



Studying Light Verbs in two Persian Specialized Corpora

Elham Alayiaboozar¹ 

Computational linguistic research group.
Information Science Research Department. Iranian
Research Institute for Information Science and
Technology (IranDoc). Tehran.Iran.

Abstract

One of the topics examined within the framework of corpus linguistics is the compound verb. Compound verbs consist of two parts: a "non-verbal element" and a "light verb." In this study, the frequency of light verbs was extracted from two specialized corpora, "PAZHUHESHNAME" and "PEKA" (the corpus of IranDoc books), and the frequency of the top three light verbs in these two corpora was compared with their frequency in a general corpus, namely the "Persian Text Corpus."

In these two specialized corpora, which together contain nearly eight million words, the light verbs "zadan" (to hit), "dashtan" (to have), "kardan" (to do), "sepordan" (to give), "gereftan" (to take), "amadan" (to come), "dadān" (to give), "oftadan" (to fall), "khordan" (to eat), "keshidan" (to pull), "avardan" (to bring), "nemoudan" (to do), "raftan" (to go), "bordān" (to carry), and "andakhtan" (to throw) were analyzed. Among these, the light verbs "kardan" with a frequency of 5848 and "dadān" with a frequency of 5037 ranked first and second in terms of frequency. The frequency of these two light verbs is also high in the general corpus, where they rank among the top. The light verb "gereftan" with a frequency of 3246 ranks third in the specialized corpora. However, in the general corpus, the light verb "gereftan" ranks after the light verbs "kardan," "dadān," "dashtan," and "namoudan" in terms of frequency.

1. Alayi@irandoc.ac.ir (Corresponding Author)

How to cite: Alayiaboozar, E. (2024). Studying Light Verbs in two Persian Specialized Corpora. *Language and Linguistics*, 20(39), 97- 122. doi: 10.30465/LSI.2025.49517.1772

After extracting and examining the frequency of light verbs, compound verbs formed with these three light verbs (*kardan*, *dadān*, and *gereftan*) were also analyzed in terms of the likelihood of a gap between the non-verbal element and the light verb within a defined context (five tokens before and five tokens after the target light verb). The analysis of the behavior of high-frequency compound verbs in specialized corpora showed that authors of specialized texts, mainly researchers, students, and professors, are less inclined to place a gap between the non-verbal element and the light verb. Therefore, in most compound verbs, the non-verbal element and the light verb appear together without any syntactic category or group intervening. In contrast, in general texts, the likelihood of authors inserting syntactic groups or various categories between the non-verbal element and the light verb is higher than in specialized texts.

Keywords: compound verb; corpus; frequency of light verbs; light verb; specialized corpus

1. Introduction

Nowadays, many linguistic studies are conducted using corpora. One of the topics investigated in corpus linguistics is “*compound verbs*”, which refer to verbs composed of two independent words; the first word is a noun, adjective, or adverb and the second word is a verb. Examples in Persian include *tamiz kardan* (cleaning) and *harf zadan* (talking). The first part is called *NON-VERBAL ELEMENT* or *HAMRAH* and the second part is called *LIGHT VERB* or *HAMKARD*. The distinctive feature of the light verb is that it loses its original meaning and it usually acts like a morphological element that gives the compound its verbal identity. In corpus-based studies of compound verbs, frequency is an important parameter. As an example, the study of compound verbs in a general corpus (*PERSIAN TEXT CORPUS*) shows that among the compound verbs that are formed with the light verbs *zadan* (hit), *dāshtan* (have), *kardan* (do), *sepordan* (give), *gereftan* (take), *āmadan* (come), *dādan* (give), *oftādan* (fall), *khordan* (eat), *keshidan* (pull), *āvardan* (bring), *nemoodan* (do), *raftan* (go), *bordan* (take away) and *andākhtan* (throw), the frequency of compound verbs that are formed with the light verbs *kardan*, *dāshtan* and *dādan* is higher than other compound verbs. There are numerous corpus-based studies on compound verbs in Persian which are based on general corpora. Since corpus-based research relying on specialized corpora has been rarely done on compound verbs, this

study tries to examine the behavior of compound verbs, whose frequency is high in specialized corpora.

Research Question(s)

1. Is there a significant difference between the frequency of light verbs in general and specialized corpora?
2. What is the likelihood of creating a distance between the verbal and non-verbal elements in compound verbs formed with high-frequency light verbs in the specialized corpus?

2. Literature Review

Compound verbs have been examined from structural and computational perspectives. Structural studies are based on written or spoken data and rely on the linguistic intuition of the researcher or linguistic information collected from Persian speakers in the written form. This method is considered traditional. From a methodological perspective, both Iranian and non-Iranian grammarians have paid special attention to the category of compound verbs in Persian describing and classifying them in line with formal or semantic criteria to distinguish them from simple verbs (Dabirmoghadam, 1995; Karimi, 1997; Khanlari, 1987; Moradi & Karimidoostan, 2013; Sharif & Shadmehr, 2022; Tabatabaei, 2005; Zafarabadi & Rahemeyan, 2020).

Most traditional studies have focused on the semantic and syntactic behavior of compound verbs. Each of these approaches has been subject to numerous critiques accompanied by counterexamples. Corpus-based studies and computational approaches to language processing have not overlooked the complex area of examining and identifying compound verbs and their syntactic and semantic behaviors (Alayiaboozar & Bijankhan, 2012; Asgarian, 2012; Cheraghi & Karimi Doostan, 2013; Eshaghi & Karimi Doustan, 2021; Rasooli et al, 2011; Taslimipoor, 2012; Vahdatzade, 2012).

3. Methodology

In this study, first the frequency of the light verbs *zadan*, *dāshtan*, *kardan*, *sepordan*, *gereftan*, *āmadan*, *dādan*, *oftādan*, *khordan*, *keshidan*, *āvardan*, *nemoodan*, *raftan*, *bordan* and *andākhtan* were extracted from two Persian specialized corpora; *PAZHUHESHNAME*, which is made out of the published articles of the Iranian journal of information processing and management and *PEKA*, which is made out of the digital books of IranDoc. These corpora contain specialized and interdisciplinary texts. The result is shown in Table 1 below.

Table 1

The Frequency of Light Verbs in Two Persian Specialized Corpora

Persian light verbs	PAZHUHESHNAME	PEKA	The sum
kardan	3427	2421	5848
dādan	4024	1013	5037
gereftan	2697	549	3246
nemoodan	1587	789	2376
dāshtan	1406	775	2181
āmadan	576	171	747
āvardan	339	242	481
bordan	272	159	431
raftan	106	123	229
zadan	55	42	97
keshidan	16	25	41
oftādan	18	18	36
khordan	17	10	27
andākhtan	7	11	18
sepordan	3 cases, none of which are used in the structure of compound verbs	1 case (be faramooshi sepordan (oblivion))	4

Persian compound verbs are divided into two categories: separable and non-separable. Compound verbs whose components are separable, can have a word or words between the verbal element and the non-verbal element (Barjesteh, 1983). In this study, the possibility of putting distance between the non-verbal element and the light verb has also been taken into account.

4. Results

In these two specialized corpora, among the light verbs, the frequency of the two light verbs of *kardan* and *dādan* is high, as expected and compared to

their frequency in a general corpus. But the light verb *gereftan* had a higher frequency in specialized corpora compared to the general corpus; i.e. in a general corpus the frequency of the light verbs *kardan*, *dādan*, *dāshtan*, and *nemoodan* is higher than *gereftan*, but in specialized corpora, the frequency of the light verb *gereftan* is higher than *dāshtan* and *nemoodan*. After extracting the frequency of light verbs, compound verbs that can be formed with these three light verbs were also studied in terms of the possibility of distance between the non-verbal element and the light verb in the specified context. It seems that the authors of the specialized texts are less inclined to put distance between the non-verbal element and the light verb. Therefore, in most compound verbs, the non-verbal element and the light verb are placed next to each other and no category or syntactic group is placed between them.

5. Conclusion


In this study, the frequency of the light verbs *zadan*, *dāshtan*, *kardan*, *sepordan*, *gereftan*, *āmadan*, *dādan*, *oftādan*, *khordan*, *keshidan*, *āvardan*, *nemoodan*, *raftan*, *bordan*, and *andākhtan* was extracted from two specialized corpora, namely *PAZHUHESHNAME* and *PEKA*. It was expected that the frequency of light verbs in these two specialized corpora would not significantly differ from their frequency in *PERSIAN TEXT CORPUS*, as a general corpus. In the *PERSIAN TEXT CORPUS*, the frequency of compound verbs formed with the three light verbs *kardan*, *dādan*, and *dāshtan* was higher than others. In the specialized corpora encompassing 20,799 instances of light verbs, *kardan* with a frequency of 5,848 and *dādan* with a frequency of 5,037 were ranked first and second. The frequency of these two light verbs is also high in the general corpus. The light verb *gereftan* with a frequency of 3,246 ranks third in the specialized corpora. In other words, about 28% of the light verb *kardan*, 24% of the light verb *dādan*, and 15% of the light verb *gereftan* are used in specialized corpora. However, in the *PERSIAN TEXT CORPUS*, the light verb *gereftan* ranks after the light verbs *kardan*, *dādan*, *dāshtan*, and *nemoodan* in terms of frequency. It was also looked at how likely it was that there would be a gap between the nonverbal element and the light verb in the given context when these three light verbs were used to make compound verbs. It was revealed that authors of specialized texts are less inclined to place a gap between the non-verbal element and the light verb. In general texts, on the other hand, authors are more likely than in specialized texts to

put syntactic groups or other categories between the non-verbal element and the light verb.



