



## Applications of Corpus Linguistics in Teaching Persian to Non-Persian Speakers

Reza Morad Sahraei<sup>1</sup>, Samira Mirzaei<sup>2</sup>

1. Professor of Linguistics, Department of Linguistics and Teaching Persian to Non-Persian Speakers, Faculty of Persian Literature and Foreign Languages, Allameh Tabataba'i University, Tehran, Iran. E-mail: sahraei@atu.ac.ir
2. Corresponding Author, M.A. of Teaching Persian to Non-Persian Speakers, Department of Linguistics, Faculty of Persian Literature and Foreign Languages, Allameh Tabataba'i University, Tehran, Iran. E-mail: s.mirzaei@gmail.com

---

### Article Info

**Article type:**  
Research Article

**Article history:**

**Received:** 29 May 2023

**Received in revised form:**

01 Aug 2023

**Accepted:** 12 Aug 2023

**Available online:** 22 Dec 2023

**Keywords:**

corpus linguistics,  
pedagogic corpus,  
language learner corpus  
annotation,  
metadata.

### ABSTRACT

Today, corpus-based language teaching is the focus of many researchers in the field of language education, and language corpora are used in different ways in this field. In this article, we intend to introduce various types of corpora useful in language teaching and point out their structural features, and discuss the efficiency of corpora and their role in language teaching. Based on this, we have generally introduced two types of corpus, pedagogic corpus, and the other language learner corpus, and we have mentioned examples of these corpora in the Persian language. In addition, we have listed some tagged linguistic data of the Persian language and showed how they can be used in teaching the Persian language. It is clear that the main question of this research is focused on the effect and applications of corpora in language teaching, and to answer it, we had to refer to the background of theoretical and practical studies in this field with a comprehensive descriptive view. The results of the review and research studies of this study show that each of the mentioned bodies can be useful in language teaching in a special way according to the content infrastructure they have, and the data analysis of this article confirms the positive effect of linguistic data on the production of appropriate educational content.

---

**Cite this article:** Sahraei, R. M., & Mirzaei, S. (2024). Applications of Corpus Linguistics in Teaching Persian to Non-Persian Speakers *Research in Western Iranian Languages and Dialects*, 11 (4), 113-140.



© The Author(s).

Publisher: Razi University.

DOI: 10.22126/JLW.2023.9189.1703

---

## 1- Introduction

This research emphasizes the importance of corpus linguistics in the discussion of teaching the Persian language. Therefore, the concept of corpus, corpus linguistics, its impact on language teaching, and the types of useful corpora are examined. The use of corpora, extracting information from them, corpus tools, and types of Persian learner corpora are the main focus of this research. Linguistic corpora include learner corpora and pedagogical corpora, both of which are effective in this field. The use of these corpora for the preparation and editing of teaching materials based on real data at different levels has provided useful knowledge to teachers and researchers in the field of education. In this research, the introduction and examination of the use of available corpora, syntactically tagged corpora, and non-syntactic tagged corpora in the Persian language, as well as the influential studies of corpora, were discussed. However, there have been relatively few corpus-based studies conducted in the Persian language. The use of the AntConc software for examining content words and prepositions has been practical for the Persian language from the Hamshahri corpus, and subsequently, searching the syntactically tagged corpus yielded acceptable results.

## 2- Methodology and literature review

Research indicates that the use of language corpora not only opens up new outlets of knowledge but also enhances the speed, accuracy, and credibility of the work. Learner and educational language corpora are two types of corpora that serve the field of language teaching. Learner corpora and pedagogical corpora are two types of corpora that serve the field of language education. The learner corpus is derived from the language productions of language learners and is used for error analysis and, consequently, educational planning. Pedagogical corpora are a suitable educational resource for writers, teachers, and language learners (Gavioli & Aston, 2001: 244), providing learners with the opportunity to learn language in the face of natural language productions. In this study, to examine the impact of corpora on language education, we first review the background of corpus-based language teaching studies. Then, we will examine the keywords in this field. We will also introduce data related to Persian language learning and corpus-based studies of Persian language education. Additionally, we will attempt to explain how to use three types of Persian learner corpora, the Persian pedagogical corpus (along with the characteristics of such a corpus) and annotated corpora in teaching the Persian language. In the process, while referring to practical examples, we will discuss the limitations and boundaries of different types of corpora to provide a criterion for corpus-based language teaching research. Some of the corpus-based studies on the Persian language include Safari (1395), Bakneov et al. (1398), Rezapour and Mouchani (1400). Similarly, in the field of corpus-based Persian language learning studies with a focus on vocabulary, we can mention Bijankhan et al. (1393), Sahraei et al. (1396), Zulfaqr Kandari et al. (1398), Mirzaei (1398), and Navidi et al. (1400).

## 3- Findings

As mentioned, language corpora can be useful in three forms: learner corpora, pedagogical corpora, and annotated corpora in language education. To create a learner

corpus, after gathering the data and storing learners' language productions in the database, annotating is carried out. The types of annotations can be categorized into two groups, determined based on the purposes of creating the corpus. The first group of tags is error tags, which indicate the areas where the learner's language production contains errors and the type of error. The other tags are linguistic.

In the third phase, tools for preparing the corpus are developed. Zero (1991), Mushtaq et al. (2019), Johansson (2008), and Sánchez (2013) have presented different categorizations of errors.

