



<https://jrl.ui.ac.ir/?lang=en>

Journal of Researches in Linguistics

E-ISSN: 2322-3413

14(1), 173-198

Received: 15.11.2022 Accepted: 18.01.2023

Research Paper

A Corpus of Light Verb Constructions in Persian

Mahdieh Eshaghi 

Post-doctoral student of Linguistics, University of Tehran, Tehran, Iran
mahdie_eshaghi@ut.ac.ir

Gholamhossein Karimi Doostan 

Linguistic professor, University of Tehran, Tehran, Iran
gh5karimi@ut.ac.ir

Abstract

A linguistic corpus is a collection of linguistic data derived from language texts, which represent the real patterns of language use to the researchers. The priority of the corpus over other linguistic resources stems from the amount of data it represents and the possibility of computer use in linguistic studies. In the present study, an annotated monolingual linguistic corpus of Light Verb Constructions (LVCs) of Persian language (LCP) developed by the authors was introduced. The corpus contained more than 6000 LVCs, which were used in more than 2000000 linguistic contexts. Just a comparison of the number of LVCs with the number of simple verbs in Persian is enough to indicate the importance of these types of language resources. This annotated corpus presented LVCs formed by 21 Persian Light Verbs (LVs) that are used in real contexts. This unprecedented work has the capacity to easily provide a large computational bulk of various data for the researchers to assess the existing hypotheses and put forward the new ones.

Keywords: Persian Language, Language Resources, Linguistic Corpus, Light Verb Constructions, Natural Language Processing

Introduction

Light verbs are a group of verbs that have lost part of their semantic contents during language evolution. These so-called light verbs in combination with a preverbal element like a noun, adjective, or prepositional phrase form Light Verb Constructions (LVCs) in Persian. The study of LVCs is important not only theoretically, but also practically. The verbal system of Persian largely consists of LVCs and it doubles the importance of their study in this language. Nevertheless, many studies have pointed out the challenges that Persian LVCs pose for computational systems. They have emphasized the lack of appropriate computer resources and the necessity of studies that provide the researchers with their standard language patterns in this language (Maerefat, 2004; Hasas Sediqi, 2010; Taslimipour, 2012; Askariyan, 2012, and Barfi, 2016 among others). Although there are already valuable Persian corpora developed by specialists like Bijan Khan (2004, 2018), Asi (2005), and Al-e-Ahmad et al. (2010) in this field, there is no corpus to comprehensively represent LVCs of all productive Persian Light Verbs (LVs). The only available corpus dealing with Persian LVCs is PresPred (Samvellian & Faqiri, 2013), which represents those consisting of one of the twenty-one

*Corresponding author



This is an open access article under the CC BY-NC-ND/4.0/ License (<https://creativecommons.org/licenses/by-nc-nd/4.0/>).



<https://doi.org/10.22108/jrl.2023.135758.1685>

productive Persian LVs (Zadan). To address this need, we developed the first corpus for Persian LVCs.¹ This annotated corpus presented the LVCs formed by 21 Persian LVs that are used in real contexts. The present unprecedented work has the capacity to readily provide a large computational bulk of various data for researchers.

Materials and Methods

Development of the present corpus experienced the following steps: designing the structure of the corpus, selecting a corpus as a basis, normalizing the texts, defining the search nodes, writing macro codes in Visual Basic Analysis (VBA) language for preparing the search software, extracting all the sentences containing the verbs under investigation (regardless of being light or lexical verbs), extracting the sentences with LVCs, and finally selecting an annotation model and applying it to the results. It was designed to be a synchronic monolingual corpus of Persian LVCs. We chose a corpus developed by Bijan Khan (2018) as a basis. It was developed in the Research Institute of Information and Communication Technology and contained 950000 text files. First, we normalized the texts and then used VBA macro codes to extract the LVCs consisting of 21 Persian LVs (da:shtan: have, kardan: do, shodan: become, gashtan: turn, goza:shtan: put, keshidan: pull, didan: see, da:dan: give, bakhshidan: give, grant, gereftan: get, yaftan: obtain, ?a:madan: come, ?a:vardan: bring, residan: arrive, raftan: go, ?ofta:dan: fall, ?anda:khtan: throw, bordan: take, khordan: collide, zadan: hit, and bastan: tie). then, constituency test (^{topicalization}, coordination, deletion, and substitution) was applied to distinguish LVCs from lexical verbs. Annotation of LVCs has been done at the word level within a Distributed Morphology setting (Halle & Marantz, 1993 and Marantz, 2013). Preverbal elements and LVs were considered as categoryless elements (annotated as Pre-Verbs (PVs)) and categorizers (annotated as LVs), respectively. In addition, the present and past lemmas of each LVC were given and their separability/inseparability was annotated as SEP/INSEP. It should be noted that in line with Karimi-Doostan (2011), the cases, in which preverbal elements and LVs were broken by a negative particle (neg), the imperfective morpheme (mi), modals and auxiliaries, such as ba:yad (should, must), xa:stan (will) as a future auxiliary verb, and da:s'tan (to have) as a progressive auxiliary verb, as well as clitic pronouns like -es' (it), were annotated as INSEP. Table 1 represents these tags and the colors used for each of them.

Discussion of Results and Conclusion

Light Verb Constructions (LVCs) as a subset of complex or multi-word predicates are among the most challenging topics of language. The present study developed a monolingual corpus of Persian LVCs with the aim of providing the researchers with a large computational bulk of data related to these challenging constructions and improving the authenticity of the studies conducted in this field. The present corpus included about 6000 LVCs in more than 2000000 contexts. In contrast, the number of Lexical verbs in Persian is about 200. The comparison highlighted how significant this kind of linguistic resource could be for a language and its researchers. They can be used in machine translation, artificial intelligence and language processing programs, data recovery programs, language learning, grammar books, and dictionaries.

References

- Acquaviva, P. (2008). Roots and lexicality in distributed morphology. In A. Galani, D. Redinger and N. Yeo (Eds), *Special issues of York working papers in linguistics* (pp.1-21) New York: University of New York.
- AleAhmad, A., Amiri, H., Rahgozar, M., and Oroumchian, F. (2009). Hamshahri: A Standard Persian Text Collection. *Knowledge-Based Systems* 22(5), 382–387.
- Aronoff, M. (1976). *Word formation in generative grammar*. Cambridge, MA: MIT Press.
- Askariyan, N. (2012). *Automatic identification of Persian compound verbs*. Master thesis, University of Shiraz. [In Persian]
- Assi, S. M. (1997). Farsi linguistic database (FLDB). *International Journal of Lexicography* 10(3), 265.
- Barfi, V. (2016). Discovering the semantic space of Persian light verbs in the writing of Persian-foreign students from a cognitive point of view. Master thesis, University of Semnan. [In Persian]
- Beard, R. (1995). *Lexeme-morpheme base morphology*. New York: State University of New York Press.
- BijanKhan, M. (2004). The role of linguistic corpora in writing the grammar of language: An introduction to a computer software. *Journal of Linguistics* 19(2), 48-67. [In Persian]
- Bijan Khan, M. (2018). *Native search engine project*. Tehran: Research Institute of Information and Communication Technology. [In Persian]
- Bonet, E. (1991). *Morphology after syntax: Pronominal clitics in Romance languages*. PhD dissertation, MIT.

¹ The corpus of Light Verb Constructions of Persian is available at <https://literature.ut.ac.ir/compound-verb/>



- Borer, H. (2003). Exo-skeletal vs. endo-skeletal explanations: Syntactic projections and the lexicon. In J. C. Moore and M. Polinsky (Eds.), *The nature of explanations in linguistic theory* (pp. 37-67). Chicago: Chicago University Press.
- Borer, H. (2013). The category of roots. In R. Alexiadou, H. Borer and F. Schafer (Eds.), *The syntax of roots and the roots of syntax* (pp.112-149). Oxford: Oxford University Press.
- DabirMoghaddam, M. (1997). Persian compound verb. *Journal of Linguistics* 23, 31-46. [In Persian]
- Embick, D., and Marantz, A. (2008). Architecture and blocking. *Linguistic Inquiry* 39, 1-53.
- Embick, D., and Noyer, R. (2001). Movement operations after syntax. *Linguistic Inquiry* 32, 555-595.
- Eshaghi, M., and Karimi-Doostan, G. (2021). The productivity of Persian light verbs. *Journal of Language Researches* 12, 1-28. [In Persian]
- Family, N. (2006). Explorations of semantic space: The case of light verb constructions in Persian. PhD dissertation, Ecole des Hautes Etude en Sciences Sociales.
- Folli, R., Harley, H., and Karimi, S. (2005). Determinant of event type in Persian complex predicates. *Lingua* 115(10), 1365-1401.
- Goldberg, A. E. (1996). Words by default: Optimizing constraints and the Persian complex predicate. In D. Librik and R. Belear (Eds.), *Proceedings of Berkeley Linguistic Society* (pp. 132-146). Berkeley: Berkley University Press.
- Halle, M., and Marantz, A. (1993). Distributed morphology and the pieces of inflection. In K. Hale and S. J. Keyser (Eds.), *The view from building* (pp. 117-176). Cambridge: MIT Press.
- Halle, M. (1990). An approach to morphology. *North Eastern Linguistic Society* 20(1), 150-184.
- Harley, H. (2009). Compounding in Distributed morphology. In R. Lieber and P. Stekauer (Eds.), *Oxford Handbook of Compounding* (pp. 129-144). Oxford: Oxford University Press.
- HasaSedighi, P. (2010). *Teaching Persian to non-Persian speakers: problems and solutions*. Master thesis, Alame Tabatabaei University. [In Persian]
- Jespersen, O. (1965). *A modern English grammar on historical principles*. London: George Allen and Unwin Ltd.
- Karimi, S. (1997). Persian complex verbs: Idiomatic or compositional. *Lexicology* 3, 273-318.
- Karimi-Doostan, G. (1997). Light verb constructions in Persian. PhD dissertation, University of Essex.
- Karimi-Doostan, G. (2005). Light verb and structural case. *Lingua* 115(12), 1737-1756.
- Karimi-Doostan, G. (2008). Event structure of verbal nouns and light verbs. In S. Karimi, V. Samiian and D. Stilo (Eds.), *Aspects of Iranian linguistics* (pp.206-226). New Castle: Cambridge Scholars Publishing.
- Karimi-Doostan, G. (2011). Separability of light verb constructions in Persian. *Studia Linguistica* 65(1), 70-95.
- Khazaeifar, A. (2005). Transation theory: today and in the past. *Academy of Persian Language and literature journal* 28, 69-79. [In Persian]
- Kiparsky, P. (1982). Lexical morphology and phonology. In S. Yang (Ed.), *Linguistics in the Morning Calm* (pp. 3-91). Seoul: Hansin.
- Kiparsky, P. (1997). Remarks on denominal verbs. In A. Alsina, J. Bresnan and P. Sells (Eds.), *Argument Structure* (pp. 473-499). Stanford: Center for the Study of Language and Information.
- Marantz, A. (1997). No escape from syntax: Don't try morphological analysis in the privacy of your own lexicon. In A. Dimitriadis, L. Siegel, C. Surek-Clark, & A. Williams (Eds.), *University of Pennsylvania working papers in linguistics* (pp. 201-225). Philadelphia: University of Philadelphia.
- Marantz, A. (2000). *Roots: The universality of roots and pattern morphology*. Presented at the *Conference on Afro-Asiatic Language*. Paris University.
- Marantz, A. (2013). Verbal argument structure: Events and participants. *Lingua* 130, 152-168.
- Marefat, F. (2005). Written errors of Kurdish learners of Persian: A case of Mahabadi dialect. *Literary Text Research* 9(26), 10-37. [In Persian]
- Megerdooian, K. (2001). Event structure and complex predicates in Persian. *Canadian Journal of Linguistics* 46, 97-125
- Natel Khanlari, P. (1986). *The history of Persian language* (Vol. 2). Tehran: Nashrenow.[In Persian]
- Noyer, R. (1997). *Features, positions and affixes in autonomous: Morphological structure*. New York, NY: Garland.
- Panagiotidis, P. (2015). *Categorial features: A general theory of word class categories*. Cambridge: Cambridge University Press.
- Pestesty, D. (1982). Complementizer-trace phenomena and the nominative island condition. *The linguistic review* 1(3), 297-344
- Pestesty, D. (1995). *Zero syntax: Experiencers and cascades*. Cambridge MA: The MIT Press.
- Rasooli, M. S., Kouhestani, M., and Moloodi, A. S. (2013). Development of a Persian syntactic dependency treebank. In H. Hua, J. Lin, & A. Lopez (Eds), *Proceedings of the 2013 Conference of the North American Chapter of Association for Computational Linguistics: Human Language Technologies* (pp. 306-314). Atlanta: Association for Computational Linguistics.