بررسی افعال سبک در دو پیکره تخصصی فارسی

استادیار گروه پژوهشی زبان‌شناسی رایانشی، پژوهشکده علوم اطلاعات، پژوهشگاه علوم و فناوری اطلاعات ایران (ایراندک)، تهران، ایران

الهام علائی ابوذری 

چکیده

یکی از مقوله‌هایی که در چارچوب زبان‌شناسی پیکره‌ای مورد بررسی قرار گرفته، فعل مرکب است. افعال مرکب از دو بخش «جزء غیرفعلی» و «فعل سبک» تشکیل شده‌اند. در این پژوهش فراوانی افعال سبک از دو پیکره تخصصی «پیکره پژوهش‌نامه» و «پیکا» (پیکره کتاب‌های ایراندک) استخراج شده است و فراوانی سه فعل سبک اول در این دو پیکره با فراوانی افعال سبک در یک پیکره عمومی، یعنی «پیکره متنی فارسی» مقایسه شده است. در این دو پیکره تخصصی، که در مجموع نزدیک به هشت میلیون واژه در این پیکره‌ها موجود است، از میان افعال سبک «زدن»، «داشتن»، «کردن»، «سپردن»، «گرفتن»، «آمدن»، «دادن»، «افتادن»، «خوردن»، «کشیدن»، «آوردن»، «نمودن»، «رفتن»، «بردن» و «انداختن»، دو فعل سبک «کردن» با فراوانی ۵۸۴۸ و فعل «دادن» با فراوانی ۵۰۳۷، در رتبه اول و دوم از نظر فراوانی قرار دارند؛ فراوانی این دو فعل سبک در پیکره عمومی نیز بالا است و در رتبه‌های نخست قرار دارند. فعل سبک «گرفتن» با فراوانی ۳۲۴۶ در پیکره‌های تخصصی در رتبه سوم قرار دارد. حال آنکه در پیکره عمومی، فعل سبک «گرفتن»، از نظر فراوانی بعد از افعال سبک «کردن»، «دادن»، «داشتن» و «نمودن» قرار دارد. پس از استخراج فراوانی افعال سبک و بررسی آن‌ها، افعال مرکبی که با این سه فعل سبک (کردن، دادن و گرفتن) ساخته می‌شوند از نظر احتمال فاصله میان جزء غیرفعلی و فعل سبک در بافت تعریف‌شده (پنج نگاشت قبل و پنج نگاشت بعد از فعل سبک موردنظر) نیز مورد بررسی کلی قرار گرفتند. بررسی رفتار افعال مرکب پربسامد در پیکره‌های تخصصی نشان داد نویسندگان متون تخصصی، عمدتاً پژوهشگران، دانشجویان و استادان، کمتر تمایل دارند بین جزء غیرفعلی و فعل سبک فاصله بیاندازند و بنابراین، در بیشتر افعال مرکب جزء غیرفعلی و فعل سبک در کنار هم قرار می‌گیرند و مقوله و گروه نحوی میان آن‌ها قرار نمی‌گیرد. در حالیکه در متون عمومی احتمال اینکه نویسندگان گروه/گروه‌های نحوی و مقولات گوناگون میان جزء غیرفعلی و فعلی قرار دهند بیشتر از متون تخصصی است.

کلیدواژه: فعل مرکب، پیکره، افعال سبک، پیکره تخصصی، فراوانی افعال سبک.

۱- مقدمه

امروزه بسیاری از پژوهش‌های زبانی به کمک پیکره‌های زبانی انجام می‌شود. پیکره^۱، مجموعه‌ای نظام‌مند و رایانه‌ای شده از داده‌های زبانی است که برای پژوهش‌های زبان‌شناسی مورد استفاده قرار می‌گیرد (سینکلر^۲، ۲۰۰۴) و زبان‌شناسی پیکره‌ای نیز روشی است آماری در تجزیه و تحلیل داده‌های پیکره که برگرفته از داده‌های واقعی زبان است. زبان‌شناسی پیکره‌ای را می‌توان گذار از رویکردهای ذهن‌گرایانه به پژوهش‌های زبانی دانست. در این میان، هم از روش‌های کمی و هم از روش‌های کیفی برای تحلیل زبان بهره می‌برد و رایانه را برای انجام تحلیل‌های پیچیده به کار می‌گیرد. آنچه زبان‌شناسی پیکره‌ای را به کار زبان می‌آورد، توانمندی آن در بررسی کاربرد واقعی زبان است تا پژوهشگر از زیاده‌روی در نظریه‌پردازی دور نگه داشته شود. همین توانمندی نیز واکاوی پایگاه‌های بزرگ داده‌های زبانی را به زبان‌شناسی پیکره‌ای می‌سپارد (کوهستانی، ۱۳۸۹).

یکی از مقوله‌هایی که در چارچوب زبان‌شناسی پیکره‌ای می‌توان مورد بررسی قرار داد، فعل است. مقوله^۳ دستوری فعل در دستور زبان به دلیل باز بودن فهرست واژه‌های متعلق به آن، گوناگونی ساختاری اعضای این مقوله^۳ دستوری و ویژگی‌های معنایی که هر ساختار با خود به همراه دارد، از بحث‌برانگیزترین مباحث دستوری بوده است. در این میان بررسی افعال مرکب جایگاه ویژه‌ای در پژوهش‌های زبانی داشته است. افعال مرکب^۳ به افعالی اطلاق می‌شود که از دو واژه مستقل ترکیب یافته‌اند؛ واژه اول اسم یا صفت یا قید و واژه دوم فعل است، مانند «تمیز کردن»، «حرف زدن»، «بهانه گرفتن» و غیره. بخش اول، «همراه» و بخش دوم «همکرد» نامیده می‌شود (طباطبایی، ۱۳۸۴). البته در بیشتر متون مربوطه بخش اول، با عنوان «جزء غیرفعلی^۴» و بخش دوم با عنوان «فعل سبک^۵» به کار رفته است.

اصطلاح فعل سبک هم به دسته‌ای از افعال اطلاق می‌شود که بخشی از بار معنایی خود را در روند فرایندی به نام سبک‌سازی از دست داده‌اند، اما به دلیل وجود نشانگرهای شخص و شمار و زمان در طبقه^۶ افعال قرار می‌گیرند. اصطلاح فعل سبک را اولین بار یسپرسن^۶ (۱۹۶۵)، نقل از اسحاقی و کریمی‌دوستان، (۱۴۰۰) برای اشاره به این افعال در زبان انگلیسی به کاربرد

1. corpus
2. Sinclair
3. compound verb
4. non-verbal element
5. light verb
6. O. Jespersen

و پس از آن مطالعه گریمشاوا^۱ و مستر^۲ (۱۹۸۸، نقل از اسحاقی و کریمی دوستان، ۱۴۰۰) آغازگر بررسی‌های ساخت فعل سبک در زبان‌های مختلف و در زبان‌شناسی نوین شد. در زبان فارسی ساخت‌های فعل سبک به گونه‌ای زایا به کار برده شده‌اند، تا بدان جا که مفاهیم فعلی جدیدی که به زبان راه می‌یابند نیز از این الگو تبعیت می‌کنند، مانند «سرچ کردن» و «بلاک کردن» (اسحاقی و کریمی دوستان، ۱۴۰۰).

در پژوهش‌های پیکره-بنیاد مربوط به افعال مرکب، فراوانی پارامتری مهم برای کشف الگوهای متناوب و بالفعل است، بنابراین، می‌توان توزیع فراوانی الگوی افعال مرکب در پیکره‌های متنی زبان فارسی را مورد بررسی قرار داد. به عنوان نمونه، بررسی افعال مرکب در یک پیکره عمومی، مانند پیکره متنی فارسی نشان می‌دهد از میان افعال مرکبی که با افعال سبک «زدن»، «داشتن»، «کردن»، «سپردن»، «گرفتن»، «آمدن»، «دادن»، «افتادن»، «خوردن»، «کشیدن»، «آوردن»، «نمودن»، «رفتن»، «بردن»، و «انداختن» ساخته می‌شوند، فراوانی افعال مرکبی که با افعال سبک «کردن»، «داشتن» و «دادن» ساخته می‌شوند، بیشتر از سایر افعال مرکب است (علائی ابوذر و بی‌جن‌خان، ۱۳۹۱).

در پژوهش‌های گوناگون، رفتار افعال مرکب در پیکره‌های عمومی مورد بررسی قرار گرفته است، اما تا کنون از پیکره‌های تخصصی برای این منظور استفاده نشده است. در صورتیکه این مسئله مهم است بدانیم در متون تخصصی چه میزان از افعال مرکب استفاده می‌شود. محتوای پیکره‌های استفاده شده در این پژوهش عمومی نیست و حاوی متون تخصصی و میان‌رشته‌ای است. در این پژوهش سعی می‌شود افعال مرکبی که فراوانی بالایی در پیکره‌های عمومی دارند در پیکره‌های تخصصی نیز مورد بررسی قرار گیرند. بنابراین، در این پژوهش سه فعل سبک «کردن»، «دادن» و «گرفتن» در دو پیکره تخصصی پیکره پژوهش‌نامه و پیکا مورد بررسی قرار می‌گیرند. پیکره پژوهش‌نامه، از متون مقاله‌های پژوهش‌نامه پردازش و مدیریت اطلاعات ساخته شده است و پیکا از متون کتاب‌های دیجیتال ایرانداک ساخته شده است. هر کدام به ترتیب، شامل بیش از چهار میلیون و ۷۸۰ هزار واژه و سه میلیون و ۳۲۹ هزار واژه هستند.

هدف از بررسی افعال سبک در پیکره‌های تخصصی، مقایسه میزان استفاده از افعال سبک پربسامد در متون تخصصی در مقایسه با متون عمومی است و پرسش اصلی این پژوهش این است که آیا تفاوت معناداری در نوع افعال مرکبی که در متون تخصصی وجود دارند با افعال مرکبی که در متون عمومی مورد استفاده قرار گرفته‌اند وجود دارد؟ به عبارت دیگر، آیا نویسندگان متون تخصصی به اعتبار اینکه متن تخصصی می‌نویسند از افعال مرکب خاصی

1. J. Grimshaw
2. A. Mester

هنگام نگارش استفاده می‌کنند تا گونه رسمی‌تر زبان را نمایش دهند یا تفاوت معناداری در این زمینه وجود ندارد؟ برای پاسخ به این سوال ابتدا فراوانی همه افعال سبک به‌کاررفته در پیکره‌های تخصصی پژوهش‌نامه و پکا مورد بررسی قرار گرفته است و این فراوانی با فراوانی افعال سبک در پیکره متنی فارسی، که یک پیکره عمومی است مقایسه شده است. مراحل انتخاب این سه فعل سبک در بخش مربوطه توضیح داده شده است. علاوه بر بررسی فراوانی افعال مرکبی که با این سه فعل سبک ساخته می‌شوند، امکان ایجاد فاصله میان جزء غیرفعلی و فعل سبک نیز مورد بررسی قرار گرفته است.