So far, 4 corpora have been compiled for Persian, including the Saadi Foundation Corpus, Allameh Tabataba'i University Corpus, Salam Farsi Corpus, and Persian Learner Corpus.

Pedagogical corpora are collected from the language productions of native speakers, making them a valuable resource for language learners. These corpora provide data that can be used by teachers, learners, and writers during language learning (Gavioli & Aston, 2001) and are themselves considered as teaching resources. Subsequently, concepts such as collocation and language teaching, vocabulary in context and language teaching, word sequences, and basic and specialized vocabulary based on frequency lists of words and keywords were discussed.

Annotated corpora such as the Persian Dependency Treebank, Persian Proposition Bank, Persian Discourse Bank, Named Entity Corpus, and Coreference Corpus were introduced.

The interaction of corpus linguistics with language teaching was also examined, focusing on research conducted in the field of teaching Persian language and the applications of the Persian learner corpus in teaching Persian, the applications of pedagogical corpora in teaching Persian, and the applications of annotated Persian corpora in teaching Persian.

#### **4- Discussion and conclusion**

The results indicated that the use of annotated corpora can be effective in teaching grammar and syntactic and semantic structures. Additionally, the Persian Dependency Treebank provides suitable patterns for teaching prepositions and compound verbs. The use of pedagogical corpora in developing lesson content for vocabulary and teaching vocabulary in context is useful. Both of these corpora can serve as educational aids for language learners and assist them in learning the language. It was also stated that learner corpora and their annotated errors are resources used in producing educational materials and can even indicate the learning trajectory of language topics. It can be said that various types of language corpora, including learner corpora, pedagogical corpora, and annotated corpora, are all useful and illuminating in language teaching in various ways. The development of suitable educational corpora, considering the goals of language learners, can be seen as the prospect of future research. Furthermore, making the most of the annotated corpora of the Persian language in presenting educational patterns can be part of future studies. Tagging and analyzing the errors of learner corpora in Persian and then utilizing them according to the learner's level, first language, language learning goals, and the like, are other subjects suggested for future studies.

The research section of this article focused on extracting multi-layered content and grammatical data from different Persian language sources and their application in language teaching. It also explored the use of Persian syntactic corpora in Persian language teaching, with a focus on the role of prepositions in verb structure and the categorization of verbs based on their grammatical capacity.



## کاربردهای زبان‌شناسی پیکره‌ای در آموزش زبان فارسی به غیرفارسی‌زبانان

رضامراد صحرائی<sup>۱</sup>، سمیرا میرزائی<sup>۲</sup>

۱. استاد زبان‌شناسی و آموزش زبان فارسی به غیرفارسی‌زبانان، گروه زبان‌شناسی، دانشکده ادبیات فارسی و زبان‌های خارجی، دانشگاه علامه طباطبائی، تهران، ایران. رایانامه: [sahraei@atu.ac.ir](mailto:sahraei@atu.ac.ir)
۲. نویسنده مسئول، کارشناسی ارشد آموزش زبان فارسی به غیرفارسی‌زبانان، گروه زبان‌شناسی، دانشکده ادبیات فارسی و زبان‌های خارجی، دانشگاه علامه طباطبائی، تهران، ایران. رایانامه: [s.mrzaei@gmail.com](mailto:s.mrzaei@gmail.com)

### چکیده

### اطلاعات مقاله

امروزه آموزش زبان مبتنی بر پیکره، مورد توجه بسیاری از پژوهشگران حوزه آموزش زبان است و استفاده از پیکره‌های زبانی در این حوزه به شیوه‌های مختلفی صورت می‌گیرد. در این مقاله بر آن بودیم تا ضمن معرفی انواع پیکره‌های مفید در امر آموزش زبان و اشاره به ویژگی‌های ساختاری آن‌ها، به کارآمدی پیکره‌ها و نقش آن‌ها در آموزش زبان پردازیم. براساس این، به‌طور کلی، دو نوع پیکره، یکی پیکره آموزشی زبان و دیگری پیکره زبان‌آموز را معرفی کرده، به نمونه‌هایی از این پیکره‌ها در زبان فارسی اشاره کرده‌ایم. علاوه بر آن، برخی داده‌های زبانی برچسب‌خورده زبان فارسی را فهرست کردیم و نشان دادیم که چگونه می‌توان از آن‌ها در آموزش زبان فارسی بهره برد. مشخص است که پرسش اصلی این پژوهش بر نحوه اثرگذاری و کاربردهای پیکره‌ها در آموزش زبان متمرکز است و برای پاسخ به آن ناگزیر بوده‌ایم با نگاه جامع توصیفی به پیشینه مطالعات نظری و عملی این حوزه مراجعه کنیم. نتیجه بررسی‌های مروری و پژوهشی این مطالعه نشان داد که هریک از پیکره‌های مورد اشاره باتوجه به زیرساخت محتوایی که دارند به شکل خاصی می‌توانند در آموزش زبان مفید باشند و بررسی داده‌بنیاد این مقاله بر تأثیر مثبت داده‌های زبانی بر تولید محتوای آموزشی مناسب صحه می‌گذارد.