- Samvelian, P., and Faghiri, P. (2013). Persian complex predicates: How compositional are they? *Semantics-Syntax Interface* 1, 43-74.
- Shamsfard, M., Hesabi, A., Fadaei, H., Mansoory, N., Noor, P., Famian, A., Bagherbeigi, S., Fekri, E., and Monshizadeh, M. (2010). Semiautomatic development of Farsnet; the Persian wordnet. *Proceedings of 5th Global WordNet Conference* 9(2), 35-44.
- Siddiqi, D. (2009). *Syntax within word: Economy, allomorphy, and argument selection in Distributed Morphology*. Amsterdam: John Benjamins.
- Tabibzade, O. (2005). Dictionary and grammar writing. *Academy of Persian Language and literature journal* 28, 69-79. [In Persian]
- Taslimipoor, S. (2012). *Automatic semantic processing of Persian compound verbs*. Master thesis, University of Shiraz. [In Persian]
- Vahedi Langrudi, M. (1996). *The syntax, semantics and argument structure of complex predicates in modern Farsi*. PhD dissertation, University of Ottawa



مقاله پژوهشی

پیکره ساخت‌های فعل سبک زبان فارسی

*مهديه اسحاقی 

**غلامحسین کریمی‌دوستان 

چکیده

پیکره زبانی مجموعه‌ای بزرگ از داده‌های زبانی مبتنی بر کاربرد سخنوران زبان‌هاست که الگوهای واقعی کاربرد زبانی را در اختیار پژوهشگران قرار می‌دهند. برتری پیکره‌ها در مقایسه با سایر منابع داده‌ای علاوه بر حجم زیاد داده، ایجاد امکان به کارگیری رایانه در بررسی‌های زبانی است. مقاله حاضر به معرفی اولین پیکره ساخت‌های فعل سبک زبان فارسی می‌پردازد. آشنایی با ماهیت این ساخت‌ها و دسترسی به فهرستی از آن‌ها، علاوه بر اهمیت نظری به لحاظ کاربردی نیز حائز اهمیت است. این یافته‌ها در حوزه بررسی‌های هوش مصنوعی مرتبط با پردازش زبان‌های طبیعی، ترجمه ماشینی، آموزش زبان فارسی، دست‌نویسی و فرهنگ‌نگاری کاربرد می‌یابد. پیکره هدف این پژوهش «پیکره زبانی ساخت‌های فعل سبک زبان فارسی» یا LCP نام دارد. برای ایجاد آن پیکره تک‌زبان پژوهشگاه ارتباطات و فناوری اطلاعات (بی‌جن‌خان، ۱۳۹۷) که حاوی ۹۵۰۰۰۰ فایل متنی است، به‌عنوان پیکره مبنای برگزیده شد. ساخت‌های فعلی مرکب مربوط به ۲۱ فعل سبک زبانی زبان فارسی از آن استخراج شده است و پس از برجسب‌زنی در چارچوب صرف توزیعی (Marantz, 2013Halle & Marantz, 1993)؛ در قالب پیکره‌ای مشتمل بر بیش از ۶۰۰۰ ساخت فعل سبک در بیش از ۲۰۰۰۰۰۰ بافت زبانی ارائه شده است که در بیش از ۲۰۰۰۰۰۰ بافت زبانی ارائه شده‌اند. مقایسه تعداد فعل‌های واژگانی زبان فارسی با تعداد ساخت‌های فعل سبک موجود در پیکره حاضر، بدیهی‌ترین عاملی است که وجود چنین پیکره‌ای در میان منابع زبان فارسی را ضرورت می‌بخشد. از سوی دیگر، ماهیت این پیکره، یعنی نمایش ساخت‌های فعل سبک در بافت‌های زبانی متفاوت، می‌تواند به پژوهشگران در یافتن پاسخ پرسش‌های موجود در رابطه با این ساخت‌ها، رد یا تأیید فرضیه‌ها و طرح نظریه‌های جدید کمک کند.

کلیدواژه‌ها: زبان فارسی، منابع داده‌ای، پیکره زبانی، ساخت‌های فعل سبک، پردازش زبان طبیعی



۱. مقدمه

زبان از ارکان اصلی جوامع بشری و ابزار بیان هویت هر ملتی است؛ از این رو، کوشش در زمینه شناخت و حفظ آن ضرورت می‌یابد. پیکره‌های زبانی از جمله ابزارهای کارآمد زبانی هستند که نقش مؤثری در حفظ و شناخت زبان‌ها بر عهده دارند. مقاله حاضر به معرفی پیکره زبانی ساخت‌های فعل سبک زبان فارسی می‌پردازد که حاصل پژوهش نگارندگان مقاله است.^۱ پژوهشی که در آن تلاش شده است با ایجاد این ابزار زبانی راه را برای شناخت بهتر یکی از اجزای چالش‌برانگیز زبان، یعنی ساخت‌های فعل سبک،^۲ هموار گردد. فعل سبک به دسته‌ای از فعل‌ها اطلاق می‌شود که در روند تحولات زبانی بخشی از بار معنایی خود را از دست داده‌اند و به اصطلاح سبک شده‌اند، این اصطلاح را اولین بار **یسپرسن**^۳ (۱۹۶۵) ابداع کرد. این فعل‌ها در بیان مفاهیم فعلی با عنصر زبانی دیگری از مقوله اسم، صفت یا حرف اضافه هم‌نشین شده و ساخت‌های فعلی مرکب را ایجاد کرده‌اند.^۴ مطالعه نحوی و معنایی این ساخت‌ها به یکی از چالش‌برانگیزترین مسائل مورد بررسی زبان‌شناسی امروز تبدیل شده است و در زبان‌های مختلف از جنبه‌های مختلف و در قالب رویکردهای متفاوت به آن‌ها پرداخته شده است. زبان فارسی از جمله زبان‌هایی است که در آن تعداد کمی از فعل‌ها ساده هستند؛ **ناتل خانلری** (۱۳۶۵) به وجود حدود ۲۷۹ فعل ساده واژگانی در زبان فارسی اشاره کرده است.^۵ **کریمی**^۶ (۱۹۹۷) به وجود ۱۱۵ فعل ساده واژگانی و **کریمی دوستان** (۱۹۹۷) به وجود ۱۵۰ فعل ساده واژگانی سبک‌نشده اشاره کرده‌اند. سایر افعال مورد استفاده در این زبان حاصل کنار هم قرار گرفتن یک پیش فعل و یک فعل سبک هستند، این ویژگی ضرورت پرداختن به این ساخت‌های فعلی را دو چندان ساخته است.

ساخت‌های فعلی سبک زبان فارسی به دلیل ماهیت ویژه‌ای که دارند، از جمله ساخت‌های زبانی هستند که سبب ایجاد چالش در حوزه‌های مختلف از جمله آموزش زبان، فرهنگ‌نگاری، دستورنویسی، هوش مصنوعی، ترجمه ماشینی، ایجاد جویشتگرهای بومی و تحلیل‌های زبان‌شناختی داده‌محور شده‌اند. **تسلیمی پور** (۱۳۹۱) به چالش‌های موجود در برخورد با ساخت‌های فعلی مرکب در سیستم‌های محاسباتی اشاره کرده است. وی بر کمبود منابع محاسباتی زبان فارسی و ضرورت بررسی‌های پیکره‌ای برای قرار دادن احتمالات معیار در اختیار پژوهشگران حوزه هوش مصنوعی تأکید کرده است. او استخراج خصوصیات معنایی و ساختاری ساخت‌های فعل سبک را برای استفاده در عملیات‌های پردازش زبانی مانند ترجمه، برچسب‌گذاری، خلاصه‌برداری ضروری دانسته است. **عسکریان** (۱۰:۱۳۹۱) نیز اولین گام در پردازش زبانی را شناخت اجزای پایه‌ای زبان، یعنی واژه‌ها، دانسته است و در این میان به چالش‌های موجود در رابطه با واژه‌های مرکب اشاره کرده است. وی دلیل وجود این چالش‌ها را ویژگی‌های نحوی و معنایی خاص در این دسته از واژه‌ها می‌داند که حاصل ترکیب ویژگی‌های اجزای تشکیل‌دهنده آن‌هاست. ویژگی‌هایی که برای هر نوع سیستمی که به نوعی با فهم زبان در ارتباط است مانند سیستم‌های ترجمه، خلاصه‌برداری و سیستم‌های محاوره مورد نیاز است. افزون بر این، یافته‌های حاصل از چنین پژوهش‌هایی می‌تواند در رفع مشکلات مربوط به آموزش ساخت‌های فعل سبک در دستور زبان مدارس و آموزش زبان فارسی زبانان نیز مؤثر واقع

^۱ این پیکره حاصل پژوهشی است که به عنوان طرح پسادکتری با حمایت معاونت علمی و فناوری ریاست جمهوری، صندوق حمایت از پژوهشگران و فناوران کشور با شماره 99030152 در دانشگاه تهران انجام شده است. پیکره مذکور در وبگاه دانشکده ادبیات و علوم انسانی دانشگاه تهران به آدرس <https://Literature.ut.ac.ir/compound-verb> بارگذاری شده است.

^۲ به ساخت‌های فعل سبک، فعل مرکب هم گفته می‌شود.

^۳ O. Jespersen

^۴ لازم به ذکر است که پیکره حاضر دسته خاصی از گزاره‌های مرکب با عنوان ساخت‌های فعل سبک را مدنظر قرار داده است و در شناسایی آن‌ها پیش از هر معیار دیگری سبک‌شدگی معنایی آن‌ها را مدنظر قرار داده است، ویژگی‌ای که در سایر دسته‌بندی‌ها از جمله فعل‌های مرکب انضمامی از جمله «غذا خوردن»، «ماهی گرفتن»، «زمین نشستن» و غیره دیده نمی‌شود. ^۵ ناتل خانلری در این فهرست تمامی فعل‌های ساده واژگانی زبان فارسی را آورده است اعم از آن‌ها که دستخوش سبک‌سازی شده‌اند و آن‌ها که صورت واژگانی را حفظ کرده‌اند.

برخی از افعال در فهرست خانلری در فارسی امروزی کاربرد ندارند.

^۶ S. Karimi

شود و در امر تهیه و تدوین مواد آموزشی مرتبط با این ساخت‌ها تأثیرگذار باشد. مطالعات انجام‌شده در زمینه مشکلات آموزش زبان فارسی به غیرفارسی‌زبانان به کرات به وجود مشکلاتی در رابطه با آموزش ساخت‌های فعلی مرکب اشاره کرده‌اند؛ برای مثال، معرفت (۱۳۸۴) به انتخاب جز فعلی نامناسب به‌عنوان یکی از خطاهای مشاهده‌شده در کاربرد زبان فارسی از سوی فارسی‌آموزان غیرفارسی‌زبان اشاره کرده است. حساس صدیقی (۱۳۸۹) نیز به مشکلات غیرفارسی‌زبانان در رویارویی با ساخت‌های فعلی مرکب پرداخته است. از دیگر پژوهش‌های این حوزه می‌توان به برفی (۱۳۹۵) اشاره کرد که از دیدگاه شناختی به مشکلاتی پرداخته که غیرفارسی‌زبانان سه مرکز آموزشی زبان فارسی در رویارویی با این ساخت‌ها با آن‌ها مواجه بوده‌اند. در حوزه فرهنگ‌نگاری نیز بسیاری از نویسندگان در پی یافتن راه‌حلی برای چگونگی مدخل کردن چنین الگوهایی در فرهنگ‌ها هستند. خزاعی‌فر (۱۳۸۴) و طیب‌زاده (۱۳۸۴) از جمله پژوهشگرانی هستند که به مشکلات مربوط به فعل مرکب در فرهنگ‌نگاری پرداخته‌اند.

مشکلاتی از این دست که نتیجه کمبود داده‌های موجود از این ساخت‌های زبانی است، نگارندگان مقاله را بر آن داشت که برای سهولت بخشیدن به انجام پژوهش‌های این حوزه و کمک به گشایش رمزی از رمزهای زبان فارسی به ایجاد اولین پیکره از این ساخت‌های زبانی زایا پردازند. در دست داشتن یک پیکره زبانی حاوی این ساخت‌ها به‌صورت نمونه‌های واقعی که الگوهای رفتاری آن‌ها را هم به‌لحاظ معنایی و هم به‌لحاظ نحوی به تصویر می‌کشد، ابزاری خواهد بود به‌سوی شناخت بهتر این ساخت‌های فعلی و رفع چالش‌های موجود در رابطه با آن‌ها.

ساخت‌های فعلی سبک در زبان فارسی نیز مانند سایر زبان‌ها مورد توجه بوده‌اند، از جمله مهمترین آثار موجود در این حوزه می‌توان به دبیرمقدم (۱۳۷۶)؛ کریمی (۱۹۹۷)؛ واحدی لنگرودی (۱۹۹۶)؛ گلدبرگ^۱ (۱۹۹۶)؛ کریمی‌دوستان (۲۰۱۱، ۲۰۰۸، ۲۰۰۵، ۱۹۹۷)؛ مگردومیان^۲ (۲۰۰۱)؛ فولی،^۳ هارلی^۴ و کریمی (۲۰۰۵)؛ فامیلی^۵ (۲۰۰۶)؛ سامولیان^۶ و فقیری^۷ (۲۰۱۳) اشاره کرد، که از میان آن‌ها تنها سامولیان و فقیری (۲۰۱۳) به تهیه پیکره‌ای مربوط به بیش از ۶۰۰ فعل مرکب شامل فعل سبک «زدن» پرداخته و آن را واژگان نحوی و معنایی افعال مرکب فارسی (PersPred) نامیده‌اند. این مجموعه چندزبانه شامل اطلاعات نحوی و معنایی افعال مرکب زبان فارسی با جز فعلی «زدن»، ترجمه انگلیسی و فرانسوی افعال و حداقل یک جمله مثال برای هر فعل است. در زبان فارسی پیش از این پیکره‌هایی از جمله پیکره متنی زبان فارسی (بی جن خان، ۱۳۸۳)، پایگاه دادگان زبان فارسی (Assi, 1997)، پیکره همشهری (AleAhmad et al., 2009) و پیکره فارس نت (Shamsfard et al., 2010) ساخته شده‌اند. اما پیکره‌ای که این پژوهش در پی ساخت آن بوده است، تنها یک نوع خاص از ساخت‌های زبانی به نام ساخت‌های فعلی سبک در زبان فارسی را هدف قرار داده است.