۲. پیشینه پژوهش

افعال مرکب از جنبه‌های گوناگونی مورد بررسی قرار گرفته‌اند. بسیاری از پژوهش‌ها روی تشخیص افعال مرکب متمرکز شده‌اند که خود به دو دسته ساختاری و ماشینی تقسیم می‌شوند. منظور از ساختاری، پژوهش‌هایی است که مبتنی بر داده‌های نوشتاری یا گفتاری و متکی بر شم زبانی پژوهشگر (که گویشور زبان فارسی است) یا اطلاعات زبانی جمع‌آوری شده از گویشوران زبان فارسی به صورت مکتوب است؛ این روش بررسی به اصطلاح، روش سنتی بررسی زبان محسوب می‌شود. از نظر روش‌شناسی به اصطلاح سنتی، دست‌نویسان و زبان‌شناسان ایرانی و غیرایرانی توجه ویژه‌ای به مقوله فعل مرکب در زبان فارسی داشته‌اند و هر یک ضمن اینکه آن را توصیف و طبقه‌بندی نموده‌اند، معیاری نیز برای متمایز ساختن آن از فعل ساده ارائه داده‌اند که برخی از آن‌ها صوری و برخی معنایی بوده است. برخی از این پژوهش‌ها شامل موارد زیر می‌شود:

خانلری (۱۳۶۶) اصطلاح فعل مرکب را به فعلی اطلاق می‌کند که از دو کلمه مستقل ترکیب شده است: کلمه اول اسم یا صفت است و تغییر نمی‌کند (صرف نمی‌شود) و کلمه دوم فعلی است که صرف می‌شود که وی آن را «همکرد» می‌نامد، و از مجموع این دو قسمت معنی واحدی دریافت می‌شود. او دو معیار برای تعیین فعل مرکب برمی‌شمرد: اول اینکه همکرد در فعل مرکب معنی اصلی را از دست داده و معنی ثانویه پیدا می‌کند و یا فقط جزء صرفی می‌شود. دومین معیار خانلری تفاوتی است که در صرف بعضی زمان‌های فعل مرکب و فعل ساده وجود دارد. در عبارت «باید عجله کنیم» پیشوند «ب-» بر سر ساخت التزامی نمی‌آید، یعنی نمی‌گوییم «باید عجله بکنیم»^۱، زیرا فعل مرکب است. طباطبائی (۱۳۸۴) معیارهایی برای تشخیص فعل مرکب ارائه می‌کند که شامل این موارد می‌شود: فعل مرکب یک واحد

۱- البته به نظر نگارنده، در مورد همه افعال مرکب نمی‌توان این ادعا را کرد، چرا که می‌توان گفت «باید زنگ بزنیم».

معنایی است و جزء فعلی آن از محتوای معنایی خود تهی شده است و بخش اعظم معنا را جزء غیرفعلی حمل می‌کند، مانند «جنگ کردن».

معیار دوم تشخیص فعل مرکب این است که اگر عبارتی از یک اسم و یک فعل متعدی ساخته شده باشد و این دو در مجموع به یک مفعول «رایبی» (مفعول همراه با «را») احتیاج داشته باشند، با فعل مرکب سروکار داریم، مانند «اتاق را جارو کردم». معیار سومی که وی ارائه می‌دهد این است که اگر عبارتی از یک فعل متعدی و اسم ساخته شده باشد و این دو در مجموع همچون فعل لازم عمل کنند، یعنی به مفعول نیاز نداشته باشند، با فعل مرکب سروکار داریم، مانند «رنج کشیدن» و «دوش گرفتن».

به عقیده دبیرمقدم (۱۳۷۴)، دو نوع فعل مرکب در زبان فارسی وجود دارد: فعل مرکب حاصل از ترکیب و فعل مرکبی که با انضمام ساخته می‌شود. ترکیب از اجتماع فعل کمکی با صفت یا اسم یا گروه حرف اضافه‌ای یا قید و یا از اسم مفعول با فعل کمکی مجهول‌ساز ساخته می‌شود. صفت با فعل «بودن» و «شدن» و «کردن» می‌تواند ترکیب شود و فهرست بازی از فعل‌های مرکب مانند «دلخور بودن» و «دلخور کردن» و «دلخور شدن» را بسازد.

تعداد فعل‌های کمکی که با اسم ترکیب می‌شوند و فعل مرکب می‌سازند هفت عدد است که عبارت‌اند از «کردن»، «زدن»، «دادن»، «گرفتن»، «کشیدن»، «داشتن» و «خوردن». تعداد فعل‌های مرکبی که با این فعل‌های کمکی ساخته شده‌اند و دبیر مقدم آن‌ها را استخراج کرده، در مجموع ۱۲۹۹ فعل مرکب است. این افعال یا متعدی‌اند یا لازم. در مورد انضمام، دبیر مقدم دو نوع انضمام قائل می‌شود: انضمام مفعول صریح و انضمام گروه حرف‌اضافه‌ای.

کریمی (۱۹۹۷) معتقد است به نظر می‌رسد در زبان فارسی شکل‌گیری افعال مرکب جایگزین شکل‌گیری افعال ساده شده است. استفاده از فعل مرکب به جای افعال ساده حتی در مورد کلمات قرضی هم دیده می‌شود، مانند «تلفن کردن»، «تایپ کردن». او افعال سبک زبان فارسی را فهرست می‌کند، این افعال شامل: «زدن»، «داشتن»، «کردن»، «سپردن»، «گرفتن»، «آمدن»، «دادن»، «فتادن»، «خوردن»، «کشیدن»، «آوردن»، «نمودن»، «رفتن»، «بردن» و «انداختن» است.

نکته قابل توجه این است که کریمی (۱۹۹۷) در فهرستی که از افعال سبک ارائه می‌دهد «شدن» را نیز لحاظ می‌کند، اما افعالی که با این فعل سبک ساخته می‌شوند معادل معلوم نیز دارند که با فعل سبک «کردن» ساخته می‌شوند. شاید کریمی «شدن» را به این علت در فهرست افعال سبک قرار داده است که معتقد به وجود ساخت مجهول در زبان فارسی نیست. وی همچنین عنصر غیرفعلی که همراه فعل سبک می‌آید در طبقات مختلفی قرار می‌دهد، مانند اسم (دعوت کردن، کتک زدن، شکست دادن)، صفت (تمیز کردن، سبک کردن، کم

کردن)، قید (بیرون کردن، پس دادن، پایین آوردن)، گروه حرف اضافه‌ای (از دست دادن، از یاد بردن، به نظر آمدن). درباره معنی جزء فعلی (فعل سبک) و غیرفعلی افعال مرکب تاکنون تحلیل‌های مختلفی صورت گرفته است. اغلب پژوهشگران کوشیده‌اند معنی فعل سبک را در یک یا چند ویژگی خلاصه کنند.

مرادی و کریمی‌دوستان (۱۳۹۲) ماهیت افعال مرکب زبان فارسی را در چارچوب نظریه نمود معنایی-واژی لیبر بررسی می‌کنند. این نظریه با استفاده از هفت مشخصه معنایی و یک اصل به نام اصل هم‌نمایگی، به بررسی نقش معنایی عناصر واژی و عملکرد آن‌ها در فرایندهای ترکیب، اشتقاق و تغییر مقوله می‌پردازد و برای هر عنصر واژی، حتی وندها، یک اسکلت معنایی و یک بدنه معنایی در نظر می‌گیرد. اسکلت از یک یا چند مشخصه معنایی و یک یا چند موضوع ساخته شده است و بدنه، خود از دو لایه تشکیل شده است. در پژوهش ایشان ادعا می‌شود که فعل همه جمله‌ها در اصل ساده بوده است، اما به دلیل اینکه بدنه معنایی هم‌کرد تغییر نموده، موضوع رویدادی خود را از دست داده و به تنهایی توانایی نشان دادن محمول را نداشته است و با حفظ ساخت موضوعی، یکی از موضوع‌هایش با موضوع یکی از موضوع‌های محمول، هم‌نمایه شده و یک واحد معنایی مرکب تشکیل داده است. مرادی و کریمی‌دوستان نشان می‌دهند که فعل مرکب از لحاظ نحوی، یک ترکیب وابسته و از لحاظ معنایی، درون‌مرکز است که در آن هم‌کرد، هسته نحوی و معنایی و پیش‌فعل وابسته و در ساخت موضوعی فعل مرکب است.

شریف و شادمهر (۲۰۲۲) پس از نقد آرای مطرح‌شده درباره معنی فعل سبک، فرضیه‌ای درباره تشکیل فعل مرکب ارائه می‌دهند که در آن همه افعال مرکب به طور مستقیم یا غیرمستقیم ریشه در فرایند انضمام دارند. این فرایند با جمله‌ای کامل آغاز و با انضمام دو عنصر از آن و حذف سایر عناصر ختم می‌شود. تشکیل هر فعل مرکب، منجر به شکل‌گیری طرحواره ساختی $X + LVO$ می‌شود که امکان ساخت افعال مرکب جدیدی با همان فعل سبک و در همان حوزه معنایی را فراهم می‌آورد. بر این اساس، معنی هم‌کرد در حالت انضمام مستقیم، همان معنی سنگین آن و در حالت انضمام غیرمستقیم، منحل در معنی ساختی طرحواره می‌شود. معنی فعل سبک در هر یک از این دو حالت می‌تواند متفاوت باشد و هر یک از این معانی برگرفته یا مشابه یکی از معانی سنگین آن است. حالت سوم نیز هست که در آن فعل سبک، نقش فعل‌سازی را ایفا و تصریف فعلی را برای یک عنصر غیرفعلی ممکن می‌کند. بدین ترتیب افعال سبک روی طیفی قرار می‌گیرند که از موارد کاملاً طرحواره و فاقد معنی تماتیک تا معنی سنگین آن‌ها را در برمی‌گیرد.