نوع مقاله: پژوهشی

تاریخ دریافت: ۸ خرداد ۱۴۰۲

تاریخ بازنگری: ۱۰ مرداد ۱۴۰۲

تاریخ پذیرش: ۲۱ مرداد ۱۴۰۲

دسترسی برخط: ۱ دی ۱۴۰۲

### واژه‌های کلیدی:

زبان‌شناسی پیکره‌ای،  
پیکره آموزشی زبان،  
پیکره زبان آموز،  
برچسب‌گذاری،  
اطلاعات فراداده.

استناد: صحرائی، رضامراد؛ میرزائی، سمیرا (۱۴۰۲). کاربردهای زبان‌شناسی پیکره‌ای در آموزش زبان فارسی به غیرفارسی‌زبانان. *مطالعات زبان‌ها و گویش‌های غرب ایران*، ۱۱ (۴)، ۱۱۳-۱۴۰.



© نویسنده‌گان.

ناشر: دانشگاه رازی

DOI: 10.22126/JLW.2023.9189.1703

## ۱- مقدمه

مطالعات نشان می‌دهد که استفاده از پیکره‌های زبانی علاوه‌بر آنکه دریچه‌های جدیدی از دانش را پیش‌رو قرار می‌دهد، به سرعت، دقت و اعتبار کار می‌افزاید. امروزه در آموزش زبان استفاده از پیکره‌ها بسیار مورد اقبال است. پیکره‌های زبان‌آموز و آموزشی زبان دو نوع از پیکره‌هایی هستند که به حوزه آموزش زبان خدمات می‌دهند. پیکره زبان‌آموز برگرفته از تولیدات زبانی زبان‌آموزان است و برای تحلیل خطا و در نتیجه، برنامه‌ریزی‌های آموزشی استفاده می‌شود. پیکره آموزشی زبان که برای نویسندگان، مدرسان و زبان‌آموزان منبع آموزشی مناسبی به شمار می‌رود (گاوپولی و استون<sup>۱</sup>، ۲۰۰۱: ۲۴۴)، از داده‌های طبیعی گویشوران زبان گردآوری می‌شوند. این داده‌ها این امکان را برای زبان‌آموز فراهم می‌آورند که در مواجهه با تولیدات طبیعی، زبان را بیاموزد. در این پژوهش برای بررسی چگونگی و میزان تأثیر پیکره‌ها بر آموزش زبان، ابتدا پیشینه مطالعات آموزش زبان پیکره‌یار را مرور می‌کنیم؛ سپس کلیدواژه‌های این حوزه را بررسی خواهیم کرد. همچنین داده‌های مرتبط با فارسی‌آموزی و بررسی‌های داده‌بنیاد آموزش زبان فارسی را معرفی می‌کنیم. علاوه‌بر آن، در این بررسی می‌کوشیم نحوه استفاده از سه نوع پیکره فارسی‌آموز، پیکره آموزشی زبان فارسی (همراه با بیان ویژگی‌های چنین پیکره‌ای) و پیکره‌های برجسب‌خورده زبان فارسی را در آموزش زبان فارسی شرح دهیم. در این‌بین، ضمن اشاره به نمونه‌های کاربردی، به بیان محدودیت‌ها و حد و حدود انواع پیکره‌ها می‌پردازیم تا متر و معیاری برای پژوهش‌های آموزش زبان پیکره‌یار ارائه کنیم.

مرور پژوهش‌های مرتبط با آموزش پیکره‌یار زبان در دو بخش پژوهش‌های فارسی‌آموزی و پژوهش‌های زبان‌آموزی قابل بررسی است. از پژوهش‌های پیکره‌یار زبان فارسی می‌توان به صفری (۱۳۹۵) اشاره کرد که ضمن تعریف پیکره زبان‌آموز و مرور روش‌های تولید آن با معرفی پیکره سلام فارسی، تولید پیکره زبان‌آموز را به شکل عملی معرفی می‌کند. در پژوهشی دیگر بکنوف<sup>۲</sup> و دیگران (۱۳۹۸) خطاهای نوشتاری فارسی‌آموزان قزاق (۲۴ دختر و ۱۹ پسر) را در پیکره‌ای ۴۰۰۰ واژه‌ای بررسی و انواع خطاها را شناسایی کردند. یافته‌ها نشان داد قزاق‌ها بیشتر خطاهای دستوری دارند. در همین زمینه، رضاپور و موچانی (۱۴۰۰) تفاوت کاربرد افزوده‌های وجهی در ۲۴۸ دست‌نوشته فارسی‌آموزان سطح پیشرفته زن و مرد را در چارچوب دستور نقش‌گرای نظام‌مند بررسی کردند. نتایج

نشان داد در این داده‌ها افزوده‌های وجه‌نمایی و در وجه‌نمایی‌ها، قطعی‌ها فراوان‌ترند. همچنین مشخص شد زن‌ها بیشتر از افزوده‌های اظهاری در نوشته‌هایشان استفاده می‌کنند.

