتکا نبود.
- ابزار پیش‌پرداز متون فارسی، به رایگان قابل بارگذاری است اما نتایج بررسی‌ها گویای آن بود که نتایج این پژوهش چندان معتبر نخواهد بود.
- پارسی‌پرداز که شاید بهترین نمونه از انواع فوق بود، به صورت برخط و استفاده از وبگاه آن سنجیده شد، که آن هم نمی‌توانست پاسخگوی نیازهای این پژوهش باشد.
- تمام نتایج ذکرشده درباره نرم‌افزارهای مذکور بر اساس بررسی‌های محققان بر روی ورودی‌های همان تحقیق بوده و چنانچه در مقیاس وسیع‌تر بررسی شود، نتایج آن‌ها ضعیف‌تر خواهد بود.
- الگوریتم‌های موجود مطابق با بررسی‌های آزمایشگاهی نمی‌توانند صورت‌های مختلف صرفی برای مصدرهای حرف اضافه‌ای مرکب، پیش‌وندی مرکب و حرف اضافه‌ای پیش‌وندی مرکب نظیر: به‌هم-

زدن، رخت‌برستن و به‌زاند آمدن را به‌خوبی شناسایی کند. این را می‌توان به افعال مرکب و یا پیش‌وندی نیز کمابیش تعمیم داد.

- صرف استفاده از ریشه‌یاب و آن هم ریشه‌یاب‌های یادشده، برای شناسایی افعال چندکلمه‌ای و یا جداشونده، حتی به همراهی برچسب‌زن صرفی آنان کارآمد نیست؛ چرا که به شناسایی اجزای غیرفعلی به صورت وابسته عنصر فعلی نمی‌پردازد و سبب ایجاد ابهام است که فعل شناخته‌شده، ساده است و یا صورت‌های دیگر آن نظیر مرکب و پیش‌وندی. کارآمدی این سیستم‌ها در مقیاس عملیاتی، نیازمند تحقیقی جداگانه است.

- تنها پیمایشگرهای دستور وابستگی برای این منظور مفید هستند، اما بنا به دانسته‌های کنونی، سیستمی مبتنی بر شبکه عصبی و یادگیری عمیق که به صورت آزاد در دسترس محققان باشد، هنوز وجود ندارد و پیمایشگرهای وابستگی نیز دقت مطلوبی ندارند. آموزش یک مدل پیمایشگر وابستگی بر اساس روش‌های یادگیری ماشینی مستلزم سامانه‌ای است که متن را متناسب با پیکره آموزش داده‌شده نرمال‌سازی، تقطیع، ریشه‌یابی و برچسب‌زنی صرفی کند. بنابر یافته‌های فعلی، هیچ‌یک از سامانه‌های آزاد با این نیاز سازگار نیستند؛ زیرا پیکره‌های برچسب‌خورده و سامانه‌ها، رویکردهای متفاوتی در برچسب‌زنی صرفی و نحوی دارند.

با توجه به نتایج فوق، اقدام به نگارش سامانه‌ای به زبان پایتون گردید که با استفاده از عبارات باقاعده به صورت تضمینی به شناسایی تمام صورت‌های صرفی مصدرهای زبان فارسی می‌پردازد. این نگارش شامل برخی محدودیت‌ها نیز هست از جمله آن‌که صورت‌های چسبیده افعال اسنادی نظیر «خوبم» به معنی خوب هستم را شناسایی نمی‌کند. همچنین در این نسخه به علت نیاز به داشتن سیستمی غربالگر، به شناسایی افعال جداشدنی که یک یا چندین کلمه میان جزء غیرفعلی و فعلی آنها فاصله انداخته، پرداخته نشده است.

به علت وجود حالات صرفی متعدد برای افعال زبان فارسی، و نیز لزوم تقدم شناسایی و غربال‌گری برخی افعال مرکب بر ساده، نظیر «ادامه دادن» بر «دادن» سامانه نگاشته شده دارای پیچیدگی اجرایی بسیار بالایی است و حتی با روش‌های پردازش موازی، سیستم در دسترس نیز قادر به انجام پردازش‌ها بر روی تمامی داده‌های موجود نبوده است. تنها پیکره‌ای که به صورت کامل مورد بررسی قرار گرفت، پیکره بیجن‌خان بود و برای سایر پیکره‌ها نیز به علت حجم بسیار بالای آنها، هم‌حجم پیکره بیجن‌خان اقدام به نمونه‌برداری تصادفی شده است. شایان ذکر است که مقاله‌ای که به معرفی این نرم‌افزار فعل‌یاب می‌پردازد، در حال نگارش است.

سامانه‌ها و پیکره‌های مورد استفاده در این پژوهش عبارتند از: فرهنگ ظرفیت نحوی افعال فارسی (نسخه‌ی ۳،۰) (رسولی و همکاران، ۲۰۱۱)، مجموعه‌ی همشهری (نسخه‌ی ۲) (آل احمد و همکاران، ۲۰۰۹)، پرسیکا (اقبال‌زاده و همکاران، ۲۰۱۲)، پیکره‌ی بی‌جن‌خان (بیجن‌خان، ۱۳۸۳) و پیکره‌ی به‌روزشونده‌ی متون فارسی.

(د) محاسبه‌ی بسامد افعال و طبقه‌بندی آنها با توجه به موضوعات متون: در این مرحله بسامد کاربرد هر فعل در هریک از طبقات موضوعی محاسبه شد.

با توجه به محدودیت‌های ذیل، نمونه‌گیری تصادفی صورت گرفت:

- بروز مشکلات پیش‌بینی‌نشده در استخراج افعال (یک‌دست نبودن پیکره‌ها، مشکلات فاصله و نیم‌فاصله و ...) که گاهی به استخراج چندباره افعال منجر می‌شد. در برخی موارد به دلیل همپوشانی بین فعل ساده و افعال مرکب، نخست افعال

مرکب بازیابی شدند که این امر زمانبر و مشکل آفرین بود (به عنوان مثال، فعل «کردن» که در حالت فعل ساده بودن، همراه با صفات می آید و در غیر این صورت، فعل مرکب خواهد بود).

- در بسیاری از موارد هرچند که ظاهراً کار ماشینی و قاعده مند است، اما نیاز به بازیابی های دستی بسیار زمانبر برای تشخیص افعال ساده از افعال مرکب وجود داشت.

ه) ارائه ی فهرست موضوعی افعال: در این مرحله افعال بر اساس میزان کاربرد در هر طبقه، فهرست شدند و فهرست موضوعی افعال پیشوندی و مرکب تهیه گردید.

حوزه های مختلف برچسب خورده در هر یک از پیکرها به صورت زیر یک دست سازی شده اند:

جدول ۱- موضوعات پیکره ی بیجن خان

آب و هوا	علمی	ترحیم	اجتماعی	دفاع مقدس	اجتماعی
آثار باستانی	فرهنگی	تغذیه	بهداشتی	دندانپزشکی	علمی
آگهی	اقتصادی	جامعه شناسی	علمی	دینی	فرهنگی
آگهی تسلیت	اجتماعی	جدول	فرهنگی	روانشناسی	علمی
آلودگی هوا	اجتماعی	جرم شناسی	حقوقی	ریاضی	علمی
آموزشی	علمی	جغرافیایی	علمی	زبان شناسی	علمی
اجتماعی	اجتماعی	جنایی	حقوقی	زمین شناسی	علمی
اخلاقی	اجتماعی	جنگ تحمیلی	سیاسی	زیست شناسی	علمی
ادبی	فرهنگی	جهانگردی	فرهنگی	زیستی	علمی
ادبیات	فرهنگی	جوانان	اجتماعی	ژنتیک	علمی
اطلاعات عمومی	علمی	حشره شناسی	علمی	سیاسی	سیاسی
اطلاع رسانی	اجتماعی	حقوقی	حقوقی	سیاسی اجتماعی	سیاسی
اقتصادی	اقتصادی	حوادث	اجتماعی	سیاسی ادبی	سیاسی
امنیتی	سیاسی	حیات وحش	علمی	سیاسی اقتصادی	سیاسی
انتظامی	سیاسی	خاطرات سیاسی	سیاسی	سیاسی امنیتی	سیاسی
انرژی	اقتصادی	خاطره	فرهنگی	سیاسی فرهنگی	سیاسی
ایران شناسی	علمی	خاطره سیاسی	سیاسی	سیاسی مذهبی	سیاسی
باستان شناسی	علمی	خدمات	اقتصادی	سیاسی نظامی	سیاسی
باستانی	فرهنگی	خدماتی	اقتصادی	سیاسی و اقتصادی	سیاسی
بهداشتی	بهداشتی	خطبه نماز	سیاسی	سیما	فرهنگی
پزشکی	بهداشتی	جمعه		سینما	فرهنگی
تاریخ	علمی	داستان	فرهنگی	شناسنامه کتاب	علمی
تاریخی	فرهنگی	داستانی	فرهنگی		