ظفرآبادی و رحیمیان (۱۳۹۹) معتقدند بررسی در زمانی میزان کاربرد فعل سبک فارسی در دوره‌های مختلف می‌تواند ظهور و شکل‌گیری فعل سبک زبان فارسی را نشان دهد. به این

منظور در این پژوهش، آن‌ها ۱۵۰۰۰ فعل ساده و مرکب را از متون روزنامه‌ای استخراج کرده‌اند و سپس پنج فعل سبک «داشتن»، «آمدن»، «آوردن»، «گرفتن» و «دیدن» که از افعال سبک فارسی به حساب می‌آیند، را در طول یک قرن (۱۲۲۰-۱۳۲۰) در پنج دوره مورد بررسی قرار داده‌اند. با تحلیل افعال مرکب به دست آمده در هر دوره، میزان افعال هم به لحاظ «نوع» و هم به لحاظ «رخداد» مشخص گردیده است. این پژوهش نشان می‌دهد که فعل «داشتن» فعل سبک اصلی در هر پنج دوره مورد بررسی بوده است. همچنین فعل «دیدن» در سه دوره آخر کاربردی از آن به عنوان فعل سبک مشاهده نشده است؛ در صورتیکه در دو دوره اول رواج داشته است، که نشان می‌دهد در دوره‌های اولیه به عنوان فعل سبک رواج داشته ولی به مرور زمان کاربرد فعل واژگانی پیدا کرده است. همچنین در بسامد فعل سبک در دوره‌های پنج‌گانه مورد بررسی نوساناتی نیز مشاهده می‌شود.

عمده پژوهش‌های سنتی انجام شده در این حوزه از گذشته تا کنون، روی رفتار معنایی و نحوی افعال مرکب متمرکز بوده‌اند. در مورد هر کدام از این رویکردها نقدهای بسیاری همراه با مثال‌های نقض ارائه شده است. بررسی رویکردهای تشخیص افعال مرکب فارسی و نقدهایی که بر آن‌ها وارد است، در حوزه پژوهش حاضر نیست. بررسی‌های پیکره-بنیاد و رویکردهای رایانشی پردازش زبان از حوزه پیچیده بررسی افعال مرکب و تشخیص آن‌ها و بررسی رفتاری نحوی و معنایی آن‌ها غافل نبوده‌اند و پژوهش‌های پیکره-بنیاد و ماشینی در مورد افعال مرکب انجام شده است که از میان آن‌ها می‌توان به موارد زیر اشاره کرد:

وحدت‌زاده (۱۳۹۱) ابزار تشخیص و معیار تشخیص جدیدی برای شناخت فعل مرکب در قالب نظریه زبان‌شناسی پیکره‌ای ارائه داده است. وی بر این اساس ۱۴ فعل گروهی که در کتاب‌های درسی برای شناخت فعل مرکب به عنوان نمونه آورده شده‌اند را مبنای مطالعه قرار داده است و همه نمادهای صورت‌های مختلف صرفی این ۱۴ فعل گروهی از یک پیکره زبانی استخراج شده است و فاصله جزء غیرفعلی از جزء فعلی شمارش و میانگین و واریانس این فاصله نیز محاسبه شده است.

رسولی و همکاران (۲۰۱۱) معتقدند یکی از ویژگی‌های افعال مرکب در فارسی این است که به لحاظ نحوی انعطاف‌پذیر هستند؛ به این معنا که محدودیتی در مورد فاصله بین فعل سبک و عنصر غیرفعلی وجود ندارد. علاوه بر این، عنصر غیرفعلی می‌تواند صرف شود. چنین ویژگی‌هایی تشخیص ماشینی افعال مرکب در فارسی را دشوار می‌کند. آن‌ها دو روش مختلف بدون نظارت^۱ برای تشخیص خودکار افعال مرکب در فارسی پیشنهاد می‌کنند؛ در روش اول با

بسط مفهوم سنجش PMI (اطلاعات متقابل نقطه‌به‌نقطه^۱)، روش خودگردانی^۲ را به کار می‌برند و در رویکرد دوم الگوریتم خوشه‌بندی k-means را به کار می‌برند. نتایج به‌کارگیری دو رویکرد نشان می‌دهد رویکردهای پیشنهادی نتایجی به مراتب بهتر از به‌کارگیری رویکردهای پایه‌ای دارند که سنجش PMI را به عنوان معیار خود به کار می‌برند.

تسلیمی‌پور (۱۳۹۱) مجموعه‌ای متداول از افعال مرکب فارسی که شکل اسم+ فعل دارند را مورد بررسی قرار داده است و معیاری برای تعیین میزان قابل قبول بودن عبارات به عنوان افعال مرکب ارائه داده است که مبنی بر کلاس معنایی است. در این پژوهش در تعیین گروه‌های معنایی کلمات از مدل فضای برداری استفاده شده است و معنای کلمات در فضای کلمات به صورت بردار نشان داده شده است. با مشارکت دادن گروه‌های معنایی به‌دست‌آمده از نمایش برداری در معیار وابسته به کلاس معنایی، نتایج بهتری در تشخیص افعال مرکب فارسی به‌دست آمده است.

عسگریان (۱۳۹۱) برای شناسایی افعال مرکب از معیارهای استاندارد همبستگی، معیار پایستگی و ترجمه‌ای برای آموزش دسته‌بندی‌کننده SVM^۳ استفاده کرده است. معیارهای پایستگی به گونه‌ای طراحی شده‌اند که ویژگی‌های نحوی افعال را مورد آزمون قرار دهند. معیارهای ترجمه‌ای نیز بر اساس رابطه فعل‌های مرکب فارسی و ترجمه آن‌ها به انگلیسی طراحی شده‌اند. نتایج دسته‌بندی، کارایی مطلوب معیارهای پیشنهادی را نشان می‌دهد.

چراغی و کریمی‌دوستان (۱۳۹۲) در پژوهش خود به طبقه‌بندی افعال زبان فارسی بر اساس ساخت رویدادی و نمودی پرداخته‌اند. در این پژوهش، پیکره‌ای متشکل از ۶۰ فعل به پیروی از نظریه کرافت (۲۰۱۲) مورد بررسی قرار گرفته است و افعال به چهار طبقه اصلی و ده طبقه فرعی تقسیم شده‌اند. همچنین اسحاقی و کریمی‌دوستان (۱۴۰۰) ابتدا با رویکردی پیکره‌بنیاد میزان زایایی ۲۱ فعل سبک زبان فارسی را مورد سنجش قرار داده‌اند؛ زایایی در زبان به معنی توانایی تولید صورت‌های جدید بر اساس قواعد موجود است، که یکی از مهم‌ترین ویژگی‌های زبان به شمار می‌رود. این پدیده بیشتر در حوزه واژه‌سازی مطرح است. پس از بررسی پیکره‌بنیاد، آن‌ها در چارچوب انگاره صرف توزیعی هله^۴ و مرننز^۵ (۱۹۹۳)، نقل از اسحاقی و کریمی‌دوستان، (۱۴۰۰) و مرننز (۲۰۱۳)، نقل از اسحاقی و کریمی‌دوستان، (۱۴۰۰) که از نظریه‌های بر پایه اصول نوساختگرایی است، به تبیین علت تفاوت در میزان زایایی این فعل‌ها

1. pointwise mutual information (PMI)

2. bootstrapping

3. Support Vector Machine (SVM)

4. Halle

5. Marantz

پرداخته‌اند. نتایج بیانگر این است که زایایی فعل‌های سبک به میزان سبک‌شدگی معنایی، امکانات درج و بازداری‌های درج هر یک از آن‌ها بستگی دارد. میزان سبک‌شدگی هر یک از فعل‌ها از طریق مقایسه صورت‌های سبک و واژگانی هر یک از فعل‌ها در داده‌های مستخرج از پیکره مورد ارزیابی قرار داده شده است. سپس امکانات معنایی و ساختاری درج، یعنی بافت‌های معنایی و ساختاری مجاز برای درج هر یک از فعل‌ها را محاسبه کرده‌اند و در پایان، بازداری‌های معنایی و ساختاری درج هر یک از فعل‌ها، یعنی بافت‌های معنایی و ساختاری که فعل مجوز درج در آن‌ها را نخواهد داشت، مورد محاسبه قرار داده‌اند. بر این اساس هرچه فعل به لحاظ معنایی سبک‌تر شده باشد، امکانات درج بیشتر و بازداری‌های کمتری داشته باشد، زایاتر است. برخی از پژوهشگران مانند علایی‌ابوذر و بی‌جن‌خان (۱۳۹۱) به بررسی فاصله میان جزء غیرفعلی و فعل سبک در زبان فارسی پرداخته‌اند. آن‌ها تلاش کرده‌اند گروه/گروه‌های نحوی که میان جزء غیرفعلی و فعل سبک فاصله می‌اندازند، تعیین کنند و همچنین این فاصله را بر حسب مقوله و گروه‌های نحوی که در این فاصله قرار می‌گیرند، نمایش دهند. فاصله میان جزء غیرفعلی و فعل سبک افعال مرکب با استفاده از آزمون تجزیه پراش محک خورده است و این نتیجه بدست آمده است که این فاصله وابسته به نوع گروه/گروه‌ها، مقوله/مقوله‌های نحوی به‌کاررفته میان جزء غیرفعلی و فعل سبک است؛ بدین‌گونه که بیشترین فاصله میان جزء غیرفعلی و فعل سبک زمانی ایجاد می‌شود که ترکیبی از گروه‌ها/مقوله‌های نحوی در این فاصله قرار گیرند. همچنین نوع جزء غیرفعلی نیز بر این فاصله تأثیر می‌گذارد؛ به این صورت که اگر جزء غیرفعلی یکی از کلمات «احتمال»، «اشاره»، «اظهار»، «برخورد»، «بو»، «تأکید»، «تعهد»، «توجه»، «حق»، «دخالت»، «راه» و «محبت» که بیشتر با فعل سبک «داشتن» بکار می‌روند، باشد بیشترین فاصله (برحسب تعداد گروه‌ها/مقوله‌های نحوی) میان جزء غیرفعلی و فعل سبک در افعال مرکب جداشدنی زبان فارسی ایجاد می‌شود.