از پژوهش‌های فارسی‌آموزی مبتنی بر پیکره آموزشی زبان و با موضوع واژه می‌توان به بی‌جن‌خان و دیگران (۱۳۹۳) که به نقش واژگان بسامدی در مهارت واژگانی پرداختند و جهانگردی و دیگران (۱۳۹۵) که واژه‌های پُرسامد پایگاه داده‌های زبان فارسی را با کتاب‌های فارسی‌آموزی مقایسه کردند و از عدم همپوشانی این فهرست با محتوای واژگانی کتاب‌ها پرده برداشتند، اشاره کرد. همچنین صحرایی و دیگران (۱۳۹۶) واژه‌های پُرسامد پیکره‌ای ۱۲۰۰ هزار واژه‌ای برگرفته از روزنامه‌های کثیرالانتشار را با فهرست واژه‌های پُرسامد پژوهش‌های دیگر مقایسه کردند و از این مقایسه به اختلاف ۳۰ درصدی رسیدند. پژوهش دیگر، صحرایی و دیگران (۱۳۹۷) است که از پیکره‌ای شامل ۱۲۰۳۵۹۸ واژه برگرفته است، از متون مطبوعاتی به دست آمد. فهرست واژه‌های حاصل از این بررسی ۲۱۵۰ واژه با بسامد بیش از ۵۰ بار تکرار در پیکره بود.

ذوالفقارکندری و دیگران (۱۳۹۸) نیز به استخراج خودکار واژگان پایه علوم پزشکی پرداختند و برای این منظور از دو پیکره تخصصی (محقق‌ساخته) و پیکره همشهری (نسخه دوم) به‌عنوان پیکره عمومی استفاده کردند. نویدی و دیگران (۱۴۰۰) واژه‌های آموزشی کتاب‌های گام اول، مینا، آموزش کاربردی واژه و آموزش فارسی در سطح مقدماتی را با واژه‌های پُرسامد پیکره در فرهنگ بسامدی، همشهری، فرهنگ بسامدی فارسی و فرهنگ زبان‌آموز پیشرفته فارسی مقایسه کردند. نتیجه پژوهش انطباق ۵ تا ۲۶ درصد را نشان داد که مقدار کمی است.

موضوع دیگر پژوهش میرزائی (۱۳۹۸) است که به بررسی چگونگی تأثیر پیکره‌ها بر آموزش واژه‌های محتوایی و نقشی پرداخته و با استفاده از پیکره همشهری نشان داد که هر پیکره‌ای با هر محتوایی نمی‌تواند به‌عنوان پیکره آموزشی در آن زبان مورد استفاده قرار گیرد.

از پژوهش‌های زبان‌آموز پیکره‌یار می‌توان به بولتون<sup>۱</sup> (۲۰۰۹) اشاره کرد که کوشید میزان اثربخشی پیکره را بسنجد. او ۱۳۲ دانشجوی سال اول کالج مهندسی در فرانسه را در سطوح تقریباً متوسط و پایین انگلیسی در دو گروه قرار داد و یکی را به روش سنتی (با گزیده‌هایی از فرهنگ لغت دوزبانه یا دستورالعمل دستور) و گروه دیگر را به‌صورت پیکره‌یار آموزش داد. یافته‌ها نشان داد آموزش پیکره‌ای

مؤثرتر از روش سنتی است. وایاتکینا<sup>۱</sup> (۲۰۱۶) نیز که به میزان اثربخشی آموزش داده‌بنیاد باهم‌آیی‌های واژگانی-دستوری (و اینجا باهم‌آیی فعل-حرف اضافه در زبان آلمانی) در مقایسه با روش‌های سنتی اشاره کرده، نشان داد هر دو روش برای آموزش مؤثر است؛ ولی آموزش داده‌بنیاد باهم‌آیی‌ها توفیق بیشتری دارد.

پژوهش دیگر، پاکر و اوزکان<sup>۳</sup> (۲۰۱۷) است که ۲۸ دانشجوی سطح متوسط با زبان مادری ترکی و میانگین سن بین ۱۷ تا ۲۰ سال را انتخاب کردند. گروه آزمایش، واژگان را با داده‌بنیاد آموختند و گروه کنترل، آموزش واژه را به صورت سنتی و با کتاب و فرهنگ لغت یاد گرفتند. نتایج نشان داد گروه آزمایش نمره بالاتری نسبت به گروه کنترل کسب کرده‌اند. در ضمن، یافته‌ها نشان از رضایت دانشجویان از روش پیکره‌یار داشت. بونتام و فوچارونسیل<sup>۴</sup> (۲۰۱۸) نیز آموزش داده‌بنیاد حروف اضافه «در طول، میان و بین» را در میان دانش‌آموزان تایلندی بررسی کردند. نتیجه آموزش ۳۰ دانش‌آموز تایلندی به مدت شش هفته، با روش داده‌بنیاد و بر اساس نمرات پیش‌آزمون و پس‌آزمون نشان از پیشرفت دانش‌آموزان در نمره پس‌آزمون داشت. همچنین شرکت‌کنندگان بعد از آموزش می‌توانستند جملات معنایی و دستوری بیشتری را با پیچیدگی‌های گوناگون تولید کنند.

لین<sup>۵</sup> (۲۰۲۱) در آموزش داده‌بنیاد، گرامر را به دو روش داده‌بنیاد و سنتی بررسی کرد. نتایج دو گروه موردپژوهش نشان داد هیچ‌یک تغییر قابل‌توجهی در نگرش یادگیری نداشتند. این یعنی آموزش پیکره‌بنیاد فقط برای تقویت زبان‌شناختی زبان‌آموزان گرامر مناسب بوده است. همچنین مشخص شد بیشترین پیشرفت آموزش پیکره‌بنیاد برای دانشجویان سطح پایین است. لین استدلال می‌کند احتمالاً دانشجویان پیشرفته بیشتر تجربه یادگیری گذشته را مورد استفاده قرار داده باشند.