پیکره حاصل از پژوهش یک پیکره همزمانی تک‌زبانه از ساخت‌های فعلی سبک در زبان فارسی با قابلیت جستجوی رایانه‌ای است. داده‌های موجود در این پیکره زبانی مستخرج از پیکره تک‌زبانه پژوهشگاه ارتباطات و فناوری اطلاعات است که توسط بی‌جن خان (۱۳۹۷) برای طرح جویشگر بومی مرکز تحقیقات مخابرات ایران ایجاد شده است و حاوی ۹۵۰۰۰۰ فایل متنی است. ساخت‌های فعلی مرکب مربوط به ۲۱ فعل سبک زبانی فارسی (شدن، کردن، خوردن، بردن، آوردن، آمدن، انداختن، افتادن، گرفتن، دادن، بخشیدن، رفتن، رسیدن، گشتن، کشیدن، یافتن، دیدن، گذاشتن، بستن، زدن و داشتن) از این پیکره استخراج و در قالب جمله‌هایی که این ساخت‌های فعلی در آن‌ها به کار رفته است، با عنوان اولین پیکره ساخت‌های فعلی مرکب زبان فارسی (LCP)^۸ ارائه می‌شود.

¹ A. E. Goldberg

² K. Megerdooonian

³ R. Folli

⁴ H. Harley

⁵ N. Family

⁶ P. Samvelian

⁷ P. Faghiri

⁸ Light Verb Constructions of Persian

برچسب‌زنی این ساخت‌ها در چارچوب نظری صرف‌توزیعی^۱ صورت گرفته است. صرف‌توزیعی چارچوبی نظری است که در دهه ۹۰ ظهور یافت، از منادیان این نحله نظری می‌توان به هله^۲ (۱۹۹۰)، بونت^۳ (۱۹۹۱)، نویر^۴ (۱۹۹۷) و پستسکی^۵ (۱۹۹۵) اشاره کرد؛ اما اولین بار هله و مرتنز^۶ (۱۹۹۳) عنوان صرف‌توزیعی را برای این انگاره دستوری مطرح کردند. صرف‌توزیعی نظریه‌ای مختص صرف نیست و این نام از این روی بر آن نهاده شده است که وظیفه ساخت واژه بین بخش‌های مختلف انگاره دستوری توزیع شده است. این رویکرد مبتنی بر تعامل میان بخش‌های مختلف دستور از جمله صرف، نحو و واج‌شناسی است. تنها یک بخش زیاده در آن مفروض است و آن نحو است، ساختار واژه‌ها نیز مانند ساختار گروه و جمله در نحو شکل می‌گیرد. انگاره صرف‌توزیعی در تقابل با فرضیه‌های واژگان‌گرا بر این فرض استوار است که واژه‌ها با استفاده از ریشه‌ها و مشخصه‌های صوری در نحو تولید می‌شوند، نه اینکه به صورت واحدهای پیش‌ساخته درون‌داد نحو شوند. بنابراین، در این رویکرد اشتقاق با عملکرد فرایندهای نحوی بر مجموعه‌ای از مشخصه‌های صرفی نحوی و ریشه‌ها در حوزه نحو آغاز می‌شود و سپس، در سطح بازنمون^۷ که کار نحو پایان یافته است اشتقاق در دو مسیر صورت‌آوایی^۱ و صورت منطقی^۸ ادامه می‌یابد. از دیدگاه صرف‌توزیعی ریشه‌ها^۹ عناصری بدون مقوله در نظر گرفته می‌شوند. صدیقی (۲۰۰۹) ریشه‌ها را تکواژهای انتزاعی بدون مقوله و دارای مفهومی بنیادی دانسته‌اند.

درباره ماهیت صوری ریشه‌ها، عده‌ای از جمله بیرد^{۱۰} (۱۹۹۵) برای ریشه‌ها مانند واژه‌های انتزاعی اساساً ماهیت نحوی-معنایی قائل شده‌اند. آرونف^{۱۱} (۱۹۷۶) ماهیت واجی برای آن‌ها در نظر گرفت است. پستسکی (۱۹۸۲) و کیپارسکی^{۱۲} (۱۹۸۲; ۱۹۹۷) آن‌ها را ماهیتاً ترکیبی از ویژگی‌های واجی و معنایی فرض کرده‌اند. مرتنز (۱۹۹۷) به ماهیت نحوی آن‌ها پرداخته است. برر^{۱۳} (۲۰۰۳) ریشه‌ها را فاقد ویژگی‌های مؤثر بر ساختار فرض کرده است.

از دیگر مباحث مطرح‌شده درباره ریشه‌ها، میزان بار معنایی آن‌ها و نقش این بار معنایی در تعیین رفتار نحوی است. از جمله دیدگاه‌های مطرح‌شده در این زمینه دیدگاه پژوهشگرانی از جمله مرتنز (۲۰۰۰; ۱۹۹۷) و هارلی^{۱۴} (۲۰۰۹) است که معتقدند ریشه‌ها قادر به انتخاب موضوع هستند و از این طریق اطلاعاتی در رابطه با بافت نحوی خود ارائه می‌دهند. هارلی (۲۰۰۹) ریشه‌ها را عناصری بدون مقوله اما حاوی محتوای معنایی دایره‌المعارفی پیام دانسته و معتقد است هر ریشه صورت‌واژگانی شده یک مفهوم واژگانی محض است. او ریشه‌ها را دارای خوانش‌های وابسته به بافت می‌پندارد.

در مقابل عده‌ای دیگر از جمله آکوایوا^{۱۵} (۲۰۰۸) و برر (۲۰۱۳) معتقدند ریشه‌ها حاوی اطلاعات معنایی نیستند و این عدم وجود بار معنایی به معنی عدم وجود ساختار موضوعی و هر ویژگی‌گزینشی دیگری است. برر (۲۰۱۴) نیز ریشه‌ها را بدون مقوله ذاتی در نظر گرفته و از این حیث برداشت خود از ریشه‌ها را مشابه انگاره صرف‌توزیعی دانسته است.

¹ Distributed morphology

² M. Halle

³ E. Bonet

⁴ R. Noyer

⁵ D. Pestesky

⁶ A. Marantz

⁸ spell out

⁹ logical form (LF)

¹⁰ roots

¹⁰ R. Beard

¹¹ M. Aronoff

¹² P. Kiparsky

¹³ H. Borer

¹⁴ H. Harley

¹⁵ P. Acquaviva

از دیگر نخست‌های نحو در چارچوب صرف توزیعی به تعبیر ایمبیک^۱ و نویر (2001) عناصر نقشی یا مقوله‌سازها هستند که مسئولیت مقوله‌دار کردن ریشه‌های بدون مقوله را برعهده دارند. فرض مقوله‌سازی ایمبیک و مرتنز (2008) بیانگر ضرورت وجود هسته‌های نقشی مقوله‌ساز در این چارچوب است. طرح کلی رویکرد مرتنز این بود که مقوله‌های واژگانی مانند اسم و فعل حاصل ترکیب مشخصه‌های مقوله‌ای با ریشه‌ها در واژگان نیستند، بلکه ریشه‌ها بدون مقوله وارد نحو می‌شوند و محیط نحوی است که مشخص می‌کند این ریشه‌ها در جایگاه متمم کدام مقوله‌ساز جای گیرند و بر آن اساس مقوله آن‌ها تعیین گردد (Panagiotidis, 2015).

عناصر مقوله‌ساز در برخی موارد تظاهر آشکار دارند و در برخی موارد فاقد تظاهر آشکارند یا به عبارتی تهی هستند.

پاناگیوتیدس (2015) معتقد است که ریشه‌ها نه تنها بدون مقوله هستند، بلکه به لحاظ معنایی نیز کامل نیستند؛ از این رو، مقوله‌سازها علاوه بر اینکه ریشه‌ها را مقوله‌دار می‌کنند، به خوانش معنایی آن‌ها نیز کمک می‌کنند.

در برجسب‌زنی ساخت‌های فعل سبک موجود در پیکره به پیروی از مرتنز (2013) پیش‌فعل‌های به کاررفته در ساخت‌های فعل سبک را ریشه‌های بی‌مقوله و فعل‌های سبک را عناصر مقوله‌ساز در نظر می‌گیریم.

بر این اساس، پس از مقدمه حاضر در بخش دوم به معرفی مختصر مراحل ایجاد پیکره مورد بحث می‌پردازیم و در بخش سوم با ارائه نمونه‌هایی به معرفی داده‌هایی می‌پردازیم که این پیکره در اختیار کاربران قرار می‌دهد. در بخش چهارم به جمع‌بندی مطالب خواهیم پرداخت.

۲. مراحل ایجاد پیکره ساخت‌های فعلی مرکب

مراحل ایجاد این پیکره عبارت‌اند از: طراحی ساختار پیکره، گزینش پیکره مبنا، پیش‌پردازش متن‌ها، تعیین گره‌های جستجو، تصمیم‌گیری درباره بافت چپ و راست گره‌ها، نوشتن کدهای ماکرو و تهیه نرم‌افزار جستجو، استخراج تمام جملات حاوی فعل‌های مورد بررسی (فارغ از در نظر گرفتن کاربرد سبک یا واژگانی آن‌ها)، جداسازی صورت‌های واژگانی و سبک فعل‌ها، انتخاب مدل برجسب‌زنی و انجام فرایند برجسب‌زنی، ساماندهی نتایج و ایجاد امکان جستجو که در زیربخش‌های بعدی به‌طور جداگانه به چگونگی انجام هر یک از این مراحل می‌پردازیم.

۲-۱. طراحی ساختار پیکره

اولین مرحله در ایجاد یک پیکره زبانی طراحی ساختار پیکره است که در آن مشخص می‌شود که پیکره از نوع تک‌زبانه است یا چندزبانه، همزمانی است یا در زمانی، گفتاری است یا نوشتاری، شامل متن است یا جمله، محدود به موضوعی خاص است یا خیر، آیا پیکره‌ای کلی است یا هدف‌ویژه و در آخر اینکه نحوه دسترسی به آن چگونه خواهد بود. در طراحی ساختار پیکره حاضر بنا شد که پیکره به صورت یک پیکره همزمانی تک‌زبانه به زبان فارسی باشد، پیکره‌ای نوشتاری و شامل جملاتی حاوی ساخت‌های فعل سبک مربوط به ۲۱ فعل سبک در زبان فارسی. یک پیکره هدف‌ویژه به منظور ارائه داده‌ای گسترده از ساخت‌های فعل سبک زبان فارسی در بافت‌های مختلف زبانی که به صورت پیکره الکترونیکی با امکان جستجوی ماشینی در اختیار کاربران قرار گیرد.

۲-۲. گزینش پیکره مبنا و پیش‌پردازش متن‌ها

در تهیه یک پیکره به‌ویژه یک پیکره هدف‌ویژه می‌توان داده‌ها را از یک پیکره موجود در زبان استخراج کرد. چنین پیکره‌ای، پیکره مبنا نامیده می‌شود. پیکره‌ای که در ایجاد پیکره حاضر به‌عنوان مبنا در نظر گرفته شده است، پیکره تک‌زبانه پژوهشگاه ارتباطات و فناوری

¹ A. Embick

اطلاعات است که شامل ۹۵۰ هزار فایل متنی است. این پیکره توسط بی‌جن‌خان (۱۳۹۷) برای طرح جویشگر بومی مرکز تحقیقات مخابرات ایران ایجاد شده است و یک فایل اکسس شامل ۹۵۰ هزار فایل متنی است، وسعت این پیکره و تنوع متن‌ها این امکان را ایجاد می‌کند که حجم وسیعی از داده‌های زبانی مرتبط با ساخت‌های فعلی مدنظر از آن استخراج شود. هرچه پیکره بزرگتر باشد ویژگی نمایندگی^۱ خود را بهتر ایفا می‌کند و الگوهای زبانی را با دقت بیشتری به تصویر می‌کشد. پس از انتخاب پیکره مبنا و پیش از آغاز جستجو در نخستین گام برای نمایش داده‌ها و ایجاد امکان جستجو فایل اکسس پیکره را به ۱۹ فایل اکسل شکستیم که هر یک از این ۱۹ فایل شامل ۵۰ هزار فایل متنی است.

سپس داده‌ها به منظور شناسایی موارد نیازمند پیش‌پردازش به‌طور اجمالی بررسی شد. پیش‌پردازش یا نرمال‌سازی در واقع آماده‌سازی داده‌ها و ایجاد تطابق نمایش آن‌ها آغاز مرحله جستجو است. از جمله موارد نیازمند نرمال‌سازی وجود «ی» و «ک» عربی بود که در امر جستجو اختلال ایجاد می‌کرد برای حل این مشکل، برنامه یکسان‌سازی فونت‌ها به نرم‌افزار اکسل داده شد و «ی» و «ک» عربی با معادل فارسی خود جایگزین شدند. نرمال‌سازی فاصله‌ها نیز صورت گرفت. اما از اعمال پیش‌پردازش بن‌واژه‌سازی بر داده‌های پیکره خودداری کردیم. به این دلیل که ماهیت بن‌واژه‌های فعل‌های زبان فارسی به‌ویژه بن‌واژه‌های زمان حال به‌گونه‌ای است که در برخی واژه‌ها توالی‌های واجی مشابه آن‌ها یافت می‌شوند و سبب می‌شود که ابزار جستجو یافته‌های نامربوط فراوان را در نتایج ارائه دهد، در مقابل به جستجوی واژه‌ها به‌صورت توکن‌ها (موردواژه‌ها) پرداختیم.