۳. روش پژوهش

در این پژوهش ابتدا فراوانی افعال سبک «زدن»، «داشتن»، «کردن»، «سپردن»، «گرفتن»، «آمدن»، «دادن»، «افتادن»، «خوردن»، «کشیدن»، «آوردن»، «نمودن»، «رفتن»، «بردن» و «انداختن» از دو پیکره تخصصی «پیکره پژوهش‌نامه» و «پکا (پیکره کتاب‌های ایرانداک)» استخراج شده است. پیکره پژوهش‌نامه، از متون مقاله‌های پژوهش‌نامه پردازش و مدیریت اطلاعات ساخته شده است و پکا از متون کتاب‌های دیجیتال ایرانداک ساخته شده است. هر کدام به ترتیب، شامل بیش از چهار میلیون و ۷۸۰ هزار واژه و سه میلیون و ۳۲۹ هزار واژه هستند. محتوای این پیکره‌ها، متون عمومی نیست، بلکه دارای نوشته‌های بسیار تخصصی و

میان‌رشته‌ای مانند علم اطلاعات و دانش‌شناسی، فناوری اطلاعات، مدیریت دانش، زبان‌شناسی رایانشی، مدیریت اطلاعات و مانند آن‌هاست. همچنین به منظور افزایش کارایی پیکره‌ها، متون پیکره‌ها برچسب‌گذاری شده‌اند (برچسب‌گذاری اجزای کلام^۱). این نوع برچسب‌دهی، عملی کاربردی در بسیاری از حوزه‌های پیشرفته‌تر پردازش زبان طبیعی از جمله ترجمه ماشینی، خطایاب، تبدیل متن به گفتار، بازیابی اطلاعات، موتورهای جستجو و کمک به مدل‌های آماری است (علایی‌ابوذر و همکاران، ۱۴۰۰ و علایی‌ابوذر و حجت‌پناه، ۲۰۲۲). این نوع برچسب‌گذاری، یافتن افعال مرکب را در پیکره آسان‌تر می‌کند. چون در برخی موارد مانند صورت نوشتاری «کرد» (که نمونه‌ای از هم‌نگاره‌ها^۲ است)، می‌تواند اسم یا صفت (گرد) نیز باشد.

افعال مرکب فارسی خود به دو دسته «جداشدنی» و «جدانشدنی» تقسیم می‌شوند. آن‌هایی که اجزاءشان از هم جداشدنی نیست، نمی‌توان کلمه یا کلماتی را بین عنصر غیرفعلی و فعل سبک قرارداد. مانند «زنده نگه داشتن»، «تعبد کردن»، «کنار آمدن»، «برطرف کردن»، «رخ دادن»، «جا افتادن»، «پدید آمدن». افعال مرکبی که اجزاءشان از هم جداشدنی است، کلمه یا کلماتی می‌توانند بین فعل سبک و عنصر غیرفعلی قرار گیرند، مانند «راهی شدن» در جمله «وی پس از رسیدن به مقصد، راهی هتل بزرگی در میدان اصلی شهر شد» یا «وارد کردن» در جمله «نتوانستیم توپ را وارد دروازه تیم مقابل کنیم» یا «بخت داشتن» در جمله «دوبار بخت باز کردن دروازه فرانسه را داشتیم» یا «مانع شدن» در جمله «پژوهشگران تا حدی توانستند مانع از بروز این آفت بزرگ شوند» و یا «انتظار داشتن» در جمله «انتظار برانگیختگی داشت». بررسی این مسئله که آیا در پیکره‌های تخصصی، همانند پیکره‌های عمومی احتمال فاصله افتادن میان جزء غیرفعلی و فعل سبک می‌تواند بالا باشد نیز پس از استخراج فراوانی افعال سبک در دو پیکره تخصصی و انتخاب سه فعل سبک اول در فهرست استخراج‌شده و بررسی احتمال فاصله میان جزء غیرفعلی و فعل سبک مدنظر در بافت انجام شده است.

۴. تحلیل داده‌ها

همانگونه که در مقدمه اشاره شد، پیش‌فرض این پژوهش این است که همانند پیکره‌های عمومی، فراوانی افعال مرکبی که با سه فعل سبک «کردن»، «داشتن» و «دادن» ساخته می‌شوند، در پیکره‌های تخصصی نیز بیشتر از سایر افعال مرکبی باشد که با سایر افعال سبک ساخته می‌شوند. برای اطمینان، پیش از بررسی سه فعل سبک «کردن»، «داشتن» و «دادن»، ابتدا فراوانی افعال سبک فارسی از دو پیکره پژوهش‌نامه و پکا استخراج شد. افعال سبک

1. Part Of Speech (POS) tagging

2. homograph (صورت‌های نوشتاری یکسان، اما معنی یا تلفظ متفاوت)

بررسی شده شامل این افعال هستند: «زدن»، «داشتن»، «کردن»، «سپردن»، «گرفتن»، «آمدن»، «دادن»، «افتادن»، «خوردن»، «کشیدن»، «آوردن»، «نمودن»، «رفتن»، «بردن» و «انداختن».

نکته قابل توجه این است که برخی پژوهشگران مانند کریمی (۱۹۹۷) در فهرستی که از افعال سبک ارائه می‌دهند، فعل «شدن» را نیز لحاظ می‌کنند، اما در اکثر موارد، افعالی که با این فعل سبک ساخته می‌شوند معادل معلوم نیز دارند که با فعل سبک «کردن» ساخته می‌شوند، در این پژوهش افعالی که با فعل «شدن» ساخته شده‌اند مورد بررسی قرار نگرفته‌اند، زیرا اتفاق نظری میان زبان‌شناسان در مورد قراردادن «شدن» در فهرست افعال سبک وجود ندارد. به عنوان نمونه، لمبتون^۱ (۱۹۸۴)، کلباسی (۱۳۷۱)، دبیرمقدم (۱۳۷۴) و کریمی (۱۹۹۷)، «شدن» را در فهرست افعال سبک قرار می‌دهند. زبان‌شناسانی که «شدن» را در فهرست افعال سبک قرار می‌دهند، شاید معتقد به وجود ساخت مجهول در زبان فارسی نیستند. از آنجائیکه افعال مرکبی وجود دارد که با فعل سبک «شدن» ساخته شده‌اند، اما لزوماً ساخت معلوم معادل ندارند؛ این موارد خود می‌تواند شاهی برای این باشد که در فارسی ساخت مجهول وجود ندارد. برای مثال جملات (۱)، (۲) و (۳) را در نظر بگیرید، این مثال‌ها ساخت همتایی که با فعل سبک «کردن» یا «نمودن» ساخته شوند، ندارند.

(۱) موجب عزتمندی و سرافرازی شد

(۲) منجر به وارد آمدن خسارات مختصر شد

(۳) منشأ تحولی شگفت‌انگیز شد

بررسی این مسئله که آیا فعل «شدن» را در فهرست افعال سبک قرار دهیم یا خیر مسئله پژوهش حاضر نبوده است و در فهرست مورد بررسی، افعال سبکی لحاظ شده است که در ادبیات مربوط به بررسی افعال مرکب فارسی در فهرست افعال سبک قرار گرفته‌اند. نتیجه بررسی فراوانی افعال سبک فارسی در دو پیکره تخصصی در جدول ۴-۱ نمایش داده شده است.

در استخراج فراوانی افعال سبک از پیکره‌های تخصصی دو نکته حائز اهمیت است: با توجه به اینکه در متون علمی عمدتاً نتیجه پژوهش‌های پژوهشگر/نگارنده متن ارائه می‌شود و یا به پژوهش‌هایی که در گذشته انجام شده است استناد می‌شود، بیشتر افعال گذشته هستند و بنابراین، در این پژوهش زمان «ماضی، ساده، مثبت ۳» (سوم شخص مفرد، گذشته ساده و مثبت (به معنی غیرمنفی))، مبنای استخراج فراوانی افعال سبک از پیکره‌های تخصصی قرار گرفته است. به عنوان نمونه از میان همه صیغه‌های موجود فعل «زدن» یا «داشتن»، صیغه

1 Lambton

ماضی، ساده مثبت ۳، یعنی «زد» و «داشت» مبنای جستجو قرار گرفته است. نکته دوم این است که می‌دانیم برخی از افعال سبک خود به تنهایی یک فعل بسیط هستند و لزوماً در ساخت فعل مرکب به عنوان فعل سبک به کار نمی‌روند. به عنوان نمونه، فعل «زد» در جمله «او مرا زد»، خود یک فعل بسیط است. یا فعل «آمد»، خود در بسیاری جملات فعل بسیط است. جداکردن صورت فعل مرکب این افعال با حالتی که به صورت فعل بسیط به کار می‌روند، هدف پژوهش حاضر نبوده است. هدف این است که در هر صورت، فراوانی این افعال (چه به صورت بسیط و چه به صورت بخشی از فعل مرکب) از پیکره‌های تخصصی استخراج شوند.