## ۲- مبانی نظری

### ۲-۱ انواع پیکره‌های کارآمد در مطالعات آموزش زبان

پیکره‌های زبانی می‌توانند به سه شکل پیکره‌های زبان‌آموز، پیکره‌های آموزشی زبان و پیکره‌های برچسب‌خورده در آموزش زبان مفید باشند.

1. N. Vyatkina

2. collocation

3. T. Paker & Y. E. Ozcan

4. P. Boontam & S. Phoocharoensil

5. M. H. Lin



## ۱-۱-۲ پیکره زبان آموز

پیکره زبان آموز از تولیدات زبان آموزان تولید می‌شود. در پیکره‌های زبان آموز، پس از گردآوری داده‌ها، اطلاعات مرتبط با زبان آموزان (اعم از زبان مادری، سطح و اهداف زبان‌آموزی، سن، جنس و مانند آن) یا تولیدات زبانی (مانند زمان تولید، نوع تکلیف، نوشتاری یا گفتاری بودن آن و...) مشخص می‌شود و داده بر اساس آن امکان طبقه‌بندی می‌یابد. با این طبقه‌بندی در مجموعه داده‌ها می‌توان با امکان فیلتر درخواست کرد که مثلاً فقط داده‌های زن‌ها نشان داده شود یا سطح زبانی مشخصی درخواست می‌شود و یا داده‌های زبانی مرتبط با زبان آموزان یک سطح مشخص و با زبان مادری یکسان جست‌وجو می‌شود.

## ۲-۱-۲ مشخصات پیکره زبان آموز

برای ساخت پیکره‌های زبان آموز، در گام اول زیرساخت پیکره آماده می‌شود؛ یعنی تولیدات زبانی زبان آموزان با روش‌شناسی مشخص در پایگاه داده ذخیره می‌شود. این مرحله اگرچه در تعریف، ساده به نظر می‌رسد، پیچیدگی‌های خودش را دارد. برای مثال، در مواجهه با دست‌نوشته‌های زبان آموزان لازم است انتقال داده به رایانه صورت بگیرد. این انتقال با سختی‌های زیادی همراه است؛ چون معمولاً دست‌خط‌ها ناخوانا هستند و ثبت دقیق انشا با خطاهای آن (جهت بررسی و امکان تحلیل خطا) دشوار است. برای مثال، در بعضی از انشاها مصحح روی کلمه‌ها خط می‌کشد و کلمه دیگری می‌نویسد؛ به طوری که کلمه قبل، قابل خواندن و ثبت نیست.

گاهی بخشی به کلمه اضافه می‌شود و قابل تشخیص نیست که مصحح اضافه کرده یا از قبل بوده است. بعضی از اشتباهات را هم نمی‌توان ثبت کرد؛ چون در صفحه کلید علامتی برای آن وجود ندارد (اگرچه با کدنویسی و صرف زمان بیشتر می‌توان آن‌ها را نشان داد)؛ برای مثال، زبان آموز دندان‌های «س» را کم یا زیاد می‌گذارد یا نقطه‌های حرف «چ» را بالای آن قرار می‌دهد. گاهی دست‌خط زبان آموز خوانا نیست و گاه خط‌خطی‌ها سبب می‌شود که متن قابل خواندن نباشد. اگر داده گفتاری باشد دشواری کار به جهت پیاده‌سازی داده دوچندان می‌شود.

ثبت اطلاعات فراداده برای متن (شامل عنوان، مکان و زمان تولید متن، منابع استفاده‌شده در نگارش متن مانند فرهنگ لغت یا کتاب دستور و...) و برای زبان آموز (سن، جنسیت، ملیت، هدف زبان‌آموزی، زبان‌هایی که آموخته، مدت زبان‌آموزی، میزان تحصیلات و...)، فعالیت بعدی در ساخت پیکره است. پس از تهیه زیرساخت داده، نشانه‌گذاری و برچسب‌زنی صورت می‌گیرد. نوع برچسب‌ها که در دو

دسته قابل‌بررسی هستند بر اساس اهداف ساخت پیکره مشخص می‌شود. گروه اول برچسب‌ها از نوع برچسب‌های خطا است که مشخص می‌کند تولید زبانی زبان‌آموز در چه بخش‌هایی حاوی خطا است و آن خطا از چه نوعی است. برچسب‌های دیگر زبان‌شناختی هستند (مانند برچسب‌های صرفی و نحوی). در فاز سوم ابزارهای تهیه پیکره (شامل ابزارهای نشانه‌گذاری و ابزارهای جست‌وجو در پیکره) آماده می‌شوند. با آماده‌شدن زیرساخت داده‌ای، فهرست برچسب‌ها و ابزارهای برچسب‌زنی پیکره برچسب می‌خورد. پس از آن زمان خروجی گرفتن از داده و انجام پژوهش‌های بعدی است.