علمی	منابع طبیعی	فنی	علمی	اقتصادی	شهرسازی
علمی	مهندسی پزشکی	فهرست مطالب	علمی	اقتصادی	شیلات
بهداشتی	مواد غذایی	فهرست منابع	علمی	علمی	شیمی
اجتماعی	مواد مخدر	فیزیک	علمی	اقتصادی	صنعتی
علمی	نجوم	قضای	حقوقی	اجتماعی	طنز
سیاسی	نظامی	قضایی	حقوقی	علمی	علمی
اقتصادی	نیروگاه اتمی	کامپیوتر	علمی	علمی	علمی و پزشکی
اجتماعی	نیروی انتظامی	کشاورزی	علمی	اقتصادی	عمرانی
فرهنگی	هنری	گردشگری	فرهنگی	علمی	عنوان
اقتصادی	هوایمایی	گیاهشناسی	علمی	فرهنگی	فرهنگی
ورزشی	ورزشی	لبنان	سیاسی	فرهنگی	فرهنگی هنری
سیاسی	یونان	لیست اسامی	علمی	فرهنگی	فرهنگی و هنری
		محیط زیست	علمی	علمی	فضایی
		مذهبی	اجتماعی	علمی	فلسفی

جدول ۲- موضوعات پیکره‌ی هم‌شهری

ورزش	ورزش	ادب و هنر	فرهنگی
علمی	محیط زیست	اجتماعی	اجتماعی
اجتماعی	اجتماعی. معارف	علمی	علمی فرهنگی
اجتماعی	اجتماعی. زنان	اقتصادی	اقتصاد
ورزشی	ورزش. جام جهانی	اجتماعی	گوناگون
اقتصادی	اقتصاد. بورس و بانک	اجتماعی	گوناگون. حوادث
اقتصادی	اقتصاد. نفت	اجتماعی	گوناگون. خارجی
اقتصادی	اقتصاد. صنعت	اجتماعی	گوناگون. شهری
اقتصادی	اقتصاد. کشاورزی	سیاسی	سیاسی

بهداشتی	علمی فرهنگی. علمی. پزشکی و درمان
علمی	علمی فرهنگی. علمی. ارتباطات و فناوری اطلاعات
فرهنگی	ادب و هنر. هنر. تئاتر
سیاسی	سیاسی. سیاست داخلی
اجتماعی	گوناگون. عکس
اقتصادی	اقتصاد. انرژی
اجتماعی	اجتماعی. دانشجو

اقتصاد. بازرگانی	اقتصادی
گردشگری	فرهنگی
ادب و هنر. هنر	فرهنگی
علمی فرهنگی. علمی. کتاب	فرهنگی
علمی فرهنگی. علمی	علمی
گوناگون. شوراها و اسلامی	سیاسی
اقتصاد. ساختمان و مسکن	اقتصادی
ادب و هنر. ادبیات	فرهنگی
ادب و هنر. هنر. موسیقی	فرهنگی
ادب و هنر. هنر. سینما	فرهنگی

جدول ۳- موضوعات پیکره‌ی پرسیکا

اقتصاد و بازرگانی-اقتصادی	اقتصادی
اقتصاد-اقتصادی	اقتصادی
خودرو-اقتصادی	اقتصادی
بازرگانی-اقتصادی	اقتصادی
بورس-اقتصادی	اقتصادی
گردشگری-اقتصادی	اقتصادی
صنعت و معدن-اقتصادی	اقتصادی
کشاورزی و دامپروری-اقتصادی	اقتصادی
بهداشت و درمان-عمومی-بهداشتی	بهداشتی
بهداشت و درمان-خانواده-بهداشتی	بهداشتی
علمی-فناوری-علمی	علمی
سیاسی-سیاسی	سیاسی
سیاسی خارجی-سیاسی	سیاسی
مجلس-سیاسی	سیاسی
بین الملل-فرمانطقه‌یی-سیاسی	سیاسی
بین الملل-فلسطین-سیاسی	سیاسی

آموزشی-آموزشی	علمی
اجتماعی-خانواده-اجتماعی	اجتماعی
صنفی-آموزشی-آموزشی	علمی
اجتماعی-اجتماعی	اجتماعی
استان‌ها-اجتماعی	اجتماعی
انرژی-اجتماعی	اجتماعی
اجتماعی-شهری-اجتماعی	اجتماعی
اجتماعی-محیط زیست-اجتماعی	اجتماعی
راه و ساختمان-اجتماعی	اجتماعی
حوادث-اجتماعی	اجتماعی
صنفی-اجتماعی	اجتماعی
اجتماعی-آموزش و پرورش-اجتماعی	اجتماعی
کار و اشتغال-اجتماعی	اجتماعی
تاریخ-تاریخی	علمی

فقه و حقوق-حقوق سیاسی-سیاسی	حقوقی
بین الملل-سیاسی	سیاسی
پاکستان ، هند و چین-سیاسی	سیاسی
آسیا-خاورمیانه-سیاسی	سیاسی
جامعه اطلاعاتی-ارتباطات-علمی	علمی
جامعه اطلاعاتی-فناوری اطلاعات-علمی	علمی
پژوهشی-علمی	علمی
علمی-علمی	علمی
فناوری-علمی	علمی
فرهنگ و هنر-فرهنگی	فرهنگی
فرهنگ و هنر-سینما-فرهنگی	فرهنگی
نگاهی به وبلاگ‌ها-فرهنگی	فرهنگی
صنفی-فرهنگی-فرهنگی	فرهنگی
فرهنگ و حماسه-فرهنگی	فرهنگی
فرهنگ و ادب-فرهنگی	فرهنگی
تلویزیون و رادیو-فرهنگی	فرهنگی
فرهنگ و ادب-ادبیات-فرهنگی	فرهنگی
فرهنگ و هنر-هنرهای تجسمی-فرهنگی	فرهنگی
فرهنگ و ادب-حکمت و فلسفه-فرهنگی	علمی
فرهنگ و ادب-کتاب-فرهنگی	فرهنگی
فرهنگ و ادب-میراث فرهنگی-فرهنگی	فرهنگی
فقه و حقوق-فقه و حقوق	حقوقی
میراث فرهنگی-فرهنگی	فرهنگی
فقه و حقوق-حقوق اجتماعی-فقه و حقوق	حقوقی
فقه و حقوق-حقوق اقتصادی-فقه و حقوق	حقوقی
فقه و حقوق-فقه-فقه و حقوق	حقوقی
حج و زیارت-مذهبی	اجتماعی
ورزشی-کشتی-ورزشی	ورزشی
ورزشی-فوتبال-ورزشی	ورزشی
قهرمانی-همگانی-ورزشی	ورزشی
ورزشی-ورزشی	ورزشی
جام جهانی ۲۰۱۰-ورزشی	ورزشی
وزارت ورزش و جوانان-ورزشی	ورزشی
پارالمپیک پکن-ورزشی	ورزشی
اجتماعی-جوانان-اجتماعی	اجتماعی
ایران شناسی-اجتماعی	علمی
اجتماعی-زنان-اجتماعی	اجتماعی
آسیب‌های اجتماعی-اجتماعی	اجتماعی
اجتماعی-حوادث-اجتماعی	اجتماعی
اجتماعی-معارف-اجتماعی	اجتماعی
صنایع دستی-اقتصادی	اقتصادی
بهداشت و درمان-بهداشتی	بهداشتی
همسایگان ایران-سیاسی	سیاسی
سیاسی خارجی-ایران-سیاسی	سیاسی
انرژی هسته‌یی-سیاسی	سیاسی
افغانستان-سیاسی	سیاسی
آسیای میانه ، روسیه-سیاسی	سیاسی
حاشیه خلیج فارس-سیاسی	سیاسی

فرهنگی	بین‌الملل-گزارش و تحلیل-سیاسی	سیاسی
میراث و گردشگری-فرهنگی	سیاسی خارجی-ایران در جهان-سیاسی	سیاسی
رسانه-فرهنگی	پایان نامه-علمی	علمی
فرهنگ و هنر-مکتب اصفهان-فرهنگی	جامعه اطلاعاتی-علمی	علمی
تقریب مذاهب-مذهبی	فناوری استراتژیک-فناوری برتر-علمی	علمی
ادیان-مذهبی	فناوری استراتژیک-هوا-فضا-علمی	علمی
جام ملتهای آسیا-ورزشی	فناوری استراتژیک-نظامی-علمی	علمی
المپیاد ورزشی دانشجویان-ورزشی	فرهنگ و هنر-تئاتر-فرهنگی	فرهنگی
یونیورسیاد ازمیر-ورزشی	فرهنگ و هنر-موسیقی-فرهنگی	فرهنگی
بازیهای آسیایی دوحه-ورزشی	فرهنگ و هنر-رادیو و تلویزیون-فرهنگی	فرهنگی

جدول ۴- موضوعات پیکره‌ی به‌روزشونده

فرهنگی	فرهنگی
اقتصادی	اقتصادی
سیاسی	سیاسی
علمی	علمی
اجتماعی	اجتماعی
ورزشی	ورزشی

با توجه به حجم بسیار زیاد پیکره‌ها و عدم امکان پردازش همه‌ی آنها با رایانه‌های شخصی، از همه‌ی پیکره‌ها بجز پیکره‌ی بیجن‌خان، نمونه‌گیری صورت گرفت. با محاسبه‌ی کران بالا و پایین و ضریب اطمینان ۹۵٪، نمونه‌گیری انجام شد.