جدول ۴-۱: بررسی فراوانی افعال سبک در پیکره‌های تخصصی پژوهش‌نامه و پکا

فهرست افعال سبک	فراوانی در پیکره پژوهش‌نامه	فراوانی در پکا (پیکره کتاب‌های ایراندک)	مجموع فراوانی در دو پیکره
کردن	۳۴۲۷	۲۴۲۱	۵۸۴۸
دادن	۴۰۲۴	۱۰۱۳	۵۰۳۷
گرفتن	۲۶۹۷	۵۴۹	۳۲۴۶
نمودن	۱۵۸۷	۷۸۹	۲۳۷۶
داشتن	۱۴۰۶	۷۷۵	۲۱۸۱
آمدن	۵۷۶	۱۷۱	۷۴۷
آوردن	۳۳۹	۲۴۲	۴۸۱
بردن	۲۷۲	۱۵۹	۴۳۱
رفتن	۱۰۶	۱۲۳	۲۲۹
زدن	۵۵	۴۲	۹۷
کشیدن	۱۶	۲۵	۴۱
افتادن	۱۸	۱۸	۳۶
خوردن	۱۷	۱۰	۲۷
انداختن	۷	۱۱	۱۸
سپردن	۳ مورد که هیچ‌کدام فعل مرکب نساخته‌اند	۱ (به فراموشی سپردن)	۴

جدول ۴-۱ نشان می‌دهد، از مجموع ۲۰۷۹۹ مورد مربوط به افعال سبک بررسی شده در پیکره‌های تخصصی، دو فعل سبک «کردن» با فراوانی ۵۸۴۸ و فعل «دادن» با فراوانی ۵۰۳۷، در رتبه اول و دوم از نظر فراوانی قرار دارند؛ فراوانی این دو فعل سبک در پیکره عمومی نیز بالا است و در رتبه‌های نخست قرار دارند. فعل سبک «گرفتن» با فراوانی ۳۲۴۶ در پیکره‌های تخصصی در رتبه سوم قرار دارد. به عبارتی دیگر، حدود ۲۸٪ فعل سبک «کردن»، ۲۴٪ فعل

سبک «دادن» و ۱۵٪ فعل سبک «گرفتن» در پیکره‌های تخصصی به کاررفته است. حال آنکه در پیکره متنی فارسی (که پیکره‌ای است عمومی) فعل سبک «گرفتن»، از نظر فراوانی بعد از افعال سبک «کردن»، «دادن»، «داشتن» و «نمودن» قرار دارد (علائمی ابوذر و بی‌جن‌خان، ۱۳۹۱).

در مرحله بعد سعی شد فهرستی از افعال مرکبی که با این سه فعل سبک ساخته شده‌اند از دو پیکره استخراج شود. این فهرست شامل موارد جدول ۴-۲ می‌شود (البته همه موارد فهرست نشده است و موارد زیر به عنوان نمونه آورده شده‌اند):

جدول ۴-۲: فهرست برخی از افعال مرکب فارسی با فعل سبک «کردن»، «دادن» و «گرفتن»

نمونه‌هایی از افعال مرکب با فعل سبک «کردن»	نمونه‌هایی از افعال مرکب با فعل سبک «دادن»	نمونه‌هایی از افعال مرکب با فعل سبک «گرفتن»
سفر کردن، فراموش کردن، تبادل کردن، تبادل نظر کردن، تجدید نظر کردن، استفاده کردن، مراقبت کردن، تقدیم کردن، اهداء کردن، مطرح کردن، درخواست کردن، فتح کردن، مغلوب کردن، توصیف کردن، قدردانی کردن، تجسم پیدا کردن، دعوت کردن، اعطا کردن، یاد کردن، گزارش کردن، تباه کردن، معرفی کردن، قطع کردن، کسب کردن، عمل کردن، پیدا کردن، عبادت کردن، دیدار کردن، از آن خود کردن، کسب اطلاع کردن، سقوط کردن، معنی کردن، رعایت کردن، کردن، انتخاب کردن، اظهار نظر کردن، احساس کردن، ایفای نقش کردن، اجرا کردن، القاء کردن، ناامید کردن، فعالیت کردن، بسنده کردن، کمک کردن، توجه کردن، نگاه کردن، نقد کردن، طبقه بندی کردن، خدمت کردن، قبول کردن، خلاصه کردن، مبدل کردن، تشکر کردن، تجمع	امتیاز دادن، جهت دادن، گزارش دادن، سامان دادن، دست دادن، معنی دادن، پس دادن، تشخیص دادن، تکان دادن، خاتمه دادن، پایان دادن، عبور دادن، پوشش دادن، صیقل دادن، سفارش دادن، یاری دادن، تطبیق دادن، شکل دادن، هشدار دادن، سروسامان دادن، تن دادن، شعار دادن، توفیق دادن، انجام دادن، توضیح دادن، تغییر دادن، قرار دادن، گسترش دادن، نشان دادن، تحویل دادن، از دست دادن، ترجیح دادن، دادن، درس دادن، اجازه دادن، ادامه دادن، یاد دادن، ارائه دادن، سوق دادن، رونق دادن، فریب دادن، حرکت دادن، کفاف	درد گرفتن، اوج گرفتن، پی گرفتن، آرام گرفتن، کناره گرفتن، به دوش گرفتن، دوش گرفتن، کمک گرفتن، برده گرفتن، پاداش گرفتن، تحت پوشش گرفتن، تماس گرفتن، سرعت گرفتن، بالا گرفتن، تعلق گرفتن، به خدمت گرفتن، شتاب گرفتن، پیشی گرفتن، بر عهده گرفتن، موضع گرفتن، قرض گرفتن، بیعت گرفتن، درس گرفتن، نتیجه گرفتن، انتقام گرفتن، پند گرفتن، قرار گرفتن، فیلم گرفتن، فاصله گرفتن، نادیده گرفتن، انجام گرفتن، دست گرفتن، یاد گرفتن، بهره گرفتن، هدف گرفتن، جای گرفتن، در اختیار گرفتن، تصمیم گرفتن، ایراد گرفتن، پس گرفتن، پا گرفتن، به عهده گرفتن، دور

<p>گرفتن، در نظر گرفتن، تحویل گرفتن، شکل گرفتن</p>	<p>دادن، امکان دادن، شخصیت دادن، بازی دادن، حکم دادن، نمایش دادن، فرمان دادن، رخ دادن، تکیه دادن، پاسخ دادن، جواب دادن، تشکیل دادن، حق دادن، رأی دادن، نجات دادن، تعلیم دادن، اطمینان دادن، تسلی دادن، راه دادن، اطلاع دادن، تذکر دادن، تنزل دادن، انتقال دادن، جلوه دادن، شکست دادن، پرورش دادن، آموزش دادن، هدر دادن، ارتقاء دادن، گوش دادن، استعفاء دادن، کاهش دادن، افزایش دادن</p>	<p>کردن، تحلیل کردن، مشاهده کردن، حل کردن، شرکت کردن، تقدیر کردن، گران کردن، رد کردن، تهیه کردن، تجلیل کردن، توصیه کردن، جلوگیری کردن، درج کردن، عنوان کردن، خودداری کردن، تلقی کردن، متحیر کردن، منعقد کردن، دور کردن، مجازات کردن، درک کردن، سوال کردن، جنگ کردن، دنبال کردن، داخل کردن، خدمت کردن، پیشنهاد کردن، مجازات کردن، اشاره کردن، قربانی کردن، زندگی کردن، ارتباط برقرار کردن، محبت کردن، جذب کردن، تعهد کردن، عذرخواهی کردن، طی کردن، باز کردن، اشتباه کردن</p>
--	---	---

شایان ذکر است موارد بالا به عنوان نمونه ارائه شده‌اند و تعداد افعال مرکب استخراج‌شده از پیکره بیش از این تعداد است. برای بررسی فاصله میان جزء غیرفعلی و فعل سبک، افعال مرکب فهرست‌شده در پیکره‌های پژوهش‌نامه و پکا مورد بررسی قرار گرفتند. این دو پیکره در سامانه پیکره‌های ایرانداک (سایا) بارگذاری شده‌اند (<https://sapa.irandoc.ac.ir>) و یکی از قابلیت‌های این سامانه این است که امکان جستجوی واژه به روش چندنگاشت^۱ را فراهم می‌کند. در جستجو به صورت فهرست چندتایی/چندنگاشت، به جای نمایش واژه/عبارت در کل متن، واژه/عبارت مورد جستجو همراه چند واژه قبل یا بعد از واژه/عبارت مورد جستجو در بازیابی اطلاعات نمایش داده می‌شود که تعداد این واژه‌های قبل یا بعد از واژه جستجو را کاربر تعیین می‌کند. به عنوان نمونه، می‌توان فعل سبک «کرد» را در قسمت جستجو وارد کرد و سه یا پنج واژه قبل و بعد از این واژه را در بازیابی اطلاعات دید. در خروجی، این واژه صرفاً در بافت زبانی مدنظر در متون پیکره، در سطرهای مجزا نمایش داده می‌شود (مثلاً واژه «کرد» همراه با سه یا پنج واژه قبل و بعد از این واژه) و کاربر مجبور نیست کل متن را بررسی کند؛ فقط کافی است خروجی این نوع جستجو را دریافت کند و تنها همان فایل خروجی را بررسی کند. نمونه‌ای از جستجو به روش فهرست چندنگاشت که حاصل جستجوی فعل سبک «کرد»

1. N gram

به همراه پنج نگاشت قبل از این فعل سبک است در زیر نمایش داده شده است. در شکل ۴-۱ جستجوی این واژه بدون استفاده از فهرست چندنگاشت (بدون استفاده از آیکن Ngram) و در شکل ۴-۲، جستجوی همین واژه با مشخص کردن پنج نگاشت قبل و پنج نگاشت بعد از این فعل سبک نمایش داده شده است.

تصویر شماره ۱: نمایش جستجوی فعل سبک «کرد» بدون استفاده از فهرست چندنگاشت

نوع نگاشت	واژه	موضوع	بافت	نوع نگاشت	نگاشت
نگاشت	این واژه را در جمله، عبارت‌های خاص و ...	افعال سبکی	کتاب: ۱۳۹۸ - ۱۳۹۹	فهرست نگاشت	۱- کرد
نگاشت	فهرست نگاشت: ...	افعال سبکی	سند: ...	فهرست نگاشت	۲- کرد
نگاشت	فهرست نگاشت: ...	افعال سبکی	سند: ...	فهرست نگاشت	۳- کرد
نگاشت	فهرست نگاشت: ...	افعال سبکی	سند: ...	فهرست نگاشت	۴- کرد

مزیت اینگونه بررسی این است که مجبور نیستیم یکی یکی افعال مرکب را در پیکره جستجو کنیم و سپس، در کل پیکره و در کل جملاتی که این افعال در آن‌ها به کار رفته‌اند، به بررسی مقولاتی که میان جزء غیرفعلی و فعل سبک قرار گرفته‌اند، بپردازیم. تنها کافی است افعال سبک مدنظر (کردن، دادن و گرفتن) در سامانه وارد شوند و با انتخاب یک عدد در فهرست چندنگاشت، خروجی تهیه شود که تنها فعل سبک مدنظر به همراه چند واژه قبل و بعد از خود نمایش داده شده است. شایان ذکر است که نقطه، ویرگول و غیره همه در محاسبه لحاظ می‌شوند و برای همین منظور واژه «نگاشت» استفاده شده است. در پژوهش حاضر عدد پنج برای این منظور انتخاب شد.