### ۳-۱-۲ انواع برچسب‌های خطا در پیکره زبان‌آموز

برچسب‌های خطا شامل خطای نحو، صرف، املا، علائم نگارشی، خطاهای واژگانی و مانند آن‌هاست و برای هر برچسب، ریزبرچسب‌هایی مانند حذف، اضافه، جایگزینی و ترتیب نابجا را می‌توان نام برد (صفری، ۱۳۹۱). بدیهی است که فهرست‌های متفاوتی از انواع برچسب‌های خطا وجود دارد که در کلیت به هم شبیه‌اند. در ادامه، می‌توان برخی از این فهرست‌ها را به شکل خلاصه مرور کرد.

مشتاق<sup>۱</sup> و دیگران (۲۰۱۹) در تحلیل پیکره‌بنیاد خطا در نوشتار سطح میانی، از انواع خطاها شامل خطای درج، حذف، انتخاب اشتباه کلمات، املا، علائم نگارشی، حرف تعریف و مانند آن یاد می‌کنند و بیان می‌دارند که در بررسی‌های آن‌ها اشتباهات املایی بیشترین خطا و بعد از آن خطاهای فعل قرار دارد؛ همچنین طبق بررسی مشتاق و دیگران به نظر می‌رسد خطاهای مربوط به فعل، حرف تعریف، نقطه‌گذاری و حروف اضافه از مشکلات عمده زبان‌آموزان است. بنابراین، معلمان باید برای این مباحث برنامه‌ریزی کنند.

خطاهای زبانی در چهار طبقه قرار می‌گیرد:

- لغزش‌هایی که می‌تواند توسط خود یادگیرنده اصلاح شود؛
- اشتباهاتی که در صورت تذکر می‌تواند توسط یادگیرنده اصلاح شود؛
- خطاهایی که بدون یادگیری قابل اصلاح نیست؛
- لغزش دستوری که شامل نقض قوانین تجویزی می‌شود (جیمز<sup>۲</sup>، ۱۹۹۸: ۸۱).

از نگاه جیمز، دسته اول و آخر بیشتر به گویشور بومی مربوط است و دو خطای میانی برای زبان‌آموز است. در دو مورد آخر خطاها یا به دلیل تأثیر زبان اول بر زبان دوم است یا تأثیر زبان دوم بر

یادگیری است.

یوهانسن<sup>۱</sup> (۲۰۰۸) انواع خطاهای واژگانی و خطاهای دستوری را مورد بحث قرار می‌دهد. در خطاهای واژگانی انواع خطاهای معادل‌گزینی، مفهومی (زبان‌آموز کلمه‌ای در زبان مقصد را با کلمه‌ای مرتبط اشتباه می‌گیرد)، واژه‌سازی، خطاهای مربوط به واژه‌های دستوری (حروف اضافه)، خطاهای مربوط به باهم‌آیی کلمه‌ها، خطای مربوط به ظاهر کلمه و مربوط به حرف‌نویسی آن مورد نظر است. سانچز<sup>۲</sup> (۲۰۱۳) خطاها را در شش دسته شامل خطاهای واژگانی، دستوری، کاربردشناختی، عبارت‌پردازی، قواعد نگارشی و خطاهایی که غیرقابل‌کدگذاری هستند، دسته‌بندی می‌کند. در خطاهای واژگانی خطای مربوط به املا، مربوط به جابه‌جایی‌های واژگانی و دیگر خطاهای انتخاب واژه قرار می‌گیرد. خطای گروه‌های صفتی، قیدی، حرف‌اضافه‌ای، فعلی، بند، بند مرکب، خطای ساختارهای خاص و مانند آن در طبقه خطاهای دستوری قرار می‌گیرد. خطاهای مربوط به انسجام، پیوستگی و ژانر نیز در موضوع خطاهای کاربردشناختی مطرح است.

#### ۴-۱-۲ فهرست پیکره‌های زبان‌آموز غیرفارسی

پیکره‌های زبان‌آموز غیرفارسی بسیار متعدد و متنوع است.<sup>۳</sup> دلیل این تنوع، حجم پیکره، زبان پیکره، زبان‌آموز یا زبان اول زبان‌آموزان، گفتاری یا نوشتاری بودن داده و مانند آن است. برخی از این موارد در جدول (۱) معرفی شده است.

جدول (۱). نمونه‌هایی از پیکره زبان‌آموز

نام پیکره	زبان اول زبان‌آموزان	نوع داده پیکره
پیکره زبان‌آموز عربی ALC	۶۶ زبان	گفتاری و نوشتاری
پیکره زبان‌آموز چینی JCLC	۵۰ زبان	نوشتاری
پیکره انگلیسی Crow	۲۴ زبان	نوشتاری
پیکره انگلیسی ENNTT	۲۴ زبان	نوشتاری
پیکره زبان‌آموز انگلیسی ALEC	عربی	نوشتاری
پیکره زبان‌آموز انگلیسی YELK	کره‌ای	گفتاری

#### ۴-۱-۵ فهرست پیکره‌های زبان‌آموز فارسی

برای زبان فارسی تاکنون ۴ پیکره شامل پیکره بنیاد سعدی، علامه طباطبائی، سلام فارسی و زبان‌آموز

1. S. Johansson

2. A. M. Pérez Sánchez

3. <https://uclouvain.be/en/research-institutes/ilc/cecl/learner-corpora-around-the-world.html>

فارسی تدوین شده است:

پیکره بنیاد سعدی شامل ۱۴۷۰۰ برگه‌متن، از زبان آموزان خارجی سعدی، مرکز آموزش زبان فارسی به غیرفارسی‌زبانان دانشگاه علامه طباطبائی، جامعه‌المصطفی و کالج بین‌الملل دانشگاه علوم پزشکی تهران گردآوری شده است. این پیکره ۳۰۰ فایل صوتی هم دارد (میرزائی، ۱۳۹۹).