نشانه	توضیح
	مصدر

$n_{conf}(t)$	تعداد بروز صورت‌های صرفی مصدر
S	تعداد جمله‌ی مفروض
$R(t)$	نسبت بروز صورت‌های صرفی به جمله
$R'(t)$	نسبت نرم بروز صورت‌های صرفی به جمله
$RL(t)$	نسبت لگاریتمی بروز صورت‌های صرفی به جمله
$low(t)$	کران پایین فاصله‌ی اطمینان ۹۵ درصد برای $R(t)$
$up(t)$	کران بالای فاصله‌ی اطمینان ۹۵ درصد برای $R(t)$
$LConfInt$	فاصله‌ی اطمینان ۹۵ درصد لگاریتمی

$$R(t) = \begin{cases} \frac{n_{conf}(t)}{S}, & S \neq 0 \\ 0, & S = 0 \end{cases}$$

$$R'(t) = \begin{cases} \frac{n_{conf}(t) + 1}{S}, & S \neq 0 \\ 0, & S = 0 \end{cases}$$

$$RL(t) = \begin{cases} \frac{-1}{\log(R(t))}, & R(t) \neq 0 \\ 0.0333, & R(t) = 0 \end{cases}$$

$$low(t) = \begin{cases} R(t) - 1.96 \times \sqrt{\frac{R(t)(1-R(t))}{S}}, & R(t) > 1.96 \times \sqrt{\frac{R(t)(1-R(t))}{S}}, R(t) \neq 0 \\ R'(t) - 1.96 \times \sqrt{\frac{R'(t)(1-R'(t))}{S}}, & R(t) < 1.96 \times \sqrt{\frac{R(t)(1-R(t))}{S}}, R(t) \neq 0 \\ 10^{-30}, & R(t) = 0 \end{cases}$$

$$wp(t) = \begin{cases} R(t) + 1.96 \times \sqrt{\frac{R(t)(1-R(t))}{s}}, & R(t) \neq 0 \\ 10^{-50}, & R(t) = 0 \end{cases}$$

$$LConflnt = \left(\frac{-1}{\log(low(t))}, \frac{-1}{\log(wp(t))} \right)$$

اگر دو مصدر j و i موجود باشد و $RL(i)$ بزرگ‌تر از $RL(j)$ باشد، بدین معنی است که در شرایط همسان، اولی کاربرد بیشتری نسبت به دومی داشته است.

۴. یافته‌های پژوهش

۱.۴. آمار توصیفی

از مجموع ۹۰۴۵۴۵۴ جمله‌ی کل پیکره‌ها، ۳۰۱۰۰۴ جمله به صورت تصادفی انتخاب شده‌اند و افعال آن‌ها مورد بررسی قرار گرفته است. تعداد کل جملات و تعداد جملات نمونه در هر یک از حوزه‌ها در کل پیکره‌ها در جدول ۵ نشان داده شده است.

جدول ۵- آمار مجموع جملات پیکره‌ها و جملات نمونه

همه	سیاسی	اقتصادی	اجتماعی	ورزشی	علمی	فرهنگی	بهداشتی	حقوقی	
۳۰۱۰۰۴	۵۸۶۴۲	۳۲۶۴۷	۶۱۴۴۴	۲۵۸۵۲	۶۴۰۴۲	۴۲۶۳۰	۸۶۴۳	۷۱۰۴	جملات نمونه
۹۰۴۵۴۵۴	۱۷۴۴۴۱۲	۱۲۶۴۷۲۸	۲۸۰۴۶۸۴	۹۵۹۵۵۰	۱۰۱۷۴۵۸	۱۱۷۵۰۰۶	۶۷۷۲۸	۱۱۸۸۸	کل جملات

در جدول ۶ آمار تفصیلی تعداد کل جملات هر پیکره و جملات نمونه در هر یک از حوزه‌های هشت‌گانه ارائه شده است.

جدول ۶- آمار تعداد کل جمله‌های هر پیکره و جملات مورد بررسی

حقوقی	بهداشتی	فرهنگی	علمی	ورزشی	اجتماعی	اقتصادی	سیاسی	همه
۴۶۷	۴۰۶,۳	۲۷۶,۱۷	۵۰۵,۱۳	۹۵۳,۳	۸۹۸,۸	۳۵۳,۶	۳۹۶,۲۱	۲۵۴,۷۵

پیکره‌ی بیجن‌خان (کل پیکره و نمونه)

حقوقی	بهداشتی	فرهنگی	علمی	ورزشی	اجتماعی	اقتصادی	سیاسی	همه
-------	---------	--------	------	-------	---------	---------	-------	-----

۴۹۳,۱۲۹	۱۷۰,۱۰	۰۴۰,۸	۰۱۴,۱۵	۸۷۸,۸	۲۴۲,۵۹	۱۷۴,۹	۵۵۴,۷	۴۲۱,۱۱
---------	--------	-------	--------	-------	--------	-------	-------	--------

پیکره‌ی پرسیکا (کل پیکره)

حقوقی	بهداشتی	فرهنگی	علمی	ورزشی	اجتماعی	اقتصادی	سیاسی	همه
۶۳۷,۶	۳۹۰,۴	۳۳۱,۵	۴۲۶,۳۴	۱۵۹,۵	۷۲۵,۸	۶۷۲,۴	۹۱۰,۵	۲۵۰,۷۵

پیکره‌ی پرسیکا (نمونه)

حقوقی	بهداشتی	فرهنگی	علمی	ورزشی	اجتماعی	اقتصادی	سیاسی	همه
۰	۰	۴۸۷,۵۸۹	۵۵۱,۴۱۱	۸۷۴,۵۳۳	۱۲۸,۴۷۶	۸۳۳,۶۰۵	۳۳۵,۱۸۰,۱	۲۲۸,۷۹۷,۳

پیکره‌ی به‌روزشونده (کل پیکره)

حقوقی	بهداشتی	فرهنگی	علمی	ورزشی	اجتماعی	اقتصادی	سیاسی	همه
۰	۰	۶۸۲,۱۱	۱۵۶,۸	۵۸۰,۱۰	۴۳۵,۹	۰۰۶,۱۲	۳۹۱,۲۳	۲۵۰,۷۵

پیکره‌ی به‌روزشونده (نمونه)

حقوقی	بهداشتی	فرهنگی	علمی	ورزشی	اجتماعی	اقتصادی	سیاسی	همه
۰	۷۶۸,۵۶	۰۶۹,۵۵۹	۱۶۰,۵۳۳	۸۴۵,۴۱۲	۶۴۴,۳۰۴,۲	۵۰۲,۶۴۴	۴۹۱,۵۳۲	۴۷۹,۰۴۳,۵

پیکره‌ی همشهری (کل پیکره)

حقوقی	بهداشتی	فرهنگی	علمی	ورزشی	اجتماعی	اقتصادی	سیاسی	همه
۰	۸۴۷	۳۴۱,۸	۹۵۵,۷	۱۶۰,۶	۳۸۶,۳۴	۶۱۶,۹	۹۴۵,۷	۲۵۰,۷۵

پیکره‌ی همشهری (نمونه)

۲.۴. طبقه‌بندی افعال در مجموع حوزه‌ها

در این بخش (جدول ۷) رتبه‌بندی افعال پیشوندی و مرکب که بیش از ۵۰ بسامد داشته اند، در مجموع حوزه‌ها آمده است.