۲-۳. تعیین گره‌های جستجو

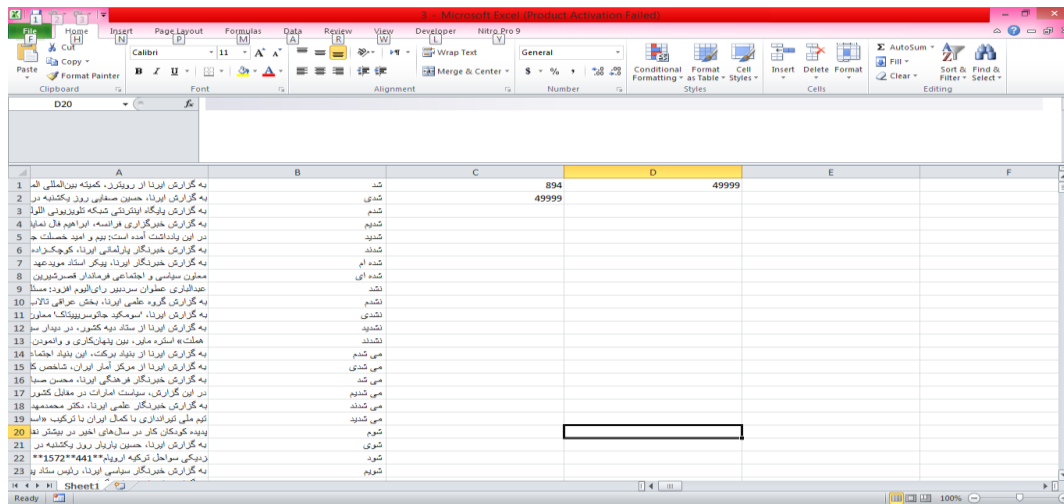
پس از آماده‌سازی داده‌ها برای جستجو وارد مرحله تعیین گره‌های جستجو می‌شویم. نظر به اینکه در این پژوهش با ساخت‌های فعل سبک روبه‌رو هستیم که متشکل از پیش‌فعل و فعل سبک هستند، جستجو برای این ساخت‌ها در دو مرحله صورت می‌گیرد.

۲-۳-۱. گره‌های جستجوی مرحله اول

در مرحله اول صورت‌های تصریف‌شده^{۲۱} فعل مورد بررسی پژوهش به‌لحاظ زمان، شخص، جهت و نمود را به‌عنوان گره‌های جستجو تعیین کردیم (تصویر ۱). آنچه ما را ناگزیر از این ساخت که تک‌تک صورت‌های تصریف‌شده فعل‌ها را به‌عنوان گره جستجو در نظر بگیریم جلوگیری از مواجهه با یافته‌های نامربوط در برون‌داد جستجو و محدود کردن نتایج جستجو بود. از این رو، موردواژه‌ها^۲ را با تعریف فاصله قبل و بعد آن‌ها به‌عنوان گره جستجو در نظر گرفتیم تا برنامه تنها جمله‌های حاوی همان واژه را جستجو کند. برای روشن شدن موضوع به این مثال توجه کنید؛ اگر بن ماضی فعل «شدن»، یعنی «شد»، را به مثابه صورت مشترک در همه تصریف‌های ماضی بدون فاصله قبل و بعد به‌عنوان گره جستجو تعریف کنیم تا صورت‌هایی از این فعل را که شامل این توالی واجی هستند از جمله «شده، شدی، شدم و نظایر آن» را بیابد، گرچه از تعداد گره‌های جستجو کاسته می‌شود، چندین برابر بر برون‌داد جستجو افزوده می‌شود و واژه‌هایی از جمله «رشد، مرشد، شداند و مانند آن» نیز در نتایج جستجو استخراج می‌شود که با حجم داده‌ای که با آن مواجه هستیم جداسازی آن‌ها بسیار مشکل‌ساز خواهد بود.

¹ representativeness

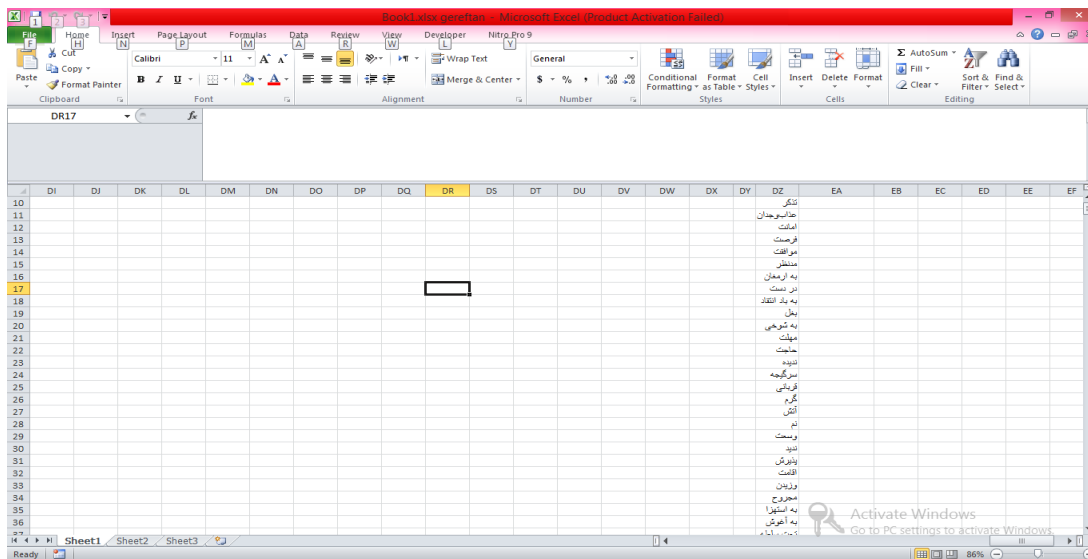
² tokens



تصویر ۱- نمونه‌ای از گره‌های جستجوی مرحله اول
Picture1- A sample of the first step search nodes

۲-۳-۲. گره‌های جستجوی مرحله دوم

گره‌های جستجوی مرحله دوم را پیش‌فعل‌های هم‌نشین با هریک از فعل‌ها تشکیل می‌دهند که پس از اتمام مرحله اول جستجو در فایل‌های مربوط به هریک از فعل‌ها به صورت غیرخودکار و در چند مرحله توسط پژوهشگر استخراج شده‌اند (تصویر ۲).



تصویر ۲- نمونه‌ای از گره‌های جستجوی مرحله دوم
Picture2- A sample of the second step search nodes

۲-۴. تعیین بافت چپ و راست گره‌ها و نوشتن الگوریتم‌های ماکرو و تهیه نرم‌افزار جستجو

بافت چپ و راست گره‌های جستجو در جستجوهای پیکره‌ای با توجه به هدف جستجو تعیین می‌شود. ما در تعیین بافت چپ و راست گره‌های جستجو دو معیار را در نظر داشتیم: اول اینکه پیکره‌ای از ساخت‌های فعل سبک را در اختیار کاربران قرار دهیم که این ساخت‌های چالش‌برانگیز زبانی را در بافتی گسترده‌تر از واژه ارائه می‌دهد تا از این طریق علاوه بر در اختیار داشتن فهرستی از ساخت‌های فعل سبک،

امکان مشاهده الگوهای رفتاری این ساخت‌ها نیز وجود داشته باشد؛ معیار دوم اینکه ویژگی جدایی‌پذیری ساخت‌های فعل سبک سبب می‌شود گاهی میان پیش‌فعل و فعل سبک فاصله بیافتد؛ از این رو، برای استخراج این ساخت‌ها در مرحله اول که تنها فعل‌ها جستجو می‌شوند باید بافت راست واژه به گونه‌ای در نظر گرفته شود که پیش‌فعل آن، حتی در نمونه‌هایی هم که از فعل فاصله گرفته‌اند، در نتایج قابل مشاهده باشد. بر این اساس، و با نگاهی گذرا به داده‌های پیکره‌ای بافت راست هر گره ۱۸ و بافت چپ ۸ در نظر گرفته شد.

مرحله بعد پس از تصمیم‌گیری درباره گره‌های جستجو و تعیین بافت چپ و راست آن‌ها، تهیه نرم‌افزار جستجو بود. نرم‌افزار جستجو را به صورت برنامه‌های ماکرو نوشته شده به زبان ویژوال بیسیک در اکسل اجرا کردیم، یک برنامه برای جستجوی مرحله اول، یعنی جستجوی فعل‌ها و یک برنامه برای جستجوی پیش‌فعل‌ها. علت استفاده نکردن از نرم‌افزارهای جستجوی موجود مانند Antconc و LancsBox این بود که در این نرم‌افزارها امکان جستجوی چندین واژه به‌طور همزمان وجود ندارد و ما بر این بودیم که در هر یک از ۱۹ فایل با حجم گسترده‌ای که دارند (هر کدام ۵۰ هزار فایل متنی) همه موردواژه‌ها به‌طور همزمان جستجو شوند و نتایج جستجو در کاربرگ بعدی در ستون‌های اکسل با عنوان هر یک از فعل‌ها به نمایش درآید. از این رو، برنامه‌ای خاص استخراج این دسته از فعل‌ها تهیه شد.

۲-۵. استخراج تمام جملات حاوی فعل‌های مورد بررسی و جداسازی صورت‌های واژگانی و سبک

استخراج ساخت‌های فعل سبک از پیکره به آسانی امکان‌پذیر نیست. از این رو، طی دو مرحله به استخراج داده‌ها پرداختیم.

۱-۵-۲. استخراج فعل‌های مورد بررسی اعم از سبک و واژگانی

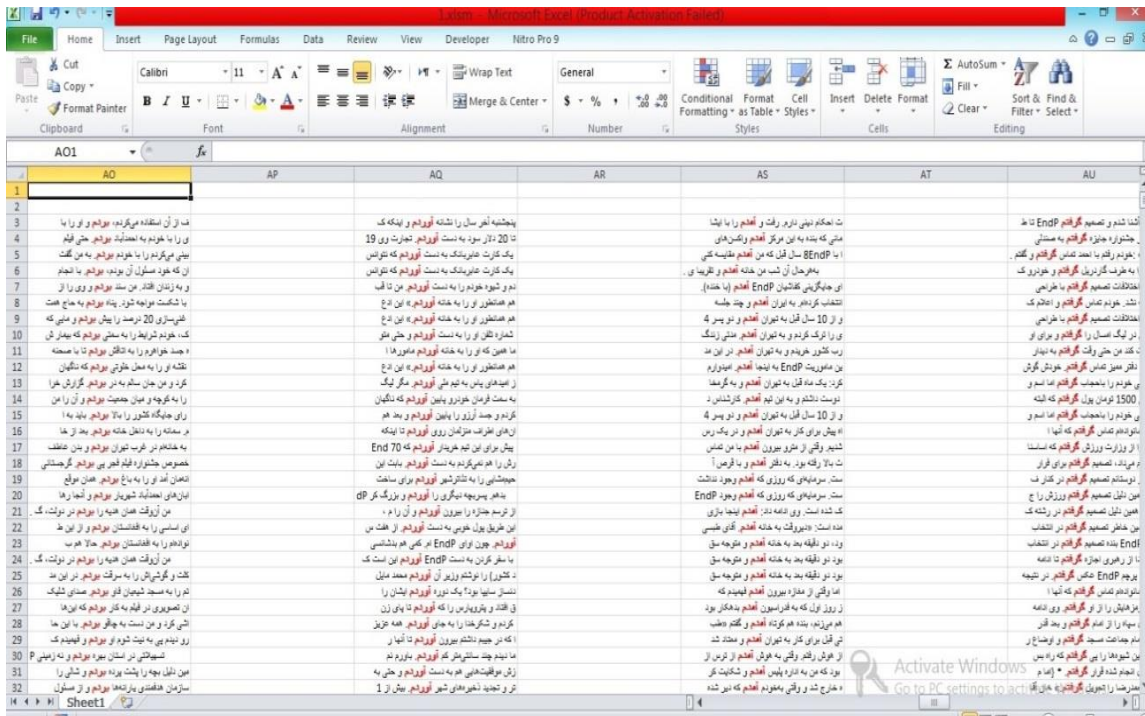
در این مرحله فعل‌های مدنظر پژوهش را، بدون در نظر گرفتن کارکرد سبک یا واژگانی آن‌ها، در هر یک از ۱۹ فایل جستجو کردیم. ستون اول هر فایل شامل ۵۰ هزار سطر است. لازم به ذکر است که هر یک از سطرهای این ستون، یک متن را شامل می‌شود نه یک جمله را. بنابراین، برنامه جستجو برای هر یک از ۱۹ فایل با ۵۰۰۰۰ فایل متنی مواجه است نه ۵۰۰۰۰ جمله (تصویر ۳).



تصویر ۳- نمونه‌ای از متن موجود در یک سطر از درون‌داد مرحله اول
 Picture3- A sample of the texts in one cell of the first step input

این متن‌ها درون‌داد برنامه جستجوی قرار گرفتند که در آن گره‌های جستجو موردواژه‌های مربوط به فعل‌های مدنظر پژوهش بودند

(صورت‌های تصریف‌شده هر فعل). پس از راه‌اندازی برنامه در هریک از فایل‌ها، نرم‌افزار به جستجوی موردواژه‌ها پرداخته و هریک را در ستونی با عنوان همان فعل ذخیره می‌سازد (تصویر ۴). برونداد این مرحله فهرستی از صورت‌های مختلف فعل‌هاست، فارغ از سبک یا واژگانی بودن آن‌ها.



تصویر ۴- نمونه‌ای از برونداد مرحله اول
 Picture4- A sample of the first step output

پس از این جستجو، نتایج جستجوی مربوط به هریک از ۲۱ فعل را در یک فایل اکسل یک‌جا کردیم. بدین ترتیب، ۲۱ فایل اکسل تهیه کردیم که در هر کدام نتایج جستجوی مرحله اول مربوط به یکی از فعل‌ها وجود داشت (تصویر ۵).