پیکره فارسی‌آموز علامه طباطبائی شامل بیش از هزار دست‌نوشته فارسی‌آموزان مرکز آرفای دانشگاه علامه طباطبائی است که پس از ثبت در رایانه و دریافت اطلاعات فراداده در فاز برجسب‌گذاری تحلیل خطا وارد شده است.

پیکره سلام فارسی (صفری، ۱۳۹۵: ۱۱۴) از متون نوشتاری فارسی‌آموزان صربستانی گردآوری شده است. سطح این متون پایه، میانی و فوق‌میانی است. برای این پیکره تمام مراحل تهیه یعنی گردآوری داده، یکسان‌سازی، نشانه‌گذاری و برجسب‌گذاری و تهیه ابزارهای جست‌وجو در پیکره فراهم شده است. این پیکره به‌منظور شناسایی خطاهای فارسی‌آموزان صربستانی تولید شده است.

پیکره زبان‌آموز فارسی (صفری، ۱۳۹۵: ۱۱۴) نیز از ۱۵۰ متن نوشتاری فارسی‌آموزان مدرسه المهدی در دو سطح میانی و پیشرفته تهیه شده است. این مجموعه داده برای پایان‌نامه کارشناسی ارشد و برپایه الگوی پیشنهادی طراحی پیکره تونو<sup>۱</sup> (۲۰۰۳) گردآوری شده است.

## ۲-۲ پیکره آموزشی زبان

پیکره آموزشی زبان از تولیدات گویشوران زبان گردآوری می‌شود؛ از این رو، می‌تواند منبع مناسبی برای برخورد زبان‌آموز با زبانی باشد که در حال یادگیری آن است. این پیکره‌ها داده‌هایی هستند که هنگام زبان‌آموزی در اختیار معلم، زبان‌آموزان و نویسندگان قرار می‌گیرند (گاویولی و آستون، ۲۰۰۱) و خود به‌عنوان منبع آموزشی مطرح می‌شوند. جانس<sup>۲</sup> (۱۹۹۱) به بهره‌گیری از پیکره‌های زبانی در آموزش زبان، یادگیری مبتنی بر داده می‌گوید.

### ۲-۲-۱ باهم‌آیی و آموزش زبان

به بیشترین هم‌وقوعی واژه‌ها در هر زبان هم‌آیی<sup>۳</sup> یا باهم‌آیی گفته می‌شود. هریک از واژه‌ها در رابطه باهم‌آیی گره و هم‌آیند نامیده می‌شوند. گره بخش اصلی و پایه گروه و هم‌آیند عضوی است که تعداد

1. Tono  
2. T. Johns  
3. collocation

هم‌وقوعی آن با گره سنجیده می‌شود (میرزائی، ۱۳۹۹). برای مثال، در عبارت آموزش زبان، «آموزش» گره است و «زبان» همایند محسوب می‌شود. برای تعیین باهم‌آیی سه ملاک را باید در نظر گرفت: الف) پراکندگی<sup>۱</sup> که به معنی فراوان بودن در کل پیکره است؛ ب) خاص بودن<sup>۲</sup> به این معنی که چقدر این همایندی در زبان تکرار شده است؛ پ) یک‌سویگی<sup>۳</sup> که می‌گوید کدام یک از کلمه‌های همایند بیشتر کلمه دیگر را یادآوری می‌کند (گابلاسوا<sup>۴</sup> و دیگران، ۲۰۱۷).

اهمیت همایندی در آموزش زبان در تهیه طیف‌های واژه‌ای و لیست‌واژه‌های مرتبط است که در آموزش واژه‌های محتوایی بسیار مهم است. دسترسی به فهرست باهم‌آهای دستوری هم بسیار اساسی است.

### ۲-۲-۲ واژه در بافت و آموزش زبان

به جمله‌های مستخرج از پیکره زبانی که در آن کاربرد و بافت هم‌نشین یک واژه مشخص بررسی می‌شود فهرست بافت‌نما<sup>۵</sup> گفته می‌شود. در پیکره‌های زبانی و با استفاده از نرم‌افزارهای بافت‌نما می‌توان برای دستیابی به رفتار بافتی واژه‌ها چند کلمه از بافت هم‌نشین سمت چپ و راست واژه را درخواست داد. دسترسی به چنین فهرست‌هایی از امکاناتی است که به سبب آن می‌توان چندمعنایی یا هم‌نام‌آی‌های مختلف را از هم متمایز کرد. به بیان دیگر، زبان‌آموز در آموزش‌های خود معنی کلمه «خوردن» را با موضوع غذا می‌آموزد. اما این زبان‌آموز اگر امکان جست‌وجو در پیکره‌های آموزشی زبان را داشته باشد و بتواند کلمه را در بافت ببیند، به‌زودی یاد می‌گیرد که «خوردن» در معنای دیگری هم به کار می‌رود. مثلاً خوردن فلز توسط اسید، خوردن به معنی اصابت کردن، خوردن به معنی مناسب بودن، خوردن به معنای نوشیدن و بسیاری معنای دیگر که با بررسی پیکره می‌توان به آن دست یافت.