جدول ۷- افعال پیشوندی و مرکب مجموع حوزه‌ها

۸۷۹	معرفی کردن
۸۷۳	ارائه دادن
۸۶۷	به وجود آمدن
۸۶۵	تصمیم گرفتن
۸۶۲	برگزار کردن
۸۶۰	بازی کردن
۸۵۵	صحبت کردن
۸۵۰	زندگی کردن
۸۲۸	اقدام کردن
۸۱۳	حمایت کردن
۸۰۵	اختصاص دادن
۷۹۵	مراجعه کردن
۷۸۸	شکل گرفتن
۷۷۳	افزایش دادن
۷۴۴	نشان کردن
۷۴۱	حرکت کردن
۷۳۸	اتفاق افتادن
705	کسب کردن
۶۸۴	توجه کردن
۶۸۳	دوست داشتن
۶۶۳	بررسی کردن
۶۶۰	به شمار رفتن
۶۳۹	انجام گرفتن
۶۳۶	صادر کردن
۶۲۴	به وجود آوردن
۶۲۲	احساس کردن

۱۵۹۲	بیان کردن
۱۵۶۳	به نظر رسیدن
۱۵۴۵	افزایش یافتن
۱۵۰۶	باز گشتن
۱۴۹۷	عمل کردن
۱۴۹۲	در نظر گرفتن
۱۴۶۵	کمک کردن
۱۳۷۰	کار کردن
۱۳۱۵	تلاش کردن
۱۲۵۶	دریافت کردن
۱۲۳۷	شرکت کردن
۱۱۷۹	آغاز کردن
۱۱۵۶	تشکیل دادن
۱۱۳۷	بر گشتن
۱۱۱۷	از دست دادن
۱۰۸۷	رخ دادن
۱۰۵۶	دست یافتن
۱۰۴۳	ادامه داشتن
۱۰۰۰	سعی کردن
۹۷۰	باز داشتن
۹۶۷	باقی ماندن
۹۴۴	ارائه کردن
۹۴۲	توضیح دادن
۹۴۰	به دست آوردن
۹۱۲	انتخاب کردن
۹۰۹	کاهش یافتن

مصدر	تعداد
قرار گرفتن	۷۸۶۵
وجود داشتن	۷۷۱۴
اعلام کردن	۶۳۳۵
ادامه دادن	۵۲۸۸
نشان دادن	۴۷۶۵
تأکید کردن	۴۲۴۷
تصریح کردن	۴۱۵۲
قرار دادن	۴۰۵۵
اظهار داشتن	۳۴۴۲
انجام دادن	۳۳۸۳
صورت گرفتن	۳۲۹۳
اظهار کردن	۳۲۷۲
استفاده کردن	۳۰۳۲
در یافتن	۳۰۱۴
بر داشتن	۲۹۷۸
قرار داشتن	۲۹۳۷
در آمدن	۲۸۹۵
خاطرنشان کردن	۲۶۶۸
اشاره کردن	۲۴۵۸
بر خوردن	۲۲۰۳
پیدا کردن	۲۰۰۷
ایجاد کردن	۱۷۹۸
فکر کردن	۱۷۹۱
عنوان کردن	۱۷۳۹
اضافه کردن	۱۵۹۶

۴۰۷	تشخیص دادن
۴۰۷	رد کردن
۴۰۴	قبول داشتن
۴۰۲	ارزیابی کردن
۴۰۱	خودداری کردن
۳۹۶	پیشنهاد کردن
۳۹۲	تأکید داشتن
۳۹۱	در رفتن
۳۸۸	فرا خواندن
۳۸۷	در نظر داشتن
۳۸۱	تهدید کردن
۳۶۸	به همراه داشتن
۳۶۳	شکست دادن
۳۶۳	برقرار کردن
۳۵۸	راه یافتن
۳۵۴	دیدار کردن
۳۵۴	توصیف کردن
۳۵۳	منتقل کردن
۳۵۲	درک کردن
۳۵۰	استقبال کردن
۳۴۹	تهیه کردن
۳۴۸	ترجیح دادن
۳۴۸	به شمار آمدن
۳۴۷	از بین بردن
۳۴۷	عبور کردن
۳۴۴	همکاری کردن
۳۴۳	تبدیل کردن

۴۸۳	یاد کردن
۴۸۰	تعیین کردن
۴۷۹	تغییر کردن
۴۷۶	انتظار داشتن
۴۶۵	پرداخت کردن
۴۶۴	شروع کردن
۴۶۲	تأیید کردن
۴۵۷	به دست آمدن
۴۵۶	انتظار رفتن
۴۵۲	مشاهده کردن
۴۵۰	ترک کردن
۴۴۹	باز گردیدن
۴۴۷	تولید کردن
۴۴۶	تصور کردن
۴۳۹	فراموش کردن
۴۳۱	حل کردن
۴۳۱	دعوت کردن
۴۳۰	بر گزیدن
۴۲۷	تغییر دادن
۴۲۷	از بین رفتن
۴۲۳	باز نگریستن
۴۲۲	گزارش دادن
۴۲۱	دفاع کردن
۴۱۳	توصیه کردن
۴۱۱	بر آمدن
۴۰۹	وا گذاردن
۴۰۸	تجربه کردن

۶۲۱	نگاه کردن
۶۰۰	فعالیت کردن
۵۹۶	برخورد کردن
۵۹۲	کاهش دادن
۵۷۶	پیش‌بینی کردن
۵۷۴	ذکر کردن
۵۶۵	سخن گفتن
۵۶۵	دنبال کردن
۵۶۴	حفظ کردن
۵۵۹	بر آوردن
۵۵۶	منتشر کردن
۵۵۵	وارد کردن
۵۵۳	در گذاشتن
۵۴۴	نام بردن
۵۲۶	بر شمردن
۵۲۳	پیش آمدن
۵۲۱	در آوردن
۵۱۹	اجرا کردن
۵۱۵	فرا گرفتن
۵۰۴	ایفا کردن
۵۰۰	جلوگیری کردن
۴۹۹	بر عهده داشتن
۴۹۱	سفر کردن
۴۸۹	تأمین کردن
۴۸۶	هشدار دادن
۴۸۶	ادامه یافتن
۴۸۳	به سر بردن

۲۴۷	نام داشتن
۲۴۶	عرض کردن
۲۴۴	به عهده داشتن
۲۳۳	کنار گذاشتن
۲۳۲	باز یافتن
۲۲۹	نظر داشتن
۲۲۹	هدایت کردن
۲۲۸	اتخاذ کردن
۲۲۸	پی بردن
۲۲۶	افزایش پیدا کردن
۲۲۶	شناسایی کردن
۲۲۵	صورت پذیرفتن
۲۲۴	مبارزه کردن
۲۲۳	ثبت کردن
۲۲۳	کشف کردن
۲۲۳	تلقی کردن
۲۲۲	سقوط کردن
۲۲۰	تحمیل کردن
۲۱۸	بیان داشتن
۲۱۳	در بر گرفتن
۲۱۳	اعتراف کردن
۲۱۱	دستور دادن
۲۰۸	تقسیم کردن
۲۰۵	همراهی کردن
۲۰۳	کنترل کردن
۲۰۳	بر خاستن
۲۰۲	تعلق گرفتن

۲۸۳	موافقت کردن
۲۸۱	چشم خوردن
۲۸۰	نادیده گرفتن
۲۷۹	واگذار کردن
۲۷۶	باز گشودن
۲۷۵	مصرف کردن
۲۷۵	شرکت داشتن
۲۷۴	ثابت کردن
۲۷۴	مخالفت کردن
۲۷۲	قبول کردن
۲۷۱	تماس گرفتن
۲۷۱	پایان دادن
۲۷۰	باز تابیدن
۲۶۸	پایان یافتن
۲۶۵	تحويل دادن
۲۶۴	درخواست کردن
۲۶۴	فرو پاشیدن
۲۶۳	حمله کردن
۲۶۲	خریداری کردن
۲۵۸	تحمل کردن
۲۵۵	روی آوردن
۲۵۴	به کار بردن
۲۵۴	تقویت کردن
۲۵۳	رشد کردن
۲۵۱	در پی داشتن
۲۵۰	بستگی داشتن
۲۴۹	انتقاد کردن

۳۴۱	پیگیری کردن
۳۴۰	اختصاص یافتن
۳۳۹	تعریف کردن
۳۳۷	ارسال کردن
۳۳۶	حضور یافتن
۳۳۴	به دنبال داشتن
۳۳۲	طی کردن
۳۳۲	نگه داشتن
۳۳۱	وا داشتن
۳۳۱	شکست خوردن
۳۲۸	به حساب آمدن
۳۲۶	به عمل آمدن
۳۲۲	به عمل آوردن
۳۲۰	تصویب کردن
۳۱۹	ادعا کردن
۳۱۵	فراهم آوردن
۳۰۵	تشکر کردن
۳۰۳	جلب کردن
۳۰۳	امضا کردن
۳۰۰	رعایت کردن
۲۹۹	سعی داشتن
۲۹۸	عرضه کردن
۲۹۶	حرف زدن
۲۹۵	بر انگیختن
۲۹۰	طول کشیدن
۲۸۵	به کار رفتن
۲۸۴	روی دادن