تصویر ۵- نمونه‌ای از فایل‌های مختص هر یک از فعل‌های مورد بررسی
 Picture 5- A sample of the files dedicated to each of the investigated verbs

۲-۵-۲. جداسازی صورت‌های سبک و واژگانی فعل‌ها

در این مرحله صورت‌های واژگانی و سبک فعل‌ها از هم جدا شد. برای تشخیص ساخت‌های فعل سبک فارغ از بحث ترکیب‌پذیری/ترکیب‌ناپذیری^۱ معنایی از معیار سازواری^۲ استفاده کردیم که براساس آن دو جزء سازنده فعل سبک تشکیل یک سازه را می‌دهند و این سازواری را در رفتار آن‌ها بررسی کردیم. بدین معنی که در مواردی که با ساخت فعل سبک مواجه هستیم کل سازه، یعنی فعل سبک و پیش فعل، در آزمون‌های تعیین سازواری صورت خوش ساخت را به دست می‌دهند. این آزمون‌ها شامل مبتداسازی، همپایه‌سازی، حذف و جانشینی^۳ هستند.

برای مثال، مبتداسازی در جمله «تیم والیبال ایران ۲۵ بر ۱۹ در این ست برتری یافت» با مبتداسازی پیش فعل و فعل سبک صورت خوش ساخت «برتری یافتن تیم والیبال ایران ۲۵ بر ۱۹ در این ست» را به دست می‌دهد و مبتداسازی فعل به تنهایی و بدون پیش فعل صورت بدساخت «یافتن تیم والیبال ایران ۲۵ بر ۱۹ در این ست برتری» را به دست می‌دهد. در همپایه‌سازی نیز هر دو عنصر سازنده ساخت فعل سبک برای ایجاد صورت خوش ساخت به کار می‌روند «تیم والیبال ایران ۲۵ بر ۱۹ در این ست برتری یافت و صعود کرد» در غیر این صورت با جمله بدساخت «تیم والیبال ایران ۲۵ بر ۱۹ در این ست برتری و صعود کرد» مواجه خواهیم شد. در حذف نیز هر دو عنصر با هم حذف می‌شوند «تیم والیبال ایران ۲۵ بر ۱۹ در این ست برتری یافت و ۲۳ بر ۲۲ در ست قبل». آزمون جانشینی نیز نشان از سازواری پیش فعل و فعل سبک دارد «تیم والیبال ایران ۲۵ بر ۱۹ در این ست شکست خورد».

در ساخت‌های فعل سبک زبان فارسی رفتارهای زبانی متنوعی یافت می‌شود و از ددرسازترین رفتارهای قابل مشاهده در این ساخت‌ها جدایی‌پذیری آن‌هاست. از آنجا که رفتار معناداری حداقل تا زمان انجام پروژه حاضر درباره نوع و نحوه قرار گرفتن پیش فعل‌های موجود در ساخت‌های فعل سبک زبان فارسی در دسترس نبود، این مرحله به‌عنوان یکی از مشکل‌ترین مراحل انجام پروژه حاضر به صورت

¹ compositionality/noncompositionality
² constituency test
³ topicalization, coordination, deletion, substitution

غیر خودکار و دستی انجام گرفت. از آنجا که حجم داده‌های موجود در برونداد مرحله اول بسیار گسترده بود، تصمیم گرفته شد استخراج پیش‌فعل‌ها به روش فیلتر کردن چند مرحله‌ای انجام گیرد. بدین معنی که تعدادی از پیش‌فعل‌ها با معیارهای بالا استخراج شد و برنامه ماکروی جدیدی برای جستجوی آن پیش‌فعل‌ها به نرم‌افزار اکسل داده شد. نرم‌افزار این برنامه را بر برونداد مرحله اول که درون‌داد این مرحله بود اجرا کرد و موارد مربوط به پیش‌فعل‌های مورد جستجو را از میان موارد موجود جدا و ذیل ستون‌هایی با عنوان همان پیش‌فعل‌ها در یک کاربرگ دیگر اکسل ذخیره می‌کرد (تصویر ۶).

	A	B	C	D	E
1	توسعه	تداوم	بهبود	تحقق	رونق
2	کیه روایط سوابی توسعه بخشد. فرانسوا	ملی را حفظ و تداوم خواهد بخشد. EndP الیت	و کنترل فشارخون وضعیت وی را بهبود بخشد.	خی است که به متابع ملت ایران تحقق بخشد.	ونی بخشد. البته ای
3	تن را عملیاتی کرد و توسعه بخشد که به‌من	ای نقل فعل خود را تداوم بخشد. به نظر م	راحی بی‌بایی فرد را بهبود بخشد و حفظ کرد	تور خود تحقق عینی خواهد بخشد. اغلب مرد	رونق بخشد و برنامه
4	ه فوتبال را به شدت توسعه بخشد. برخلاف ا	خود در این ماده را تداوم بخشد. او که بر	ت به سال‌های گذشته بهبود بخشد. EndP وزرا	م و کارشناسی را نیز تحقق بخشد و در سطح	رونق بخشد. به گزارش
5	و توسعه و پیشرفت را سرعت بخشد. باور کنی	های هفته گذشته را تداوم بخشد. اقدامات	ده و باید شرایط را بهبود بخشد. افشارزاد	اتی را در صحنه نبرد تحقق بخشد و زمینه	ونی می بخشد. شرایط
6	خود توسعه و تکمیل خواهد بخشد. آسان 17**	تور ارگ و... تداوم خواهد بخشد. 7-منطق ر	کزان تولید کشور را بهبود بخشد و از سوی	این اهداف بزرگ را تحقق بخشد. این فعل	ونی می بخشد. شرایط
7	را با ایران بیشتر توسعه بخشد. استرالی	کوگک را در خانواده تداوم بخشد. به تعویز	بروی بد پاهایتان را بهبود بخشد. تست‌سوروی	ن تجاری‌سازی طم را تحقق بخشد بنابراین	رونق بخشد ولی نتگن
8	ساخت‌ها آنها را هم توسعه بخشد. وی تصوی	چودشتر و بحران را تداوم بخشد و تا امر	داتمناسب مملتان را بهبود بخشد زیرا با ب	اده شد. این معنا را تحقق بخشد. وی افزود	ا رونق بخشد. حریری ی
9	تعال داترینیان را توسعه بخشد. وی در پا	از به سندرشتینی خود تداوم بخشد. شانگردان	نیق آبیید و اوضاع را بهبود بخشد. برای زند	ما داد EndP بود را تحقق بخشد. انشا	رونق بخشد و دولت د
10	ریگان به شدت آن را توسعه بخشد و همین سب	بین مائل اسلامی را تداوم بخشد. امیررضا ح	کم کرد و تخفیه را بهبود بخشد. این مخص	گفتان دولت اصلاحات تحقق بخشد و یک ال	قی بیشتری بخشد. گشته ن
11	ه این برنامه‌ها را توسعه بخشد. وی افزود	و معتدتر از گذشته تداوم بخشد و با جلوب	ال ارتباطات، بهبود خواهد بخشد. این سازم	و استکبارستیزی را تحقق بخشد. اوتال	ونی بخشد. وی به ی
12	می‌توان فعالیت را توسعه بخشد. وی افزود	به مناطق مسکونی را تداوم بخشد تا سرد	د اوضاع اقتصادی را بهبود بخشد. همین امر	واله (ع) بتوانم تحقق بخشد. براساس	ونی بخشد. جمشید پ
13	ا هرچه بیشتر توسعه خواهد بخشد. قلمبعا	غیرحرفه‌ای خود را تداوم بخشد. این در	افیت جامه آبیده را بهبود بخشد. وی بیان	ود را در عرصه جهانی تحقق بخشد. وی با ت	ا رونق بخشد و آسانی
14	ا هرچه بیشتر توسعه خواهد بخشد. قلمبعا	ایران را به‌راحتی تداوم بخشد. وایتید	افتعاد را می‌توان بهبود بخشد. اگر محند	فی است که می‌خواهم تحقق بخشد. می‌خواه	ونی بخشد. اگر قرار
15	ن روابط را توسعه و تکمیل بخشد. او با ب	دم بید این فضا را تداوم بخشد. وی گفت:	وان طرف مدت کوتاهی بهبود بخشد و به همین	در کنند، با کمک داعش تحقق بخشد. در واقع	را رونق بخشد و موج ب
16	ه نارگان هوایی خود توسعه بخشد تا توجه	دم بید این فضا را تداوم بخشد. وی گفت:	وان وضعیت زندگی را بهبود بخشد اما نعت	ان رویاهای آنان را تحقق بخشد. گرد پری	EndP بخشد. EndP
17	ساخت‌های فرهنگی را توسعه بخشد و افراد ر	صدا، یک فهرست" را تداوم بخشد. و-ند مش	کسب و کار را بهبود خواهد بخشد. پررحمت ت	محکم جبهه مقاومت را تحقق بخشد. وی افزود	سریع بخشد. EndP
18	فعالیت خود را توسعه بخشد. در ادامه	خود را با این مرکز تداوم بخشد. در سالها	برای سال‌های آبیده بهبود بخشد. *** خروج	توانستم به این امر تحقق بخشد. الفت گف	یشتری بخشد و باعث من
19	و یک فرهنگ را توسعه بخشد و امروز ب	درصدی جلسه قبل را تداوم بخشد. به گزارش	بتوان رتبه کشور را بهبود بخشد. شرکت اپت	موانگین منطبق باشد، تحقق بخشد. هاشمی گ	ا رونق بخشد منجر به
20	ی نیز این روابط را توسعه بخشد. EndP ای	لا باید این کار را تداوم بخشد. تا آنجا	رد ملی این ماده را بهبود بخشد. ** پنج	محمدرنگ حضرت موسی تحقق بخشد. حضرت ای	ونی بخشد چراکه وز
21	تعال داترینیان را توسعه بخشد. EndP وی د	ین رویکرد را تداوم خواهد بخشد. ما قویا	دائتم، یک کیلومتر بهبود بخشد. رحمان اد	ش‌های اسلامی است را تحقق بخشد. فروزنده	رونق بخشد ن به اقتض
22	برنامه‌های خود را توسعه بخشد تا آسینب	ین رویکرد را تداوم خواهد بخشد. ما قویا	خیابانی و ظاهری را بهبود بخشد. وی ادام	دولتی (دولتی) آردی می‌باشد، تحقق بخشد. بر این ا	بفتری بخشد. به همین
23	هنر هات، بلاستیکر، را توسعه بخشد. بیکسو	حضر مکان‌های دیگر را تداوم خواهد بخشد.	7 مرحله توانستم بهبود بخشد. رتبه ای	to activate, please click on the "File" menu, click on "Account" and click on "Sign in" to activate your account.	ونی بخشد. (ن) بخشد. بور طم

تصویر ۶- نمونه‌ای از برونداد مرحله دوم جستجو
 Picture 6- A sample of the second step output

با این روش تعداد موارد باقی‌مانده که پژوهشگر به صورت دستی و غیر خودکار باید به جستجوی بقیه پیش‌فعل‌ها در آن‌ها پردازد کمتر و امر جستجوی دوباره برای پیش‌فعل‌ها کمی آسان‌تر می‌شد (تصویر ۷).

	Q	R	S	T	U	V
1811						
1812		این تصویر داشت، اما وقتی آمد و دید با گل		همینا نو روز فل به ایران آمد و قرآن بود.		
1813				همینا نو روز فل به ایران آمد و قرآن بود.		
1814						
1815						
1816						
1817						
1818		مظفر آمدنش بود، وقتی آمد اولین کاری	کور پیاده و به سمت مکتب آمد و موتورسوار		کبه و دیوانت خارجی ترکیه، آمد و این ارائه	
1819		ارستو نام صفتش پیش گل آمد و توانست!			پیشتر و بعد حاتم پیش آمد آقای ازبک	
1820					از بعد از دستها به حزام آمد و گشت از ما	آمد، وقتی از برهان هر که آمد از همه ENDP
1821		گر شد ایستاد و فرمده آمد ولی با بین				د، یکی از نمایندگان مجلس آمد به سوی من و
1822						
1823						
1824		روی مثل واقع خنده روی آب آمد و خارج شد	بود آمد یک جوان خارجی آمد و روی آن را	هم 57 به نظر سناک ارتش آمد و پس از پای		
1825		به نشانه رسیدن یک اهل آمد، اهل را ب				وقتی شد آمد، بعد بوی وحشت آمد و با ENDP
1826						
1827						
1828						
1829						
1830						
1831		زنی تکی، آهنگ، به تهران آمد و نمودند.	لاجره این آهنگ به بزار آمد تا این گروه		فروگاه امام خمینی (ره) آمد، بخش حقوقی	
1832						
1833						
1834						
1835						
1836		بعد از آن مسئول فروگاه آمد و گفت بک تی				ن استقل به تهران خواهد آمد اما منبر ای
1837						
1838		پس وقتی گردوش به ایران آمد با این وضعی				
1839		تد وقتی اسامه به ایران آمد و ایرانیان				
1840		زنی سنامه از نیروی زمینی آمد، تصاویر ب				

تصویر ۷- تصویری از موارد باقی‌مانده پس از یک مرحله جستجوی پیش‌فعل‌ها

Picture 7- A picture of the remaining cases after one stage of Preverbal elements search

این روش به‌ویژه برای فعل‌هایی که از زایایی بالایی برخوردارند بسیار کمک‌کننده است؛ برای مثال، فعل سبک «کردن» که به همین روش در نهایت بیش از ۱۶۰۰ پیش‌فعل برای آن استخراج شد. این روش بر تک‌تک فایل‌ها تا استخراج کامل پیش‌فعل‌های همنشین هریک از فعل‌ها انجام گرفت. برونداد این مرحله ۲۱ فایل اکسل حاوی ساخت‌های فعل سبک مربوط به فعل‌های مورد بررسی بود که باید در مرحله بعد برچسب‌زنی آن‌ها صورت گیرد.^۱

۲-۶. برچسب‌زنی

برچسب‌زنی فرایند اعطای برچسب به واحدهای زبانی موجود در پیکره است. برچسب‌های مورد استفاده برای هر پیکره با توجه به هدف ساخت پیکره انتخاب می‌شوند. علاوه بر آن، برچسب‌زنی پیکره‌ها براساس مدل‌های دستوری متفاوت در سطح واژه و یا نحو صورت می‌گیرد. برچسب‌زنی در سطح واژه، دادن برچسب مقوله دستوری^۲ و یا برچسب اطلاعات معنایی به واژه‌هاست که جزئیات هریک از این انواع برچسب‌ها براساس شیوه‌نامه ساخت پیکره انتخاب می‌شود. برای مثال، دو نمونه از برچسب‌های دستوری معمول برچسب‌های مقوله دستوری تراکس^۳ و برچسب‌های درخت بانک پن^۴ هستند. پیکره بی‌جن‌خان (۱۳۸۳) از مهمترین پیکره‌های برچسب‌خورده در زبان فارسی است. برچسب‌زنی در سطح نحو به روابط میان واژه‌ها و نقش آن‌ها در جمله می‌پردازد. پیکره‌های برچسب‌خورده در سطح نحو به‌طور معمول براساس دو رویکرد دستور ساخت گروهی^۵ و دستور وابستگی^۶ بوده‌اند. درخت بانک وابستگی نحوی زبان فارسی (Rasooli et al., 2013) براساس دستور وابستگی است. در زیربخش‌های بعدی به تشریح شیوه برچسب‌زنی پیکره حاضر می‌پردازیم.