بررسی اجمالی داده‌های خروجی این مرحله نشان داد به نظر می‌رسد در متون تخصصی (پیکره‌های تخصصی) احتمال فاصله افتادن میان جزء غیرفعلی و فعل سبک کمتر از متون عمومی است. به داده‌های زیر (شکل ۴-۳، ۴-۴ و ۴-۵) که حاصل جستجوی افعال سبک «کردن»، «دادن» و «گرفتن» به همراه پنج نگاشت قبل و پنج نگاشت بعد از این افعال است توجه کنید:

تصویر شماره ۲: نمایش داده‌ها پس از انتخاب فهرست چندنگاشت (انتخاب پنج نگاشت قبل و پنج نگاشت بعد از واژه هدف)

ردیف	عنوان	نویسنده	تاریخ
۱	مقاله علمی-پژوهشی	پنج‌گانه	۱۳۹۹
۲	مقاله علمی-پژوهشی	پنج‌گانه	۱۳۹۹
۳	مقاله علمی-پژوهشی	پنج‌گانه	۱۳۹۹
۴	مقاله علمی-پژوهشی	پنج‌گانه	۱۳۹۹
۵	مقاله علمی-پژوهشی	پنج‌گانه	۱۳۹۹
۶	مقاله علمی-پژوهشی	پنج‌گانه	۱۳۹۹
۷	مقاله علمی-پژوهشی	پنج‌گانه	۱۳۹۹
۸	مقاله علمی-پژوهشی	پنج‌گانه	۱۳۹۹

تصویر شماره ۳: نمایش افعال مرکبی که با فعل سبک «کردن» ساخته شده‌اند به همراه پنج نگاشت قبل و پنج نگاشت بعد از واژه هدف در Ngram

ردیف	عنوان	نویسنده	تاریخ
۱	مقاله علمی-پژوهشی	پنج‌گانه	۱۳۹۹
۲	مقاله علمی-پژوهشی	پنج‌گانه	۱۳۹۹
۳	مقاله علمی-پژوهشی	پنج‌گانه	۱۳۹۹
۴	مقاله علمی-پژوهشی	پنج‌گانه	۱۳۹۹
۵	مقاله علمی-پژوهشی	پنج‌گانه	۱۳۹۹
۶	مقاله علمی-پژوهشی	پنج‌گانه	۱۳۹۹
۷	مقاله علمی-پژوهشی	پنج‌گانه	۱۳۹۹
۸	مقاله علمی-پژوهشی	پنج‌گانه	۱۳۹۹

تصویر شماره ۴: نمایش افعال مرکبی که با فعل سبک «دادن» ساخته شده‌اند به همراه پنج

نگاشت قبل و پنج نگاشت بعد از واژه هدف در Ngram

پیکره	عنوان مبرکه	یافته	موضوع	پیداگر	اطلاعات بیشتر
مشاهده	تحلیل کتب و انتقادی پژوهش‌های حوزه کتابخانه‌های دیجیتال در ایران	کار روند - تاریخ نشان ۱۳۹۵ موجودات - ارزیابی کتابخانه‌های دیجیتال	فناوری اطلاعات	مینا قائمی اهوری	مشاهده
مشاهده	تحلیل کتب و انتقادی پژوهش‌های حوزه کتابخانه‌های دیجیتال در ایران	مجموعه - بررسی پژوهش‌ها نشان ۱۳۹۵ که اولاً بافتی عمده تطبیقات	فناوری اطلاعات	مینا قائمی اهوری	مشاهده
مشاهده	تحلیل کتب و انتقادی پژوهش‌های حوزه کتابخانه‌های دیجیتال در ایران	روش‌شناسی تطبیقات پیشین را نشان ۱۳۹۵ که نمونه بارز آن	فناوری اطلاعات	مینا قائمی اهوری	مشاهده
مشاهده	تحلیل کتب و انتقادی پژوهش‌های حوزه کتابخانه‌های دیجیتال در ایران	تضمینها در بعد - ۱۳۹۵ رخ ۱۳۹۵ پدیدایش وب که همه	فناوری اطلاعات	مینا قائمی اهوری	مشاهده
مشاهده	تحلیل کتب و انتقادی پژوهش‌های حوزه کتابخانه‌های دیجیتال در ایران	از عوامل در گسترش قرار ۱۳۹۵ این تعریف با توجه	فناوری اطلاعات	مینا قائمی اهوری	مشاهده
مشاهده	تحلیل کتب و انتقادی پژوهش‌های حوزه کتابخانه‌های دیجیتال در ایران	را با اهمیت به‌تر گسترش ۱۳۹۵ با مراجعه به پارادایم‌ها	فناوری اطلاعات	مینا قائمی اهوری	مشاهده
مشاهده	تحلیل کتب و انتقادی پژوهش‌های حوزه کتابخانه‌های دیجیتال در ایران	رایج با موضوع پژوهش نشان ۱۳۹۵ - پژوهشی که منظور جامع	فناوری اطلاعات	مینا قائمی اهوری	مشاهده
مشاهده	تحلیل کتب و انتقادی پژوهش‌های حوزه کتابخانه‌های دیجیتال در ایران	قرار ۱۳۹۵ - یافته‌ها نشان ۱۳۹۵ در زمین محتات - اطلاعات	فناوری اطلاعات	مینا قائمی اهوری	مشاهده
مشاهده	تحلیل کتب و انتقادی پژوهش‌های حوزه کتابخانه‌های دیجیتال در ایران	نتایج پژوهش وی نشان ۱۳۹۵ که حوزه‌های موضوعی سازمان‌دهی اطلاعات	فناوری اطلاعات	مینا قائمی اهوری	مشاهده

تصویر شماره ۵: نمایش افعال مرکبی که با فعل سبک «گرفتن» ساخته شده‌اند به همراه

پنج نگاشت قبل و پنج نگاشت بعد از واژه هدف در Ngram

پیکره	عنوان مبرکه	یافته	موضوع	پیداگر	اطلاعات بیشتر
مشاهده	عشق منبرهای جمعیت شناسی در ایران - مطالعه سوادگیری‌های شناسی در فرایند بازیابی اطلاعات	بده - حوزه تهیه قرار ۱۳۹۵ یافته‌های پژوهش نشان ۱۳۹۵	شعبه و بازیابی	حسن نورانی	مشاهده
مشاهده	عشق منبرهای جمعیت شناسی در ایران - مطالعه سوادگیری‌های شناسی در فرایند بازیابی اطلاعات	یافته‌های پژوهش عنوان چنین نتیجه ۱۳۹۵ که بهیچ منبرهای جمعیت شناسی	شعبه و بازیابی	حسن نورانی	مشاهده
مشاهده	عشق منبرهای جمعیت شناسی در ایران - مطالعه سوادگیری‌های شناسی در فرایند بازیابی اطلاعات	عنوان سواد اطلاعاتی را می‌تواند ۱۳۹۵ - کاربردهای سوادگیری‌های شناسی	شعبه و بازیابی	حسن نورانی	مشاهده
مشاهده	عشق منبرهای جمعیت شناسی در ایران - مطالعه سوادگیری‌های شناسی در فرایند بازیابی اطلاعات	رایج آنها مورد تهیه قرار ۱۳۹۵ همچنین بررسی بازیابی	شعبه و بازیابی	حسن نورانی	مشاهده
مشاهده	عشق منبرهای جمعیت شناسی در ایران - مطالعه سوادگیری‌های شناسی در فرایند بازیابی اطلاعات	انتخاب شد از برای دیگر قرار ۱۳۹۵ که در نهایت بررسی	شعبه و بازیابی	حسن نورانی	مشاهده
مشاهده	عشق منبرهای جمعیت شناسی در ایران - مطالعه سوادگیری‌های شناسی در فرایند بازیابی اطلاعات	آزادی پژوهشگر حوزه جستجو قرار ۱۳۹۵ و تعداد ۱۳۹۵ -	شعبه و بازیابی	حسن نورانی	مشاهده
مشاهده	عشق منبرهای جمعیت شناسی در ایران - مطالعه سوادگیری‌های شناسی در فرایند بازیابی اطلاعات	حالات ۱۳۹۵ مورد بررسی قرار ۱۳۹۵ و مواردی که به پژوهش	شعبه و بازیابی	حسن نورانی	مشاهده
مشاهده	عشق منبرهای جمعیت شناسی در ایران - مطالعه سوادگیری‌های شناسی در فرایند بازیابی اطلاعات	در پژوهش خود بررسی قرار ۱۳۹۵ بر حسب جمعیت تطبیقات	شعبه و بازیابی	حسن نورانی	مشاهده
مشاهده	بررسی کتابخانه‌های حوزه مطالعه روش‌های جستجو شناسی	تعداد یافته‌ها عنوان در اعتبار ۱۳۹۵ پژوهشگر نشان ۱۳۹۵ در	اصطلاح شناسی	علیه تهرانی میلانی	مشاهده

۵. بحث و نتیجه‌گیری

در این پژوهش ابتدا فراوانی افعال سبک «زدن»، «داشتن»، «کردن»، «سپردن»، «گرفتن»، «آمدن»، «دادن»، «افتادن»، «خوردن»، «کشیدن»، «آوردن»، «نمودن»، «رفتن»، «بردن» و «انداختن» از دو پیکره تخصصی «پیکره پژوهش‌نامه» و «پکا (پیکره کتاب‌های ایراندک)» استخراج شد. انتظار می‌رفت فراوانی افعال سبک در این دو پیکره تخصصی تفاوت معناداری با فراوانی آن‌ها در پیکره‌های عمومی، مانند پیکره متنی، نداشته باشد.