۱۶۰	تمرین کردن
۱۵۹	مذاکره کردن
۱۵۹	جمع کردن
۱۵۸	اطلاع دادن
۱۵۶	درست کردن
۱۵۵	بر انداختن
۱۵۵	انتقال دادن
۱۵۴	نسبت دادن
۱۵۴	به دنیا آمدن
۱۵۳	حذف کردن
۱۵۱	تحقق یافتن
۱۵۱	دخالت کردن
۱۵۱	گفتگو کردن
۱۵۰	طراحی کردن
۱۴۹	کار داشتن
۱۴۸	جریان داشتن
۱۴۴	تصمیم‌گیری کردن
۱۴۴	پوشش دادن
۱۴۲	انتقال یافتن
۱۴۲	جای دادن
۱۴۲	حضور پیدا کردن
۱۴۱	قائل بودن
۱۴۰	پشتیبانی کردن
۱۴۰	گسترش دادن
۱۳۹	به نظر آمدن
۱۳۹	به تصویر کشیدن
۱۳۸	فرض کردن

۱۷۸	اعمال کردن
۱۷۸	بهره گرفتن
۱۷۷	سوق دادن
۱۷۶	رفتار کردن
۱۷۶	دیدن کردن
۱۷۵	نمایش دادن
۱۷۴	باز گرداندن
۱۷۴	باز پرسیدن
۱۷۲	امکان داشتن
۱۷۱	اداره کردن
۱۷۱	برنامه‌ریزی کردن
۱۷۰	گشایش یافتن
۱۶۸	دست پیدا کردن
۱۶۸	پدید آمدن
۱۶۸	قطع کردن
۱۶۶	قدردانی کردن
۱۶۶	پخش کردن
۱۶۶	اعلام داشتن
۱۶۵	تشویق کردن
۱۶۴	دست زدن
۱۶۳	تکذیب کردن
۱۶۳	هزینه کردن
۱۶۲	تضمین کردن
۱۶۱	رقم زدن
۱۶۱	شکایت کردن
۱۶۱	نتیجه گرفتن
۱۶۰	دستگیر کردن

۲۰۱	اختصاص داشتن
۱۹۹	فرار کردن
۱۹۸	بازدید کردن
۹۸	جذب کردن
۱۹۸	گسترش یافتن
۱۹۷	اعتراض کردن
۱۹۶	تأثیر گذاردن
۱۹۲	باور کردن
۱۹۲	بر گرداندن
۱۹۲	کاهش پیدا کردن
۱۹۱	تکرار کردن
۱۹۰	برطرف کردن
۱۸۸	آموزش دادن
۱۸۸	یاد گرفتن
۱۸۸	تشریح کردن
۱۸۶	سخنرانی کردن
۱۸۶	ادامه پیدا کردن
۱۸۵	صعود کردن
۱۸۵	سرمایه‌گذاری کردن
۱۸۵	نجات دادن
۱۸۵	رسیدگی کردن
۱۸۴	خاطر نشان ساختن
۱۸۳	گمان کردن
۱۷۹	در پیش گرفتن
۱۷۹	مقایسه کردن
۱۷۹	در دست داشتن
۱۷۹	جبران کردن

۱۱۱	ترسیم کردن
۱۱۱	ابراز کردن
۱۱۰	افتخار کردن
۱۰۹	تدوین کردن
۱۰۹	حس کردن
۱۰۹	اذعان کردن
۱۰۸	تعویق افتادن
۱۰۸	توسعه دادن
۱۰۸	تغییر یافتن
۱۰۸	نظارت کردن
۱۰۷	اشاره داشتن
۱۰۷	مدیریت کردن
۱۰۶	باز ماندن
۱۰۶	به قتل رسیدن
۱۰۵	ارجاع دادن
۱۰۵	جان باختن
۱۰۳	سکوت کردن
۱۰۳	بهبود یافتن
۱۰۳	تذکر دادن
۱۰۳	پیروی کردن
۱۰۳	تقدیم کردن
۱۰۲	فرق کردن
۱۰۱	ملاقات کردن
۹۹	راه افتادن
۹۹	تحلیل کردن
۹۹	به یاد داشتن
۹۸	بر جای گذاشتن

۱۲۴	باز رساندن
۱۲۴	از سر گرفتن
۱۲۴	مقاومت کردن
۱۲۲	احترام گذاردن
۱۲۲	جای گرفتن
۱۲۲	مطالعه کردن
۱۲۱	پدید آوردن
۱۲۱	ایجاب کردن
۱۲۰	خدمت کردن
۱۱۹	سر دادن
۱۱۹	گزارش کردن
۱۱۹	ابراز داشتن
۱۱۶	مقابله کردن
۱۱۶	ابلاغ کردن
۱۱۵	نقل کردن
۱۱۵	احتمال داشتن
۱۱۴	قضاوت کردن
۱۱۳	رقم خوردن
۱۱۳	صبر کردن
۱۱۳	راه اندازی کردن
۱۱۳	دقت کردن
۱۱۳	کشتی گرفتن
۱۱۳	انتشار یافتن
۱۱۲	اصرار داشتن
۱۱۲	تحصیل کردن
۱۱۱	صرف کردن
۱۱۱	تأسیس کردن

۱۳۷	تماشا کردن
۱۳۷	راه انداختن
۱۳۷	اهدا کردن
۱۳۷	بروز کردن
۱۳۷	سپری کردن
۱۳۶	دامن زدن
۱۳۵	خلق کردن
۱۳۴	اشتباه کردن
۱۳۴	توسعه یافتن
۱۳۳	به قتل رساندن
۱۳۲	تنظیم کردن
۱۳۲	شکل دادن
۱۳۲	رقابت کردن
۱۳۲	تدارک دیدن
۱۳۰	اصلاح کردن
۱۳۰	غلبه کردن
۱۳۰	اذعان داشتن
۱۲۸	توافق کردن
۱۲۷	در بر داشتن
۱۲۷	پرهیز کردن
۱۲۶	استعفا دادن
۱۲۶	در بردن
۱۲۶	تخمین زدن
۱۲۶	عوض کردن
۱۲۶	فاصله گرفتن
۱۲۵	تلاش داشتن
۱۲۵	ازدواج کردن

۸۰	پس گرفتن
۸۰	خاتمه دادن
۸۰	تشدید کردن
۸۰	به اجرا گذاشتن
۸۰	تکیه کردن
۸۰	فرا رسیدن
۸۰	تعجب کردن
۷۹	پیشرفت کردن
۷۹	خشتی کردن
۷۹	تمدید کردن
۷۹	قیام کردن
۷۹	جای داشتن
۷۸	حاصل کردن
۷۸	ترجمه کردن
۷۸	تضعیف کردن
۷۷	قرائت کردن
۷۷	ایستادگی کردن
۷۷	توجیه کردن
۷۷	اثبات کردن
۷۶	خاتمه یافتن
۷۶	رفع کردن
۷۶	سن داشتن
۷۶	احتمال دادن
۷۶	به خرج دادن
۷۵	صورت دادن
۷۴	زنگ زدن
۷۴	مشورت کردن

۸۶	نگاه داشتن
۸۶	شلیک کردن
۸۵	یادآوری کردن
۸۵	مشارکت کردن
۸۵	طلب کردن
۸۵	کناره گیری کردن
۸۵	راه پیدا کردن
۸۵	القا کردن
۸۴	تدریس کردن
۸۴	عهده داشتن
۸۳	مهاجرت کردن
۸۲	به میان آمدن
۸۲	عذرخواهی کردن
۸۲	تبعیت کردن
۸۲	تبلیغ کردن
۸۲	نصب کردن
۸۲	گریه کردن
۸۲	انکار کردن
۸۱	بازگو کردن
۸۱	در دادن
۸۱	حکومت کردن
۸۱	تمام کردن
۸۱	فوت کردن
۸۱	دست داشتن
۸۱	از دست رفتن
۸۱	نقض کردن
۸۰	اعزام کردن

۹۸	فریاد زدن
۹۸	چاپ کردن
۹۷	پرواز کردن
۹۷	بهبود بخشیدن
۹۶	اعتماد کردن
۹۶	قلمداد کردن
۹۶	بار آوردن
۹۶	تأثیر گذاشتن
۹۶	بحث کردن
۹۵	به خاطر داشتن
۹۵	نفوذ کردن
۹۴	پیاده کردن
۹۳	یاری کردن
۹۲	آموزش دیدن
۸۹	جستجو کردن
۸۹	حمل کردن
۸۹	تربیت کردن
۸۹	جمع آوری کردن
۸۹	گوش دادن
۸۸	انجام پذیرفتن
۸۸	تحويل گرفتن
۸۸	جا داشتن
۸۸	بیعت کردن
۸۷	ملاحظه کردن
۸۷	عقب نشینی کردن
۸۷	تقدیر کردن
۸۷	جلوه دادن