^۱ البته ناگفته نماند که داده‌های استخراج شده بار دیگر پیش از ورود به مرحله برچسب‌زنی بازبینی شد به این دلیل که با وجود اعمال همه فیلترهای گفته‌شده باز هم داده‌های نامرتب در نتایج جستجو به چشم می‌خورد و باید پیش از ورود به مرحله برچسب‌زنی حذف شود که این مرحله نیز به‌صورت غیر خودکار انجام گرفت.

^۲ POS tagging

^۳ Thrax POS tags

^۴ Penn Tree bank tags

^۵ Phrase structure grammar

^۶ Dependency grammar

۱-۶-۲. انتخاب مدل برچسب‌زنی

همان‌طور که پیش از این اشاره شد پژوهش حاضر به دنبال ایجاد پیکره‌ای از ساخت‌های فعلی سبک در زبان فارسی بوده است. گرچه این ساخت‌ها در زبان فارسی حاصل هم‌نشینی پیش‌فعل‌هایی از مقوله اسم، صفت یا عبارت حرف اضافه‌ای با یک فعل سبک هستند، هدف پژوهش حاضر شناسایی و ارائه آن‌ها به‌عنوان نمونه‌ای از عناصر فعلی در زبان فارسی است. از این رو، در این مرحله تصمیم گرفته شد که برچسب‌زنی در سطح واژه صورت گیرد. برای پرهیز از پرداختن به مقوله دستوری پیش‌فعل‌ها، همسو با چارچوب دستوری صرف توزیعی پیش‌فعل‌ها ریشه‌های بی‌مقوله^۱ در نظر گرفته شد و فعل‌های سبک مقوله‌ساز^۲، برچسب (preverbal) PV فارغ از هر مقوله دستوری برای پیش‌فعل‌ها و (Light verb) LV برای فعل‌های سبک لحاظ شد. علاوه بر این مقرر گشت که بن واژه زمان حال و گذشته (lemma) مربوط به هر یک از این ساخت‌های فعلی ارائه شود. برچسب دیگر مورداستفاده در این مرحله (separable) / INSEP (inseparable) SEP است که برای نشان دادن جدایی‌پذیری و جدایی ناپذیری (با توجه به وجود یا عدم وجود فاصله) پیش‌فعل و فعل سبک به هر یک از جملات حاوی این ساخت‌های فعلی داده شد.

۲-۶-۲. انجام فرایند برچسب‌زنی

درونداد این مرحله ساخت‌های فعلی سبکی است که به‌صورت ۲۱ فایل اکسل جداگانه از مرحله جداسازی صورت‌های سبک و واژگانی برونداد شده بودند. هر یک از این فایل‌ها به یکی از فعل‌های مورد بررسی پژوهش اختصاص داشت.

برای برچسب‌زنی این فایل‌ها نیز مانند دو مرحله پیشین پژوهش، یعنی جستجوی فعل‌ها و جستجوی پیش‌فعل‌ها، برنامه برچسب‌زنی خاص این داده‌ها طراحی و بر روی هر یک از فایل‌ها اجرا شد. گرچه برچسب‌زن‌هایی در پلتفرم NLTK و نسخه فارسی آن HAZM نیز در اختیار کاربران حوزه پیکره‌پردازد؛ اما باز هم به دلیل حجم زیاد داده و فرمت داده تهیه‌شده، طراحی برنامه‌ای خاص برچسب‌زنی داده‌های پیکره مدنظر ترجیح داده شد. برچسب پیش‌فعل‌ها و فعل‌های سبک از طریق رنگی شدن این عناصر زده شد. پیش‌فعل‌ها قرمز و فعل‌های سبک سبز. از دیگر برچسب‌هایی که تصمیم گرفته شد به این ساخت‌های فعلی زده شود، برچسب (separable) INSEP (inseparable) بود. این برچسب مبتنی بر امکان جدایی‌پذیری این ساخت‌ها در زبان فارسی است؛ به عبارت دیگر، امکان فاصله افتادن میان دو عنصر پیش‌فعل و فعل سبک. تنها چالش موجود در این مرحله مربوط به زدن برچسب SEP و INSEP بود، به این دلیل که مواردی یافت می‌شد که با وجود فاصله میان دو عنصر سازنده ساخت‌های فعل سبک با ساخت فعل سبک جدایی‌ناپذیر (inseparable) مواجه بودیم. خوشبختانه این موارد قاعده‌مند هستند و همان‌طور که کریمی دوستان (2011) به آن پرداخته است شامل مواردی می‌شوند که «خواستن» زمان آینده، «داشتن» استمرار، فعل‌های کمکی و جهی «شاید» و «باید» و همچنین، ضمیرهای متصل میان این دو عنصر قرار می‌گیرند. به‌منظور جلوگیری از زدن برچسب SEP به مواردی از این دست، این موارد به‌صورت استثنا برای برنامه تعریف شد.^۳

^۱ A category roots

^۲ Verbalizer

^۳ در صورت استفاده از برچسب «جداشده/ جدا نشده» این موارد نیز در فهرست امکان جداسازی قرار می‌گرفت. از این رو، ضمن تعریف این موارد استثنا برای برنامه، برچسب جدایی‌پذیر/جدایی‌ناپذیر در این مرحله تنها برای نشان دادن نمونه‌های دارای امکان جدایی‌پذیری یا عدم این امکان انتخاب شده و در طرحی که در دست اقدام است مسئله جدایی‌پذیری/جدایی‌ناپذیری این فعل‌ها به تفصیل بررسی خواهند شد.

DM	DN	DO	DP	DQ	DR	DS	DT	DU	DV	DW	DX	DY	DZ	EA	EB	EC
														شد	47	
														تشد	6	
														تشدن	1028	
														تشدن	که این مسائل	
														تشدن	45	
														تشدن	8	
														تشدن	18	
														تشدن	اقدامات ترومان	
														تشدن	اقدامات ترو خواهم	
														تشدن	ع انجام اقدام خواهی	
														تشدن	خواهد	
														تشدن	خواهیم	
														تشدن	خواهید	7
														تشدن	خواهند	ن
														تشدن	داریم	م
														تشدن	داری	28
														تشدن	فد	ف
														تشدن	داریم	د
														تشدن	دارید	25
														تشدن	دارند	
														تشدن	داشتیم	
														تشدن	داشتی	
														تشدن	داشت	
														تشدن	داشتیم	

تصویر ۸- تصویری از استثناهای برنامه برچسب‌زنی SEP و INSEP
Picture 8- A picture of the exceptions to the SEP/INSEP annotation program

در واقع، این استثناها مواردی هستند که ساختار فعلی زبان فارسی ورود آن‌ها را در میان دو عنصر سازنده ساخت‌های فعل سبک مجاز می‌داند و سخنوران زبان برحسب نیاز از آن استفاده می‌کنند.

در کنار این موارد سه سطر بالای هر کاربرد به ترتیب بن‌واژه گذشته و حال، صورت مصدری فعل سبک و پیش‌فعل را با کمک توابع concatenate و substitute به نمایش گذاشته است. جدول (۱) راهنمای برچسب‌های پیکره مورد بحث را نشان می‌دهد.

جدول ۱- راهنمای برچسب‌های پیکره

Table 1- The corpus tags' guide

Tag's Guide		
Lemma	Past and present lemma of each LVC	White
PV	Preverb	Red
LV	Light Verb	Green
SEP	Separable	
INSEP	Inseparable	

تصویر (۹) نمونه‌ای از برونداد مرحله برچسب‌زنی را به تصویر می‌کشد. سطر اول با رنگ سفید بن‌واژه‌های گذشته و حال ساخت فعل سبک موجود در هر ستون را نشان می‌دهد، سطر دوم با رنگ سبز فعل سبک و سطر سوم با رنگ قرمز پیش‌فعل را نشان می‌دهد. همان‌گونه که تصویر نشان می‌دهد پیش‌فعل و فعل سبک در تمام نمونه‌های مربوط به هر یک از ساخت‌های فعلی موجود در یک ستون نیز به ترتیب با رنگ قرمز و سبز نشان داده شده است و برچسب SEP و INSEP نیز در مقابل هر یک از مثال‌ها زده شده است.

	A	B	C	D
1	جمع بست # جمع بند 'Lemma	عهد بست # عهد بند 'Lemma	قرارداد بست # قرارداد بند 'Lemma	به رگیار بست # به رگیار بند 'Lemma
2	بستن LV	بستن LV	بستن LV	بستن LV
3	جمع PV	عهد PV	قرارداد PV	به رگیار PV
4	من گفتم دوستان یعنی جمع بست INSEP	روز ازل با خدای خود عهد بست INSEP	گر من با بازیکنی قرارداد بست INSEP	شوهر خالعم را به رگیار بست INSEP
5	ایش ارزش افزوده در آن جمع بست INSEP	هم با خدای خود عهد خواهم بست INSEP	روزی که با ملوان قرارداد بست INSEP	تندیم و من او را به رگیار بست INSEP
6	مله و پاراگراف کوتاه جمع بست INSEP	ادار بمانیم با مردم عهد بست INSEP	پیشگامان کویر یزد قرارداد بست INSEP	حاضران را به رگیار گزوله بست SEP
7	د که بتوان با یکدیگر جمع بست INSEP	وز عروسی‌مان با محمد عهد بست INSEP	حضور در لوگ کنتی قرارداد بست INSEP	سردار هدائی را به رگیار بست INSEP
8	پنکه گله پرچمدار را جمع بست INSEP	ده بودیم با هم عهدی بستیم INSEP	انت: زمانی که قرارداد را بست SEP	اینه رنگی را به رگیار بست INSEP
9	ب را می‌شود با بیمار جمع بست INSEP	از همه مناسب‌تر است. عهد بست INSEP	تنه با آبیوشان قرارداد بست INSEP	بچی به بست داشت به رگیار بست INSEP
10	ن طبیعت دو نفری کننده جمع بست INSEP	وقتی رقم خانه خد، عهد بست INSEP	را التماسی آمد و قرارداد بست INSEP	شهر صفی‌آباد را به رگیار بست INSEP
11		فتم: من با خدای خود عهد بست INSEP	یز طلق قرارداد: که با ما بست SEP	شهر صفی‌آباد را به رگیار بست INSEP
12		با حضرت ابراهیم (ع) عهدی بست INSEP	ای راهن قرارداد داخلی بست SEP	آباد دزفول را به رگیار بست INSEP
13		سید. روزی که دلسوزان عهد بست INSEP	ن با فلان باشگاه قرارداد بست INSEP	هاتم این شهر را به رگیار بست INSEP
14		سید. روزی که دلسوزان عهد بست INSEP	داد زیادی قرارداد خواهیم بست INSEP	شهر صفی‌آباد را به رگیار بست INSEP
15		فادار بمانیم با مردم عهد بست INSEP	. او دیوان با من قرارداد بست INSEP	هاتم این شهر را به رگیار بست INSEP
16		مام زمان (عج) عهد سربازی بستند SEP	مناسب از بازنشستگان، قرارداد خواهیم بست INSEP	ن و هو مردم را به رگیار بست INSEP
17		که بسویجان یا خاروند عهد بستند INSEP	کرانسی اوگتیار قرارداد بست INSEP	0 ساله را به رگیار گزوله بستند SEP
18		از همان زمان یا خود عهد بستند INSEP	کنور و انصارالله قرارداد بست INSEP	ساحلی سوسه تونس به رگیار بست INSEP
19		اجتماعی با خدای خود عهد بستند INSEP	1 ساله فریتالیست قرارداد بست INSEP	یش را در سراوان به رگیار بست INSEP
20		نورانی زینب گبری (بن) عهد بستند INSEP	این قرارداد این مجموعه را بست SEP	یش را در سراوان به رگیار بست INSEP
21		بیم و ما عهدی که با ولایت بستند SEP	بود، با استقلال قرارداد بست INSEP	سلمان تاثیر را به رگیار بست INSEP
22		و با عهدی که با شهیدان بستند SEP	او کفای بیمان و قرارداد بست INSEP	شده و یکدیگر را به رگیار بستند INSEP
23		د این شهیدان عهد و بیمان بستند SEP	استقلال قرارداد دوساله بست SEP	وهای انتظامی را به رگیار بستند INSEP

تصویر ۹- تصویری از فایل برچسب‌خورده ساخت‌های فعل سبک
 Picture 9- A picture of light verb construction annotated files

گام بعدی اجرای پروژه ساماندهی داده‌ها در قالب پیکره است که در زیربخش بعد به آن می‌پردازیم.