در پیکره متنی فارسی که افعال مرکب فارسی مورد بررسی قرار گرفته است، فراوانی افعال مرکبی که با سه فعل سبک «کردن»، «دادن» و «داشتن» ساخته شده‌اند بیشتر از سایر افعال مرکب است. بررسی فراوانی افعال سبک در دو پیکره تخصصی پژوهش‌نامه و پکا نشان داد از مجموع ۲۰۷۹۹ مورد مربوط به افعال سبک بررسی شده، دو فعل سبک «کردن» با فراوانی ۵۸۴۸ و فعل «دادن» با فراوانی ۵۰۳۷، در رتبه اول و دوم از نظر فراوانی قرار دارند؛ فراوانی این دو فعل سبک در پیکره عمومی نیز بالا است و در رتبه‌های نخست قرار دارند. فعل سبک «گرفتن» با فراوانی ۳۲۴۶ در پیکره‌های تخصصی در رتبه سوم قرار دارد. به عبارتی دیگر، حدود ۲۸٪ فعل سبک «کردن»، ۲۴٪ فعل سبک «دادن» و ۱۵٪ فعل سبک «گرفتن» در پیکره‌های تخصصی به‌کاررفته است.

حال آنکه در پیکره متنی فارسی (که پیکره‌ای است عمومی)، فعل سبک «گرفتن»، از نظر فراوانی بعد از افعال سبک «کردن»، «دادن»، «داشتن» و «نمودن» قرار دارد. افعال مرکبی که با این سه فعل سبک ساخته می‌شوند از نظر احتمال فاصله میان جزء غیرفعلی و فعل سبک در بافت تعریف شده (پنج نگاهت قبل و پنج نگاهت بعد از فعل سبک موردنظر) نیز مورد بررسی کلی قرار گرفتند. بررسی رفتار افعال مرکب پرسیامد در پیکره‌های تخصصی نشان داد نویسندگان متون تخصصی، عمدتاً پژوهشگران، دانشجویان و استادان، کمتر تمایل دارند بین جزء غیرفعلی و فعل سبک فاصله بیاندازند و بنابراین، در بیشتر افعال مرکب جزء غیرفعلی و فعل سبک در کنار هم قرار می‌گیرند و مقوله و گروه نحوی میان آن‌ها قرار نمی‌گیرد. در حالیکه در متون عمومی احتمال اینکه نویسندگان/گویندگان، گروه/گروه‌های نحوی و مقولات گوناگون میان جزء غیرفعلی و فعل سبک قرار دهند بیشتر از متون تخصصی است. به عنوان نمونه احتمال اینکه جمله‌ای مانند «تن به مهاجرت و انتقال ثروت داد» یا «توجه هرچه بیشتر به آثار پربرکت اقامه نماز کردند» در متون عمومی و مطبوعات دیده شود بیشتر از متون تخصصی است. بررسی دقیق‌تر رفتار این افعال در متون تخصصی، به همراه استخراج همه افعال مرکب ساخته‌شده با این افعال سبک و فاصله میان اجزای این افعال مرکب خود می‌تواند موضوع پژوهش‌های بعدی باشد.

منابع

- اسحاقی، مهدیه و کریمی‌دوستان، غلامحسین. (۱۴۰۰). «زیایی فعل‌های سبک در زبان فارسی». *پژوهش‌های زبانی*. ۱۲ (۲). صص. ۴-۲۸.
- تسلیمی‌پور، شیوا. (۱۳۹۱). *پردازش خودکار معنایی افعال مرکب فارسی*. پایان‌نامه کارشناسی ارشد. رشته مهندسی کامپیوتر (هوش مصنوعی). دانشگاه شیراز.
- چراغی، زهرا و کریمی‌دوستان، غلامحسین. (۱۳۹۲). «طبقه‌بندی افعال زبان فارسی بر اساس ساخت رویدادی نمودی». *پژوهش‌های زبانی* (دانشکده ادبیات و علوم انسانی، دانشگاه تهران). ۴ (۲). صص. ۴۱-۶۰.
- دبیرمقدم، محمد. (۱۳۷۴). «فعل مرکب در زبان فارسی». *مجله زبان‌شناسی*، س ۱۲، ش ۱ و ۲. صص. ۲-۴۶.
- طباطبائی، علاء‌الدین. (۱۳۸۴). «فعل مرکب در زبان فارسی». *نامه فرهنگستان*. ۷ (۲). صص. ۲۶-۳۴.
- ظفرآبادی، بهادر و رحیمیان، جلال. (۱۳۹۹). «بررسی تحول و تعداد وقوع فعل سبک داشتن، دیدن، آمدن، آوردن و گرفتن از سال ۱۲۲۰ تا ۱۳۲۰». *نشریه پژوهش‌های زبان‌شناسی*. ۱ (۲۲). صص. ۲۳-۴۲.
- عسگریان، نرجس. (۱۳۹۱). *شناسایی خودکار افعال مرکب فارسی*. پایان‌نامه کارشناسی ارشد. مهندسی کامپیوتر- هوش مصنوعی. دانشگاه شیراز.
- علائی ابوذری، الهام و بی‌جن‌خان، محمود. (۱۳۹۱). «بررسی افعال مرکب جداشدنی زبان فارسی در چارچوب زبان‌شناسی پیکره‌ای». *مجموعه مقالات دومین هم‌اندیشی زبان‌شناسی رایانشی*. انجمن زبان‌شناسی ایران. ۴۵-۵۸.
- علائی ابوذری، الهام، نصراله پاک‌نیت، علی‌اصغر حجت‌پناه، مجتبی زالی و محمدهادی آقالویی آغمیونی. (۱۴۰۰). *ساخت پیکره متنی از مقاله‌های پژوهش‌نامه پردازش و مدیریت اطلاعات*. پژوهشگاه علوم و فناوری اطلاعات ایران (ایرانداک).
- کلباسی، ایران (۱۳۷۱). *ساخت اشتقاقی واژه در فارسی امروز*. مؤسسه مطالعات و تحقیقات فرهنگی. کوهستانی، منوچهر. (۱۳۸۹). *بررسی خطاهای املائی و نگارشی در وبلاگ‌های فارسی و ماهیت زبان‌شناختی آن‌ها*. پایان‌نامه کارشناسی ارشد، تهران. دانشگاه تهران.
- مرادی، ابراهیم و غلامحسین کریمی‌دوستان. (۱۳۹۲). «بررسی معنایی فعل مرکب در زبان فارسی». *نشریه ادب و زبان*. دوره ۱۶. ش ۳۳، ۳۰۶-۳۲۵.
- ناتل خانلری، پرویز. (۱۳۶۶). *تاریخ زبان فارسی*. ج ۲. نشر نو.
- وحدت‌زاده، سارا. (۱۳۹۱). *بررسی پیکره‌بنیاد فعل مرکب بر مبنای کتاب‌های دستور زبان فارسی دبیرستان‌های ایران*. پایان‌نامه کارشناسی ارشد. دانشگاه پیام نور. استان تهران.
- Alayiaboozar, E., & Bijankhan, M. (2012). "Studying Persian separable compound verbs in corpus linguistics". *Second conference in computational linguistics*. Iranian linguistics society. Pp. 45-58. [in Persian].
- Alayiaboozar, E., & Hojjatpanah, A (2022). "Steps for creating two Persian specialized corpora". *International Journal of Information Science and Management (IJISM)*. Vol20, Issue 4, 231-243.

- Alayiboozar, E., Pakniat, N., Zali, M., & Aghalooyi Aghmiyooni, M.H. (2021). *Building a corpus from the published articles of Iranian Journal of Information Management and Processing*. Iranian Research Institute for Information Science and Technology (IranDoc). [in Persian].
- Asgarian, N. (2012). *Automatic identification of Persian compound verbs*. M.A. University of Shiraz. [in Persian].
- Cheraghi, Z., and Karimidoostan, Gh.H. (2013). "Classification of Persian verbs based on event and aspectual". *Journal of Language Research*. 4 (2), 41-60. [in Persian].
- Dabirmoghadam, M. (1995). "Compound verbs in Persian". *Journal of Zabanshenasi*. 12 (1 and 2), 2-46. [in Persian].
- Eshaghi, M., and Karimidoustan, Gh.H. (2021). "The productivity of Persian light verbs". *Journal of language researches*. 12 (2), 4-28.
- Kalbasi, I. (1992). *Derivational structure of words in modern Persian*. Cultural studies institute. [in Persian].
- Karimi, S. (1997). "Persian complex verbs: idiomatic or compositional". *Lexicology*. University of Arizona. Pp. 273-318.
- Kouhestani, M. (2010). *Studying spell and writing errors in Persian weblogs and their linguistics characteristics*. MA thesis. University of Tehran. [in Persian].
- Lambton, A.K.S. (1984). *Persian grammar*. Cambridge university press
- Moradi, E., & Karimidoostan, Gh.H. (2013). "Semantic study of compound verbs in Persian". *Nasrpazhuhi e adab e Farsi*. No. 33. Pp. 297-317. [in Persian].
- Natel Khanlari, P. (1987). *The history of Persian language*. Nashr e No.
- Rasooli, M.S., Faili, H., & Minaei-Bidgoli, B. (2011). Unsupervised identification of Persian compound verbs. *Advances in artificial intelligence*. In *proceeding of 10th Mexican international conference on artificial intelligence. MICAI*: 394-406.
- Sharif, B., and Shadmehr, M. (2022). "Semantics of Persian light verbs". *Journal of linguistics studies, theory and practice*. 1 (1), 191-206.
- Sinclair, J. (2004). "Corpus and Text: Basic Principles". *Developing Linguistic Corpora: a Guide to Good Practice*. M. Wynne (ed.), .ahds.literature, languages and linguistics. The Oxford Text Archive.
- Tabatabaei, A. (2005). "Compound verbs in Persian". *Naame Farhangestan*. 7 (2), 26-34. [in Persian].
- Taslimipoor, Sh. (2012). *Automatic semantic processing of Persian compound verbs*. M.A. thesis. University of Shiraz. [in Persian].
- Vahdatzade, S. (2012). *Corpus-based study of compound verbs based on Persian grammar books in Iranian high schools*. M.A. thesis. University of Payam e Noor. [in Persian].
- Zafarabadi, B., and Rahemeyan, J. (2020). "Investigating the evolution and frequency of Persian light verbs Dashtan (Have), Avardan (Bring), Gereftan (Get), and Didan (See) from 1840-1940". *Journal of researches in linguistics*. 12 (1), 23-42.