۶۲	ابداع کردن
۶۲	اجتناب کردن
۶۲	ضبط کردن
۶۲	محاسبه کردن
۶۲	سلب کردن
۶۱	اخراج کردن
۶۱	نقش کردن
۶۱	پرورش دادن
۶۱	ریشه کردن
۶۱	راه دادن
۶۱	وضع کردن
۶۰	حکم کردن
۶۰	خدا حافظی کردن
۶۰	فرو رفتن
۶۰	نام گرفتن
۵۹	روایت کردن
۵۹	خیال کردن
۵۹	به دست دادن
۵۹	لمس کردن
۵۹	به هم ریختن
۵۹	دگرگون کردن
۵۹	اشتغال داشتن
۵۹	حدس زدن
۵۹	پاس داشتن
۵۸	اجرا داشتن
۵۸	دعا کردن
۵۸	تحریک کردن

۶۸	از میان رفتن
۶۷	تکان دادن
۶۶	لحاظ کردن
۶۶	تعقیب کردن
۶۵	تفسیر کردن
۶۵	ترتیب دادن
۶۵	ارتقا دادن
۶۵	تخریب کردن
۶۴	اشغال کردن
۶۴	دریغ کردن
۶۴	گوش کردن
۶۴	درمان کردن
۶۴	تخصیص دادن
۶۴	پیشی گرفتن
۶۴	پرتاب کردن
۶۴	باز دادن
۶۳	اعطا کردن
۶۳	اخذ کردن
۶۳	پی گرفتن
۶۳	مرور کردن
۶۳	واریز کردن
۶۲	منع کردن
۶۲	خرج کردن
۶۲	موکول کردن
۶۲	به حساب آوردن
۶۲	شعار دادن
۶۲	تقاضا کردن

۷۳	عادت کردن
۷۳	تبیین کردن
۷۳	اظهار نظر کردن
۷۳	ارتقا یافتن
۷۳	تحریم کردن
۷۳	به کار بستن
۷۲	محافظت کردن
۷۲	زمان بردن
۷۲	فرود آمدن
۷۲	بروز دادن
۷۲	گوشزد کردن
۷۲	به یاد آوردن
۷۱	پناه بردن
۷۱	امتناع کردن
۷۱	شرح دادن
۷۱	تحقق پیدا کردن
۷۱	استعفا کردن
۷۰	نگهداری کردن
۷۰	حساب کردن
۷۰	نقد کردن
۷۰	تکمیل کردن
۷۰	تجاوز کردن
۷۰	ترغیب کردن
۶۹	تحقیق کردن
۶۸	اثر گذاردن
۶۸	پیشگیری کردن
۶۸	اکتفا کردن

۵۳	لازم داشتن	۵۶	سر زدن	۵۸	تزریق کردن
۵۳	کار گذاشتن	۵۶	افتتاح کردن	۵۸	بر گرفتن
۵۳	پس دادن	۵۵	هدر رفتن	۵۸	رو کردن
۵۳	تخصیص یافتن	۵۵	گلزنی کردن	۵۸	بالا گرفتن
۵۲	باز نشستن	۵۵	کنار آمدن	۵۷	راهنمایی کردن
۵۲	استراحت کردن	۵۵	کوتاهی کردن	۵۷	تغذیه کردن
۵۲	جان سپردن	۵۵	مهار کردن	۵۷	جلوه کردن
۵۱	احتمال رفتن	۵۵	خواهش کردن	۵۷	در میان گذاشتن
۵۱	فشار آوردن	۵۴	توزیع کردن	۵۷	پا شدن
۵۱	کفایت کردن	۵۴	احیا کردن	۵۷	حکایت داشتن
۵۱	گسترش پیدا کردن	۵۴	در گرفتن	۵۷	رضایت دادن
۵۱	عنوان داشتن	۵۴	بنا کردن	۵۶	آرزو کردن
۵۱	اقتضا کردن	۵۴	تجمع کردن	۵۶	دست دادن
		۵۴	یاد دادن	۵۶	ادا کردن

در این بخش، به منظور تعیین افعال عمومی (مشترک در کلیه حوزه‌ها) و افعال اختصاصی هر حوزه، ۵۰ فعل نخست کلیه حوزه‌ها که به ترتیب، بیشترین بسامد را داشته‌اند، فهرست شده است (جدول ۸).

جدول ۸- ۵۰ فعل نخست کلیه حوزه‌ها

	کل	اجتماعی	اقتصادی	بهداشتی	حقوقی	سیاسی	علمی	فرهنگی	ورزشی
۱	قرار گرفتن	قرار گرفتن	وجود داشتن	وجود داشتن	اظهار داشتن	اعلام کردن	وجود داشتن	وجود داشتن	قرار گرفتن
۲	وجود داشتن	وجود داشتن	اعلام کردن	قرار گرفتن	وجود داشتن	تأکید کردن	قرار گرفتن	قرار گرفتن	اعلام کردن
۳	اعلام کردن	اعلام کردن	قرار گرفتن	نشان دادن	قرار گرفتن	قرار گرفتن	نشان دادن	ادامه دادن	بازی کردن
۴	ادامه دادن	ادامه دادن	در آمدن	خاطر نشان کردن	ادامه دادن	وجود داشتن	ادامه دادن	در آمدن	انجام دادن

۵	نشان دادن	تصریح کردن	نشان دادن	ادامه دادن	تصریح کردن	ادامه دادن	نشان دادن	ادامه دادن
۶	تأکید کردن	نشان دادن	در یافتن	اعلام کردن	خاطرنشان کردن	قرار دادن	تصریح کردن	فکر کردن
۷	تصریح کردن	تأکید کردن	افزایش یافتن	تصریح کردن	بر خوردن	قرار دادن	قرار دادن	وجود داشتن
۸	قرار دادن	قرار دادن	بر داشتن	استفاده کردن	تأکید کردن	نشان دادن	استفاده کردن	اشاره کردن
۹	اظهار داشتن	بر داشتن	ادامه دادن	صورت گرفتن	عنوان کردن	اظهار کردن	صورت گرفتن	شکست دادن
۱۰	انجام دادن	صورت گرفتن	اظهار کردن	در یافتن	صورت گرفتن	انجام دادن	انجام دادن	قرار داشتن
۱۱	صورت گرفتن	در یافتن	صورت گرفتن	افزایش یافتن	اعلام کردن	اظهار داشتن	خاطرنشان کردن	انجام دادن
۱۲	اظهار کردن	در آمدن	تصریح کردن	تأکید کردن	قرار دادن	صورت گرفتن	تأکید کردن	بیان کردن
۱۳	استفاده کردن	قرار داشتن	تأکید کردن	قرار داشتن	در یافتن	بر خوردن	اعلام کردن	پیدا کردن
۱۴	در یافتن	اظهار کردن	قرار دادن	قرار دادن	اشاره کردن	خاطرنشان کردن	قرار داشتن	اعلام کردن
۱۵	بر داشتن	انجام دادن	استفاده کردن	اظهار کردن	انجام دادن	بر داشتن	در یافتن	قرار داشتن
۱۶	قرار داشتن	استفاده کردن	قرار داشتن	انجام دادن	باز داشتن	باز داشتن	بر داشتن	استفاده کردن
۱۷	در آمدن	اظهار داشتن	کاهش یافتن	افزایش دادن	استفاده کردن	اشاره کردن	بر خوردن	کار کردن
۱۸	خاطرنشان	بر خوردن	اشاره	کاهش	بر داشتن	قرار	اشاره	بر داشتن

	کردن		کردن	داشتن		دادن	کردن		کردن
۱۹	اشاره کردن	اشاره کردن	اظهار کردن	عمل کردن	صادر کردن	مصرف کردن	انجام دادن	اشاره کردن	برگزار کردن
۲۰	بر خوردن	خاطر نشان کردن	به نظر رسیدن	کاهش یافتن	در نظر گرفتن	استفاده کردن	پیدا کردن	صورت گرفتن	تصریح کردن
۲۱	پیدا کردن	پیدا کردن	در نظر گرفتن	بیان کردن	ایجاد کردن	تلاش کردن	ایجاد کردن	دوست داشتن	از دست دادن
۲۲	ایجاد کردن	عنوان کردن	اظهار داشتن	کمک کردن	باز پرسیدن	بیان کردن	عمل کردن	تصریح کردن	باز گشتن
۲۳	فکر کردن	افزایش یافتن	خاطر نشان کردن	جلوگیری کردن	عمل کردن	باز گشتن	در آمدن	به نظر رسیدن	قرار دادن
۲۴	عنوان کردن	ایجاد کردن	ایجاد کردن	اضافه کردن	نشان دادن	در یافتن	اضافه کردن	سعی کردن	تأکید کردن
۲۵	اضافه کردن	در نظر گرفتن	پیدا کردن	توصیه کردن	به نظر رسیدن	اضافه کردن	عنوان کردن	در گذشتن	کار کردن
۲۶	بیان کردن	فکر کردن	عنوان کردن	از دست دادن	بیان کردن	ایجاد کردن	فکر کردن	زندگی کردن	صعود کردن
۲۷	به نظر رسیدن	کمک کردن	در یافتن	ایجاد کردن	قرار داشتن	حمایت کردن	کمک کردن	آغاز کردن	کمک کردن
۲۸	افزایش یافتن	از دست دادن	افزایش دادن	مراجعه کردن	اضافه کردن	در آمدن	تشکیل دادن	نگاه کردن	اظهار داشتن
۲۹	باز گشتن	در یافتن	اضافه کردن	اظهار داشتن	پیدا کردن	پیدا کردن	به وجود آمدن	بر گشتن	تلاش کردن
۳۰	عمل کردن	ادامه داشتن	بر آوردن	پیدا کردن	در آمدن	فکر کردن	در نظر گرفتن	باز گشتن	صحبت کردن
۳۱	در نظر گرفتن	اضافه کردن	اقدام کردن	عنوان کردن	ارائه دادن	عنوان کردن	به نظر رسیدن	اضافه کردن	راه یافتن