۳. داده‌های موجود در پیکره

داده‌های حاصل از این پژوهش، ساخت‌های فعل سبک مربوط به ۲۱ فعل سبک زبان فارسی هستند که در بافت‌های زبانی به صورت برچسب‌خورده ارائه شده‌اند. این ۲۱ فعل عبارت‌اند از: آمدن، آوردن، افتادن، انداختن، بخشیدن، بردن، بستن، خوردن، دادن، داشتن، دیدن، رسیدن، رفتن، زدن، شدن، کردن، کشیدن، گذاشتن، گرفتن، گشتن و یافتن. در ادامه، اطلاعات مربوط به فراوانی داده‌های موجود در پیکره برای هریک از این فعل‌ها ارائه می‌شود.

اولین فعل از این مجموعه فعل سبک «آمدن» است. برای فعل سبک «آمدن» در داده‌های پیکره ۷۷ ساخت فعل سبک یافت شد که در حدود ۳۲۰۰۰ بافت زبانی ارائه شده‌اند. از جمله ساخت‌های فعل سبکی که فعل سبک در آن‌ها «آمدن» است می‌توان به «به عمل آمدن»، «فائق آمدن»، «به سر آمدن» و نظایر آن اشاره کرد.

دومین فعل از فعل‌های مورد بررسی، فعل سبک «آوردن» است. برای فعل سبک «آوردن» در داده‌های پیکره ۱۰۸ ساخت فعل سبک یافت شد که در حدود ۶۰۰۰۰ بافت زبانی ارائه شده‌اند. از جمله ساخت‌های فعل سبکی که فعل سبک در آن‌ها «آوردن» است می‌توان به «به ستوه آوردن»، «به چنگ آوردن»، «به خشم آوردن» و مانند آن اشاره کرد.

سومین فعل مورد بررسی فعل سبک «افتادن» است. برای فعل سبک «افتادن» در داده‌های پیکره ۹۸ ساخت فعل سبک یافت شد که در حدود ۱۵۰۰۰ بافت زبانی ارائه شده‌اند. از جمله ساخت‌های فعل سبکی که فعل سبک در آن‌ها «افتادن» است می‌توان به «دور افتادن»، «به - شک افتادن»، «کارگر افتادن» و مانند آن اشاره کرد.

چهارمین فعل از این مجموعه فعل سبک «انداختن» است. برای فعل سبک «انداختن» در داده‌های پیکره ۱۱۳ ساخت فعل سبک یافت شد که در حدود ۷۰۰۰۰ بافت زبانی ارائه شده‌اند. از جمله ساخت‌های فعل سبکی که فعل سبک در آن‌ها «انداختن» است می‌توان به «از پا انداختن»، «خط انداختن»، «از کار انداختن» و نظایر اشاره کرد.

پنجمین فعل مورد بررسی، فعل سبک «بخشیدن» است. برای فعل سبک «بخشیدن» در داده‌های پیکره ۱۱۱ ساخت فعل سبک یافت شد که در حدود ۵۰۰۰۰ بافت زبانی ارائه شده‌اند. از جمله ساخت‌های فعل سبکی که فعل سبک در آن‌ها «بخشیدن» است می‌توان به «زینت - بخشیدن»، «عزت بخشیدن»، «تسلی بخشیدن» و نظایر آن اشاره کرد.

ششمین فعل از این مجموعه فعل سبک «بردن» است. برای فعل سبک «بردن» در داده‌های پیکره ۶۹ ساخت فعل سبک یافت شد که در حدود ۱۹۰۰۰ بافت زبانی ارائه شده‌اند. از جمله ساخت‌های فعل سبکی که فعل سبک در آن‌ها «بردن» است می‌توان به «لذت بردن»، «رنج - بردن»، «یورش بردن» و نظایر آن اشاره کرد.

فعل سبک «بستن» هفتمین فعل سبک از این مجموعه است. برای فعل سبک «بستن» در داده‌های پیکره ۳۳ ساخت فعل سبک یافت شد که در حدود ۳۰۰۰ بافت زبانی ارائه شده‌اند. از جمله ساخت‌های فعل سبکی که فعل سبک در آن‌ها «بستن» است می‌توان به «همت بستن»، «عقد بستن»، «دل بستن» و نظایر آن اشاره کرد.

هشتمین فعل از این مجموعه، فعل سبک «خوردن» است. برای فعل سبک «خوردن» در داده‌های پیکره ۱۲۱ ساخت فعل سبک یافت شد که در حدود ۲۰۰۰۰ بافت زبانی ارائه شده‌اند. از جمله ساخت‌های فعل سبکی که فعل سبک در آن‌ها «خوردن» است می‌توان به «قل - خوردن»، «زخم خوردن»، «قسم خوردن» و مانند آن اشاره کرد.

نهمین فعل مورد بررسی، فعل سبک «دادن» است. برای فعل سبک «دادن» در داده‌های پیکره ۳۵۹ ساخت فعل سبک یافت شد که در حدود ۵۰۰۰۰ بافت زبانی ارائه شده‌اند. از جمله ساخت‌های فعل سبکی که فعل سبک در آن‌ها «دادن» است می‌توان به «چاک دادن»، «نشر دادن»، «جان دادن» و مانند آن اشاره کرد.

فعل دهم از مجموعه فعل‌های مورد بررسی، فعل سبک «داشتن» است. برای فعل سبک «داشتن» در داده‌های پیکره ۵۱۶ ساخت فعل سبک یافت شد که در حدود ۳۰۰۰۰ بافت زبانی ارائه شده‌اند. از جمله ساخت‌های فعل سبکی که فعل سبک در آن‌ها «داشتن» است می‌توان به «انس داشتن»، «رونق داشتن»، «تمنا داشتن» و مانند آن اشاره کرد.

یازدهمین فعل از این مجموعه فعل سبک «دیدن» است. برای فعل سبک «دیدن» در داده‌های پیکره ۴۳ ساخت فعل سبک یافت شد که در حدود ۴۰۰۰۰ بافت زبانی ارائه شده‌اند. از جمله ساخت‌های فعل سبکی که فعل سبک در آن‌ها «دیدن» است می‌توان به «شایسته دیدن»، «فراهم دیدن»، «سان دیدن» و مانند آن اشاره کرد.

دوازدهمین فعل مورد بررسی فعل سبک «رسیدن» است. برای فعل سبک «رسیدن» در داده‌های پیکره ۱۷۶ ساخت فعل سبک یافت شد که در حدود ۶۵۰۰۰ بافت زبانی ارائه شده‌اند. از جمله ساخت‌های فعل سبکی که فعل سبک در آن‌ها «رسیدن» است می‌توان به «به وفاق رسیدن»، «به قطعیت رسیدن»، «به ارث رسیدن» و نظایر آن اشاره کرد.

فعل سیزدهم از این مجموعه فعل سبک «رفتن» است. برای فعل سبک «رفتن» در داده‌های پیکره ۷۲ ساخت فعل سبک یافت شد که در حدود ۴۰۰۰۰ بافت زبانی ارائه شده‌اند. از جمله ساخت‌های فعل سبکی که فعل سبک در آن‌ها «رفتن» است می‌توان به «لو رفتن»، «سجده رفتن»، «از یاد رفتن» و مانند آن اشاره کرد.

چهاردهمین فعل مورد بررسی فعل سبک «زدن» است. برای فعل سبک «زدن» در داده‌های پیکره ۲۳۷ ساخت فعل سبک یافت شد که در حدود ۱۹۰۰۰ بافت زبانی ارائه شده‌اند. از جمله ساخت‌های فعل سبکی که فعل سبک در آن‌ها «زدن» است می‌توان به «ناخنک زدن»، «لطمه زدن»، «گریز زدن» و مانند آن اشاره کرد.

پانزدهمین فعل از مجموعه فعل‌های مورد بررسی، فعل سبک «شدن» است. برای فعل سبک «شدن» در داده‌های پیکره حدود ۱۱۳۷ ساخت فعل سبک یافت شد که در حدود ۱۰۰۰۰۰۰ بافت زبانی ارائه شده‌اند. از جمله ساخت‌های فعل سبکی که فعل سبک در آن‌ها «شدن» است می‌توان به «عاجز شدن»، «مجازات شدن»، «اثبات شدن» و نظایر آن اشاره کرد.

فعل شانزدهم از فعل‌های مورد بررسی فعل سبک «کردن» است. برای فعل سبک «کردن» در داده‌های پیکره حدود ۱۶۶۹ ساخت فعل سبک یافت شد که در حدود ۱۲۰۰۰۰۰ بافت زبانی ارائه شده‌اند. از جمله ساخت‌های فعل سبکی که فعل سبک در آن‌ها «کردن» است می‌توان به «کشت کردن»، «ترقی کردن»، «چک کردن» و مانند آن اشاره کرد.

هفدهمین فعل از فعل‌های مورد بررسی، فعل سبک «کشیدن» است. برای فعل سبک «کشیدن» در داده‌های پیکره ۱۷۱ ساخت فعل سبک یافت شد که در حدود ۱۱۰۰۰ بافت زبانی ارائه شده‌اند. از جمله ساخت‌های فعل سبکی که فعل سبک در آن‌ها «کشیدن» است می‌توان به «زوزه کشیدن»، «به نظم کشیدن»، «زجر کشیدن» و مانند آن اشاره کرد.

هجدهمین فعل مورد بررسی، فعل سبک «گذاشتن» است. برای فعل سبک «گذاشتن» در داده‌های پیکره حدود ۲۱۱ ساخت فعل سبک یافت شد که در حدود ۴۵۰۰۰ بافت زبانی ارائه شده‌اند. از جمله ساخت‌های فعل سبکی که فعل سبک در آن‌ها «گذاشتن» است می‌توان به «قیمت گذاشتن»، «تأکید گذاشتن»، «تمایز گذاشتن» و نظایر آن اشاره کرد.

فعل نوزدهم از این مجموعه فعل سبک «گرفتن» است. برای فعل سبک «گرفتن» در داده‌های پیکره حدود ۲۷۷ ساخت فعل سبک یافت شد که در حدود ۳۸۶۰۰۰ بافت زبانی ارائه شده‌اند. از جمله ساخت‌های فعل سبکی که فعل سبک در آن‌ها «گرفتن» است می‌توان به «بغل گرفتن»، «حاجت گرفتن»، «گوشه گرفتن» و مانند آن اشاره کرد.

بیستمین فعل از فعل‌های مورد بررسی، فعل سبک «گشتن» است. برای فعل سبک «گشتن» در داده‌های پیکره ۳۱۰ ساخت فعل سبک یافت شد که در حدود ۴۹۰۰۰ بافت زبانی ارائه شده‌اند. از جمله ساخت‌های فعل سبکی که فعل سبک در آن‌ها «گشتن» است می‌توان به «ترکیب گشتن»، «سرکوب گشتن»، «عازم گشتن» و نظایر آن اشاره کرد.

در نهایت، بیست و یکمین فعل از فعل‌های مورد بررسی این پژوهش، فعل سبک «یافتن» است. برای فعل سبک «یافتن» در داده‌های پیکره ۲۲۶ ساخت فعل سبک یافت شد که در حدود ۶۸۰۰۰ بافت زبانی ارائه شده‌اند. از جمله ساخت‌های فعل سبکی که فعل سبک در آن‌ها «یافتن» است می‌توان به «تکوین یافتن»، «هدایت یافتن»، «مهلت یافتن» و مانند آن اشاره کرد.

جدول ۲- فراوانی ساخت‌های فعل سبک و بافت‌های مرتبط در پیکره حاضر

Table 2- The frequency of light verb constructions and their related context in the present corpus

فعل سبک	فراوانی ساخت‌های فعل سبک	فراوانی بافت‌های زبانی حاوی ساخت‌های فعل سبک
۱ آمدن	۷۷	۳۲۰۰۰
۲ آوردن	۱۰۸	۶۰۰۰۰
۳ افتادن	۹۸	۱۵۰۰۰
۴ انداختن	۱۱۳	۷۰۰۰۰
۵ بخشیدن	۱۱۱	۵۰۰۰۰
۶ بردن	۶۹	۱۹۰۰۰
۷ بستن	۳۳	۳۰۰۰۰
۸ خوردن	۱۲۱	۲۰۰۰۰۰
۹ دادن	۳۵۹	۵۰۰۰۰۰
۱۰ داشتن	۵۱۶	۳۰۰۰۰۰
۱۱ دیدن	۴۳	۴۰۰۰۰
۱۲ رسیدن	۱۷۶	۶۰۰۰۰
۱۳ رفتن	۷۲	۴۰۰۰۰
۱۴ زدن	۲۳۱	۱۱۹۰۰۰
۱۵ شدن	۱۱۳۷	۱۰۰۰۰۰۰
۱۶ کردن	۱۶۶۹	۱۲۰۰۰۰۰
۱۷ کشیدن	۱۷۱	۱۱۰۰۰۰
۱۸ گذاشتن	۲۱۱	۴۵۰۰۰
۱۹ گرفتن	۲۷۷	۳۸۶۰۰۰
۲۰ گشتن	۳۱۰	۴۹۰۰۰
۲۱ یافتن	۲۲۶	۶۸۰۰۰

همان طور که مشاهده می‌شود فراوانی داده‌های به‌دست آمده برای هر یک از این فعل‌ها متفاوت است. اسحاقی و کریمی‌دوستان (۱۴۰۰) به تفصیل به این موضوع پرداخته‌اند.