۳۲	کمک کردن	مراجعه کردن	کمک کردن	بروز کردن	بر شمردن	به نظر رسیدن	بیان کردن	بازی کردن	آغاز کردن
۳۳	کار کردن	رخ دادن	پیش‌بینی کردن	خودداری کردن	اقدام کردن	ارائه کردن	باز گشتن	توضیح دادن	صورت گرفتن
۳۴	تلاش کردن	باز گشتن	اختصاص دادن	بر داشتن	ارائه کردن	کمک کردن	نشان کردن	ایجاد کردن	خاطر نشان کردن
۳۵	دریافت کردن	زندگی کردن	کاهش دادن	در آمدن	بر گشتن	شرکت کردن	افزایش یافتن	خاطر نشان کردن	دوست داشتن
۳۶	شرکت کردن	کار کردن	بر خوردن	بر خوردن	بر خوردن	رخ دادن	کار کردن	اجرا کردن	شکست خوردن
۳۷	آغاز کردن	تلاش کردن	تشکیل دادن	دریافت کردن	مراجعه کردن	آغاز کردن	شکل گرفتن	اظهار داشتن	پیدا کردن
۳۸	تشکیل دادن	به نظر رسیدن	کار کردن	در نظر گرفتن	وارد کردن	کار کردن	سعی کردن	در نظر گرفتن	در یافتن
۳۹	بر گشتن	تشکیل دادن	تأمین کردن	توضیح دادن	رسیدگی کردن	ادامه داشتن	ارائه دادن	دریافت کردن	سعی کردن
۴۰	از دست دادن	آغاز کردن	خریداری کردن	تشکیل دادن	سعی کردن	در نظر گرفتن	آغاز کردن	صحبت کردن	دریافت کردن
۴۱	رخ دادن	اقدام کردن	به شمار رفتن	اشاره کردن	نشان کردن	باقی ماندن	توضیح دادن	شرکت کردن	تمرین کردن
۴۲	دست یافتن	عمل کردن	ذکر کردن	به نظر رسیدن	به قتل رساندن	گزارش دادن	رخ دادن	نام بردن	اشاره کردن
۴۳	ادامه داشتن	بیان کردن	پرداخت کردن	پیشگیری کردن	اتفاق افتادن	تصمیم گرفتن	انتخاب کردن	عنوان کردن	بیان کردن
۴۴	سعی کردن	باز داشتن	مراجعه کردن	انجام گرفتن	تلاش کردن	دریافت کردن	حرکت کردن	انتخاب کردن	تصمیم گرفتن
۴۵	باز داشتن	کاهش	ارائه کردن	به وجود	کمک	تشکیل	شرکت	باقی	رخ دادن

		یافتن	آمدن	کردن	دادن	کردن	ماندن		
۴۶	باقی ماندن	شرکت کردن	توضیح دادن	تشخیص دادن	تعیین کردن	رد کردن	کمک کردن	کشتی گرفتن	
۴۷	ارائه کردن	باقی ماندن	وا گذاردن	از بین بردن	به وجود آمدن	معرفی کردن	دریافت کردن	برگزار کردن	در آمدن
۴۸	توضیح دادن	اختصاص دادن	صادر کردن	به شمار رفتن	قبول داشتن	دنبال کردن	معرفی کردن	بر خوردن	به نظر رسیدن
۴۹	به دست آوردن	انتخاب کردن	آغاز کردن	کار کردن	جلوگیری کردن	به دست آوردن	معرفی کردن	باقی ماندن	
۵۰	انتخاب کردن	تصمیم گرفتن	وارد کردن	رخ دادن	به سر بردن	از دست دادن	ارائه کردن	تلاش کردن	عنوان کردن

۵. بحث و نتیجه گیری

مقایسه‌ی ۵۰ فعل پیشوندی و مرکب نخست کلیه‌ی پیکره‌ها با ۵۰ فعل نخست هر یک از حوزه‌ها نشان داد که در حوزه‌های اجتماعی، سیاسی و علمی کم‌ترین تعداد افعال متفاوت وجود داشت؛ حوزه‌ی اجتماعی ۵ فعل، حوزه‌های سیاسی و علمی هر یک ۶ فعل. در حوزه‌ی فرهنگی، ۱۰ فعل متفاوت، در حوزه‌های بهداشتی و اقتصادی ۱۵ مورد، در حوزه‌ی حقوقی ۱۷ فعل متفاوت و در حوزه‌ی ورزشی، ۱۲ فعل متفاوت وجود داشت.

از مجموع کل ۳۰۱۰۰۴ جمله‌ی نمونه‌ی بررسی شده در کل پیکره‌ها، ۲۶۰۷۷۷ فعل پیشوندی و مرکب استخراج شد. ۱۱۹۲۱۲ فعل (۴۵/۷۱٪) از کل افعال، ۵۰ فعل نخست این فهرست هستند. از بررسی و مقایسه‌ی ۵۰ فعل نخست، می‌توان به پربسامدترین افعال اختصاصی هر حوزه دست یافت. افعالی به عنوان افعال عمومی لحاظ شدند که در کل پیکره‌ها فارغ از نوع حوزه‌های موضوعی، بیشترین کاربرد را داشتند منظور از افعال اختصاصی افعالی است که در هر یک از حوزه‌ها کاربرد بیشتری داشته‌اند. از سوی دیگر افعالی نیز وجود دارند که در بیش از یک حوزه، فعل اختصاصی محسوب می‌شوند. در حوزه‌های اجتماعی، سیاسی و علمی در ۵۰ فعل نخست به افعال اختصاصی دست نیافتیم. درواقع ۵۰ فعل نخست این حوزه‌ها افعال عمومی هستند. افعال «کاهش یافتن»، «برآوردن»، «پیش‌بینی کردن»،

«کاهش دادن»، «تامین کردن»، «خریداری کردن»، «پرداخت کردن»، «واگذاری»، «صادر کردن» و «وارد کردن» در حوزه‌ی اقتصادی، فعل «مصرف کردن»، «جلوگیری کردن»، «توصیه کردن»، «بروز کردن»، «خودداری کردن»، «پیش‌گیری کردن»، «به وجود آمدن»، «تشخیص دادن» و «از بین بردن» در حوزه‌ی بهداشت. افعال «صادر کردن»، «ارائه دادن»، «اقدام کردن»، «برخورد کردن»، «مراجعه کردن»، «وارد کردن»، «رسیدگی کردن»، «نشان کردن»، «به قتل رساندن»، «اتفاق افتادن»، «تعیین کردن»، «به وجود آمدن»، «قبول داشتن»، «جلوگیری کردن» و «به سر بردن» در حوزه‌ی حقوقی. افعال «دوست داشتن»، «درگذشتن»، «زندگی کردن»، «نگاه کردن»، «بازی کردن»، «اجرا کردن»، «صحبت کردن»، «نام بردن»، «برگزار کردن» و «معرفی کردن» در حوزه‌ی فرهنگی و افعال «بازی کردن»، «شکست دادن»، «کسب کردن»، «برگزار کردن»، «صعود کردن»، «صحبت کردن»، «راه یافتن»، «دوست داشتن»، «شکست خوردن»، «تمرین کردن» و «گشتی گرفتن» در حوزه‌ی ورزشی.

از منظری دیگر می‌توان به گستره‌ی کاربرد افعال مختلف در حوزه‌های هشت‌گانه نگاهی انداخت: حوزه‌ی حقوقی از ۸۶۴ فعل پیشوندی و مرکب مختلف، حوزه‌ی بهداشتی از ۷۶۱ فعل، حوزه‌ی ورزشی از ۱۱۹۸ فعل، حوزه‌ی اقتصادی از ۱۲۶۷ فعل، حوزه‌ی سیاسی از ۱۸۴۱ فعل، حوزه‌ی علمی از ۱۹۵۷ فعل، حوزه‌ی اجتماعی از ۲۰۶۰ فعل و حوزه‌ی فرهنگی از ۱۸۲۲ فعل مختلف پیشوندی و مرکب استفاده کرده‌اند. بنابراین، حوزه‌های اجتماعی، علمی و سیاسی از گستره‌ی وسیع‌تری از افعال پیشوندی و مرکب به نسبت بقیه‌ی حوزه‌ها، برای بیان مفاهیم خود بهره گرفته‌اند.

تعداد افعال یک کاربردی در حوزه‌های هشت‌گانه نیز قابل توجه است: حوزه‌ی علمی ۴۹۷ فعل، حوزه‌ی فرهنگی ۵۰۸ فعل، حوزه‌ی اجتماعی ۵۶۵ فعل، حوزه‌ی بهداشتی ۲۸۸ فعل، حوزه‌ی سیاسی ۴۵۳ فعل، حوزه‌ی ورزشی ۳۹۱ فعل، حوزه‌ی حقوقی ۳۳۰ فعل و حوزه‌ی اقتصادی ۳۷۲ فعل. در مجموع کل پیکره‌ها نیز، ۴۹۴ فعل پیشوندی و مرکب تنها یک بار به کار رفته‌اند. به نظر می‌رسد نتایج این پژوهش می‌تواند گامی در فراگیری آسانتر زبان فارسی بردارد؛ بدین صورت که زبان آموزان غیرفارسی زبان در هر مرحله از زبان آموزی تنها نیاز به فراگیری تعداد محدودی فعل هستند. همچنین در حوزه‌ی ترجمه‌ی متون نیز این طبقه‌بندی افعال می‌تواند به مترجمان در یافتن معادل‌های قابل قبول‌تر برای افعال کمک کند.

منابع

بیجن خان، م. (۱۳۸۳). نقش پیکره‌های زبانی در نوشتن دستور زبان: معرفی یک نرم‌افزار رایانه‌ای. *مجله‌ی زبان‌شناسی*. ۱۹ (۲)، ۶۷-۴۸. <http://noo.rs/Cz6Q>.

تسلیمی‌پور، ش. (۱۳۹۱). پردازش خودکار معنایی افعال مرکب فارسی. پایان‌نامه کارشناسی ارشد منتشر نشده، دانشگاه شیراز.

چاقری، آ. و شمس‌فرد، م. (۱۳۹۲). شناسایی افعال در جملات زبان فارسی، هجدهمین کنفرانس ملی سالانه انجمن کامپیوتر ایران. (صص ۲۲-۳۰) به همت حمید سربازی آزاد. *نشریه علمی انجمن کامپیوتر ایران*. ۱ (۳)، ۳۰-۲۲. <https://csi.org.ir/product>

چراغی، ز. و کریمی‌دوستان، غ. ح. (۱۳۹۲). طبقه‌بندی افعال زبان فارسی بر اساس ساخت رویدادی نمودی. *پژوهش‌های زبانی*، ۴ (۲)، ۶۰-۴۱. <https://doi.org/10.22059/jolr.2014.50267>

شیبانی‌ا قدم، ا. و غلامی، ر. (۱۳۹۳). رویکردی پیکره‌محور بر افعال وجهی شش متن قرن ششم. *پژوهش زبان و ادبیات فارسی*، ۳ (۳۴)، ۴۴-۲۱. <http://fa.journals.sid.ir/ViewPaper.aspx?id=247094>

عسکریان، ن. (۱۳۹۱). شناسایی خودکار افعال مرکب پارسی. پایان‌نامه کارشناسی ارشد منتشر نشده، دانشگاه شیراز.

مصطفوی‌فرد، ح. (۱۳۹۳). طبقه‌بندی علوم؛ چیستی، چرایی و چگونگی. *رهیافت*، ۲۴ (۵۶)، ۸۳-۱۰۸. [20.1001.1.10272690.1393.24.56.7.6](https://doi.org/10.1027/2690.1393.24.56.7.6)

مولودی، ا. س.، کوهستانی، م.، و رسولی، م. ص. (۱۳۹۰). فرهنگ ظرفیتی فعل: تلاشی در جهت آموزش زبان فارسی به غیرفارسی‌زبانان. *مجموعه مقالات نخستین همایش بین‌المللی چشم‌انداز، راهکارها و موانع گسترش زبان و ادب فارسی* (صص. ۹۱۳-۹۳۰)، به همت عباس‌علی وفایی. شورای گسترش زبان و ادبیات فارسی <https://www.academia.edu/1224377>

پیکره به‌روزشونده پارسی. (بی‌تا). پیکره به‌روزشونده پارسی. برگرفته از www.peikare.ir

AleAhmad, A., Amiri, H., Darrudi, E., Rahgozar, M., & Oroumchian, F. (2009). Hamshahri: A standard Persian text collection. *Knowledge-Based Systems*, 22(5), 382-387. <https://doi.org/10.1016/j.knosys.2009.05.002>

Bijankhan, M., Sheykhzadegan, J., Bahrani, M., & Ghayoomi, M. (2011). *Lessons from building a Persian written corpus: Peykare. Language Resources and Evaluation*, 45(2), 143–164. <https://doi.org/10.1007/s10579-010-9132-x>

Eghbalzadeh, H., Hosseini, B., Khadivi, S., & Khodabakhsh, A. (2012). *Persica: A Persian corpus for multipurpose text mining and natural language processing*. In 2012 Sixth International Symposium on Telecommunications (IST) (pp. 1207–1214). IEEE. <https://doi.org/10.1109/ISTEL.2012.6483172>

Gahl, S., Jurafsky, D., & Roland, D. (2004). Verb subcategorization frequencies: American English corpus data, methodological studies, and cross-corpus comparisons. *Behavior Research Methods, Instruments, and Computers*, 36(3), 432–443. <https://doi.org/10.3758/BF03195591>

Grabowski, E., & Mindt, D. (1995). *A corpus-based learning list of irregular verbs in English*. *ICAME Journal*, 19, 5–22. https://icame.info/icame_static/ij19/mindt.pdf

Gross, S., Hirschmanner, M., Krenn, B., Neubarth, F., & Zillich, M. (2018). *Action verb corpus*. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)* (pp. 2147–2151). European Language Resources Association (ELRA). <https://aclanthology.org/L18-1338>

Mohseni, M., Ghofrani, J., & Faili, H. (2018). *PersianP: A Persian text processing toolbox*. In A. Gelbukh (Ed.), *Computational linguistics and intelligent text processing: 17th International Conference, CICLing 2016, Konya, Turkey, April 3–9, 2016, revised selected papers, Part I* (pp. 75–87). Springer. https://doi.org/10.1007/978-3-319-75477-2_4

Rasooli, M. S., Moloodi, A., Kouhestani, M., & Minaei-Bidgoli, B. (2011). *A syntactic valency lexicon for Persian verbs: The first steps towards Persian dependency treebank*. In *Proceedings of the 5th Language & Technology Conference (LTC): Human Language Technologies as a Challenge for Computer Science and Linguistics*

(pp. 227–231). <https://www.cs.columbia.edu/~rasooli>

Rasooli, M. S., Faili, H., & Minaei-Bidgoli, B. (2011). *Unsupervised identification of Persian compound verbs*. In *Advances in artificial intelligence: 10th Mexican International Conference on Artificial Intelligence, MICAI 2011, Puebla, Mexico, November 26–December 4, 2011, Proceedings, Part I* (pp. 394–406). Springer. https://doi.org/10.1007/978-3-642-25324-9_34

Sarabi, Z., Mahyar, H., & Farhoodi, M. (2013). *ParsiPardaz: Persian language processing toolkit*. In *2013 3rd International eConference on Computer and Knowledge Engineering (ICCCKE)* (pp. 73–79). IEEE. <https://doi.org/10.1109/ICCCKE.2013.6682862>

Shamsfard, M., Jafari, H. S., & Ilbeygi, M. (2010). STeP-1: A set of fundamental tools for Persian text processing. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*. European Language Resources Association (ELRA). https://lrec.elra.info/proceedings/lrec2010/pdf/809_Paper.pdf

Stevenson, S., & Merlo, P. (1999). *Automatic verb classification using distributions of grammatical features*. In *Proceedings of the Ninth Conference of the European Chapter of the Association for Computational Linguistics* (pp. 45–52). Association for Computational Linguistics. <https://doi.org/10.3115/977035.977043>

Vincze, V., & Csirik, J. (2010). *Hungarian corpus of light verb constructions*. In *Proceedings of the 23rd International Conference on Computational Linguistics (COLING 2010)* (pp. 1110–1118). Coling 2010 Organizing Committee. <https://aclanthology.org/C10-1125>

Schulte im Walde, S. (1998). *Automatic semantic classification of verbs according to their alternation behaviour* [Diplomarbeit, University of Stuttgart]. https://www.ims.uni-stuttgart.de/~schulte/theses/diplomarbeit/fin_thesis_short2.pdf