۴. خلاصه و نتیجه‌گیری

در این مقاله پیکره ساخت‌های فعلی سبک زبان فارسی را معرفی کردیم. ابتدا به ماهیت این ساخت‌های فعلی و اهمیت و ضرورت مطالعه آن‌ها در زبان فارسی پرداختیم. پس از اشاره به تعدادی از آثاری که به ضرورت وجود داده‌های مربوط به این ساخت‌ها تأکید کرده‌اند و چالش‌های مرتبط با در دسترس نبودن این قبیل داده‌ها را متذکر شده‌اند، وارد مبحث مراحل ایجاد پیکره شدیم. در این مراحل یکی از اولین گام‌ها انتخاب پیکره پژوهشگاه ارتباطات و فناوری اطلاعات به‌عنوان پیکره مبنای استخراج داده‌ها بود. گفته شد که استخراج ساخت‌های فعل سبک از پیکره به‌آسانی و در یک مرحله امکان‌پذیر نیست. مراحل استخراج داده‌ها یک‌به‌یک توضیح داده شد. اولین برنامه ماکرو برای استخراج فعل به‌تنهایی و بدون در نظر گرفتن صورت سبک یا واژگانی فعل بود. مرحله بعد تشخیص ساخت‌های فعل سبک با استفاده از معیار سبک‌شدگی معنایی و سازواری بود. پس از آن برنامه ماکرو جدیدی برای جستجوی پیش‌فعل‌های مربوط به

هریک از فعل‌ها به نرم‌افزار اکسل داده شد که نتیجه آن ۲۱ فایل اکسل بود که در هر یک از آن‌ها جملات استخراج‌شده از پیکره مینا حاوی فعل‌های مورد بررسی به تفکیک پیش‌فعل‌های همنشین وجود داشت. سپس، از میان مدل‌های برجسب‌زنی موجود مدلی متناسب با داده‌های به‌دست‌آمده انتخاب شد و فرایند برجسب‌زنی ساخت‌های فعلی استخراج‌شده انجام گرفت. بدین ترتیب یک پیکره همزمانی تک‌زبان با قابلیت جستجوی رایانه‌ای به دست آمد که امید است در رفع چالش‌های موجود برای این ساخت‌ها راهگشا باشد.

آنچه در مطالعات بعدی مدنظر است بررسی آماری مسئله جدایی‌پذیری / جدایی‌ناپذیری ساخت‌های فعل سبک، تهیه فرهنگ یک‌زبان از ساخت‌های فعل سبک در زبان فارسی، استخراج الگوهای رفتار نحوی و معنایی فعل‌ها، دوزبان کردن فرهنگ ساخت‌های فعل سبک زبان فارسی است.

تشریح و قدردانی

پژوهش حاضر حاصل طرح پژوهشی پسادکترای زبان‌شناسی است که در دانشگاه تهران و با حمایت مالی صندوق حمایت از پژوهشگران و فناوری‌ان کشور انجام گرفته است. بدین وسیله از صندوق حمایت از پژوهشگران و فناوری‌ان کشور نهایت قدردانی و سپاس را ابراز می‌داریم.

منابع

- اسحاقی، مهدیه و کریمی دوستان، غلامحسین. (۱۴۰۰). زیایی فعل‌های سبک در زبان فارسی. *پژوهش‌های زبانی دانشگاه تهران* (۱۲(۲)، ۱-۲۸.
- برفی، وفا. (۱۳۹۵). *کشف فضای معنایی افعال سبک زبان فارسی در نوشتار فارسی آموزان خارجی از دیدگاه شناختی*. پایان‌نامه کارشناسی ارشد، دانشگاه سمنان.
- بی‌جن‌خان، محمود. (۱۳۸۳). نقش پیکره‌های زبانی در نوشتن دستور زبان: معرفی یک نرم‌افزار رایانه‌ای. *زبان‌شناسی* (۱۹(۲)، ۶۷-۴۸.
- بی‌جن‌خان، محمود. (۱۳۹۷). *پیکره طرح جویشگر بومی*. تهران: مرکز تحقیقات مخابرات ایران.
- تسلیمی‌پور، شیوا. (۱۳۹۱). *پردازش خودکار معنایی افعال مرکب زبان فارسی*. پایان‌نامه کارشناسی ارشد، دانشگاه شیراز.
- حساس صدیقی، پریا. (۱۳۸۹). *آموزش زبان فارسی به غیرفارسی‌زبانان: مشکلات و راهکارها*. پایان‌نامه کارشناسی ارشد، دانشگاه علامه طباطبایی.
- خزاعی‌فر، علی. (۱۳۸۴). نظریه ترجمه دیروز و امروز. *نامه فرهنگستان* (۱۷(۴)، ۷۹-۲۸.
- دبیرمقدم، محمد. (۱۳۷۶). فعل مرکب در زبان فارسی. *زبان‌شناسی* (۲۳، ۲-۴۶.
- طیب‌زاده، امید. (۱۳۸۴). *رابطه فرهنگ‌نگاری و دستورنویسی*. *نامه فرهنگستان* (۱۷(۴)، ۳۱-۲۷.
- عسکریان، نرجس. (۱۳۹۱). *شناسایی خودکار افعال مرکب زبان فارسی*. پایان‌نامه کارشناسی ارشد، دانشگاه شیراز.
- معرفت، فهیمه. (۱۳۸۴). بررسی خطاهای زبانی در نوشتار دانش‌آموزان فارسی‌آموز کردزبان در سطح متوسط زبان‌آموزی. *متن‌پژوهی* (۲۶(۹)، ۱۰-۳۷.
- ناتل‌خانلری، پرویز. (۱۳۶۵). *تاریخ زبان فارسی* (جلد ۲). تهران: نشر نو.

Acquaviva, P. (2008). Roots and lexicality in distributed morphology. In A. Galani, D. Redinger and N. Yeo (Eds), *Special issues of York working papers in linguistics* (pp.1-21) New York: University of New York.

AleAhmad, A., Amiri, H., Rahgozar, M., and Oroumchian, F. (2009). Hamshahri: A Standard Persian Text Collection. *Knowledge-Based Systems* 22(5), 382-387.

Aronoff, M. (1976). *Word formation in generative grammar*. Cambridge, MA: MIT Press.

- Askariyan, N. (2012). *Automatic identification of Persian compound verbs*. Master thesis, University of Shiraz. [In Persian]
- Assi, S. M. (1997). Farsi linguistic database (FLDB). *International Journal of Lexicography* 10(3), 265.
- Barfi, V. (2016). Discovering the semantic space of Persian light verbs in the writing of Persian-foreign students from a cognitive point of view. Master thesis, University of Semnan. [In Persian]
- Beard, R. (1995). *Lexeme-morpheme base morphology*. New York: State University of New York Press.
- BijanKhan, M. (2004). The role of linguistic corpora in writing the grammar of language: An introduction to a computer software. *Journal of Linguistics* 19(2), 48-67. [In Persian]
- Bijan Khan, M. (2018). *Native search engine project*. Tehran: Research Institute of Information and Communication Technology. [In Persian]
- Bonet, E. (1991). *Morphology after syntax: Pronominal clitics in Romance languages*. PhD dissertation, MIT.
- Borer, H. (2003). Exo-skeletal vs. endo-skeletal explanations: Syntactic projections and the lexicon. In J. C. Moore and M. Polinsky (Eds), *The nature of explanations in linguistic theory* (pp. 37-67). Chicago: Chicago University Press.
- Borer, H. (2013). The category of roots. In R. Alexiadou, H. Borer and F. Schafer (Eds.), *The syntax of roots and the roots of syntax* (pp.112-149). Oxford: Oxford University Press.
- DabirMoghaddam, M. (1997). Persian compound verb. *Journal of Linguistics* 23, 31-46. [In Persian]
- Embick, D., and Marantz, A. (2008). Architecture and blocking. *Linguistic Inquiry* 39, 1-53.
- Embick, D., and Noyer, R. (2001). Movement operations after syntax. *Linguistic Inquiry* 32, 555-595.
- Eshaghi, M., and Karimi-Doostan, G. (2021). The productivity of Persian light verbs. *Journal of Language Researches* 12, 1-28. [In Persian]
- Family, N. (2006). Explorations of semantic space: The case of light verb constructions in Persian. PhD dissertation, Ecole des Hautes Etude en Sciences Sociales.
- Folli, R., Harley, H., and Karimi, S. (2005). Determinant of event type in Persian complex predicates. *Lingua* 115(10), 1365-1401.
- Goldberg, A. E. (1996). Words by default: Optimizing constraints and the Persian complex predicate. In D. Librik and R. Belear (Eds.), *Proceedings of Berkeley Linguistic Society* (pp. 132-146). Berkeley: Berkley University Press.
- Halle, M., and Marantz, A. (1993). Distributed morphology and the pieces of inflection. In K. Hale and S. J. Keyser (Eds.), *The view from building* (pp. 117-176). Cambridge: MIT Press.
- Halle, M. (1990). An approach to morphology. *North Eastern Linguistic Society* 20(1), 150-184.
- Harley, H. (2009). Compounding in Distributed morphology. In R. Lieber and P. Stekauer (Eds.), *Oxford Handbook of Compounding* (pp. 129-144). Oxford: Oxford University Press.
- HasasSediqi, P. (2010). *Teaching Persian to non-Persian speakers: problems and solutions*. Master thesis, Alame Tabatabaei University. [In Persian]
- Jespersen, O. (1965). *A modern English grammar on historical principles*. London: George Allen and Unwin Ltd.
- Karimi, S. (1997). Persian complex verbs: Idiomatic or compositional. *Lexicology* 3, 273-318.
- Karimi-Doostan, G. (1997). Light verb constructions in Persian. PhD dissertation, University of Essex.
- Karimi- Doostan, G. (2005). Light verb and structural case. *Lingua* 115(12), 1737-1756.
- Karimi-Doostan, G. (2008). Event structure of verbal nouns and light verbs. In S. Karimi, V. Samiian and D. Stilo (Eds), *Aspects of Iranian linguistics* (pp.206-226). NewCastle: Cambridge Scholars Publishing.
- Karimi- Doostan, G. (2011). Separability of light verb constructions in Persian. *Studia Linguistica* 65(1), 70-95.
- Khazaeifar, A. (2005). Transation theory: today and in the past. *Academy of Persian Language and literature journal* 28, 69-79. [In Persian]
- Kiparsky, P. (1982). Lexical morphology and phonology. In S. Yang (Ed.), *Linguistics in the Morning Calm* (pp. 3-91). Seoul: Hansin.
- Kiparsky, P. (1997). Remarks on denominal verbs. In A. Alsina, J. Bresnan and P. Sells (Eds.), *Argument Structure* (pp. 473-499). Stanford: Center for the Study of Language and Information.

- Marantz, A. (1997). No escape from syntax: Don't try morphological analysis in the privacy of your own lexicon. In A. Dimitriadis, L. Siegel, C. Surek-Clark, & A. Williams (Eds.), *University of Pennsylvania working papers in linguistics* (pp. 201-225). Philadelphia: University of Philadelphia.
- Marantz, A. (2000). *Roots: The universality of roots and pattern morphology*. Presented at the *Conference on Afro-Asiatic Language*. Paris University.
- Marantz, A. (2013). Verbal argument structure: Events and participants. *Lingua* 130, 152–168.
- Marefat, F. (2005). Written errors of Kurdish learners of Persian: A case of Mahabadi dialect. *Literary Text Research* 9(26), 10-37. [In Persian]
- Megerdooimian, K. (2001). Event structure and complex predicates in Persian. *Canadian Journal of Linguistics* 46, 97-125
- Natel Khanlari, P. (1986). *The history of Persian language* (Vol. 2). Tehran: Nashrenow. [In Persian]
- Noyer, R. (1997). *Features, positions and affixes in autonomous: Morphological structure*. New York, NY: Garland.
- Panagiotidis, P. (2015). *Categorial features: A general theory of word class categories*. Cambridge: Cambridge University Press.
- Pestesty, D. (1982). Complementizer-trace phenomena and the nominative island condition. *The linguistic review* 1(3), 297-344
- Pestesty, D. (1995). *Zero syntax: Experiencers and cascades*. Cambridge MA: The MIT Press.
- Rasooli, M. S., Kouhestani, M., and Moloodi, A. S. (2013). Development of a Persian syntactic dependency treebank. In H. Hua, J. Lin, & A. Lopez (Eds), *Proceedings of the 2013 Conference of the North American Chapter of Association for Computational Linguistics: Human Language Technologies* (pp. 306-314). Atlanta: Association for Computational Linguistics.
- Samvelian, P., and Faghiri, P. (2013). Persian complex predicates: How compositional are they? *Semantics-Syntax Interface* 1, 43-74.
- Shamsfard, M., Hesabi, A., Fadaei, H., Mansoory, N., Noor, P., Famian, A., Bagherbeigi, S., Fekri, E., and Monshizadeh, M. (2010). Semiautomatic development of Farsnet; the Persian wordnet. *Proceedings of 5th Global WordNet Conference* 9(2), 35-44.
- Siddiqi, D. (2009). *Syntax within word: Economy, allomorphy, and argument selection in Distributed Morphology*. Amsterdam: John Benjamins.
- Tabibzade, O. (2005). Dictionary and grammar writing. *Academy of Persian Language and literature journal* 28, 69-79. [In Persian]
- Taslimipoor, S. (2012). *Automatic semantic processing of Persian compound verbs*. Master thesis, University of Shiraz. [In Persian]
- Vahedi Langrudi, M. (1996). *The syntax, semantics and argument structure of complex predicates in modern Farsi*. PhD dissertation, University of Ottawa.