### ۲-۲-۳ توالی‌های واژه‌ای

در هر زبان برخی رشته‌های واژه‌ای وجود دارند که در زبان پرتکرار هستند. مثلاً «تیم ملی فوتبال ایران» یک خوشه واژگانی است که در متون ورزشی زبان فارسی بسیار پرشمار است. دسترسی به رشته‌های پربسامد در آموزش زبان بسیار مفید است. محاسبه و شمارش خوشه‌های پرتکرار یا به اصطلاح

1. dispersion
2. exclusivity of collocates
3. directionality
4. D. Gablasova
5. concordance
6. homonym

چندپشته‌ها<sup>۱</sup> (با تعداد اجزا و عضوهای متفاوت) الگوهای پرتکرار واژگانی یا دستوری زبان را به دست می‌دهد. در زبان فارسی میرزائی (۱۳۹۸) به بررسی چندپشته‌های دستوری و میرزائی (۱۴۰۰) به بررسی چندپشته‌های واژگانی زبان فارسی پرداخته و نتایج آن را در موضوع آموزش زبان به بحث گذاشته است.

#### ۴-۲-۲ واژه‌های پایه و تخصصی بر اساس فهرست بسامدی واژه‌ها<sup>۲</sup> و کلیدواژه‌ها<sup>۳</sup>

با استفاده از نرم‌افزارهای واژه‌نما می‌توان فهرست بسامدی واژه‌ها و کلیدواژه‌ها را تهیه کرد. این فهرست‌ها در آموزش زبان بسیار مفید و کارآمد هستند و پژوهش‌های زیادی در حوزه آموزش زبان بر همین اساس صورت گرفته است. دستیابی به واژگان پایه در زبان یکی از این سودمندی‌ها است. در این زمینه، می‌توان به فهرست فراوانی واژه‌ها و واژه‌های پایه در زبان انگلیسی به نقل از ثورندایک<sup>۴</sup> (۱۹۴۴) اشاره کرد. علاوه بر آن، مون<sup>۵</sup> (۲۰۱۰) در پژوهشی که روی پیکره بانک انگلیسی<sup>۶</sup> انجام داد، آماري از بسامد واژه‌ها و بن‌واژه‌های<sup>۷</sup> زبان انگلیسی به دست داده که در زیر قابل مشاهده است (۱۹۸):

جدول (۲). نمونه‌هایی از پیکره زبان آموز

نام پیکره	زبان اول زبان آموزان	نوع داده پیکره
پیکره زبان آموز عربی ALC	۶۶ زبان	گفتاری و نوشتاری
پیکره زبان آموز چینی JCLC	۵۰ زبان	نوشتاری
پیکره انگلیسی Crow	۲۴ زبان	نوشتاری
پیکره انگلیسی ENNTT	۲۴ زبان	نوشتاری
پیکره زبان آموز انگلیسی ALEC	عربی	نوشتاری
پیکره زبان آموز انگلیسی YELK	کره‌ای	گفتاری

طبق جدول بالا ۱۰ واژه اول BOE واژه‌های نقشی<sup>۸</sup> شامل *the, of, a, an, in, to* مصدری، *have to* (حرف اضافه‌ای) و *it* هستند که در هر پیکره‌ای پریسامد است. در مقابل این‌ها واژه‌های محتوایی<sup>۹</sup> قرار دارند که دستیابی به آن‌ها به تعیین واژه‌های پایه زبان که از ملزومات آموزش زبان است

1. N-grams
2. word list
3. keyword list
4. Thorndike and Lorge
5. M. R. Moon
6. Bank of English= BOE
7. token
8. function word
9. content word

کمک می‌کند. بر اساس گزارش مون، هزار واژه پرسامد *BOF*، حدود ۶۸ درصد واژه‌های پیکره و ۱۰ هزار واژه نخست، حدود ۹۰ درصد را شامل می‌شود. مسلماً با وجود چنین فهرستی می‌توان انواع منابع مفید آموزش واژه را برای زبان انگلیسی طراحی کرد که نمونه‌هایی از آن کتاب ۱۱۰۰ واژه، کتاب ۵۰۴ است که در بازار کتاب ایران هم موجود هستند.

علاوه بر فهرست واژه‌های پرکاربرد عمومی زبان، دستیابی به واژه‌های تخصصی هم در آموزش زبان اهمیت دارد. ممکن است زبان‌آموز با اهداف ویژه یا دانشگاهی به سراغ آموزش زبان بیاید و در حوزه‌ای خاص زبان را بیاموزد؛ در اینجا لزوم تدوین کتاب‌های زبان تخصصی ضروری می‌نماید و دستیابی به فهرست واژه‌های تخصصی اهمیت پیدا می‌کند. نرم‌افزارهای واژه‌پرداز به آسانی این فهرست‌ها را از پیکره استخراج می‌کنند. برای پیدا کردن واژه‌های کلیدی، پیکره مورد بررسی پژوهش با پیکره دیگری به‌عنوان پیکره مرجع مقایسه می‌شود. اگر در بسامدگیری واژه‌ای، کلمه‌ای هم در پیکره مرجع و هم در پیکره مورد نظر پرسامد باشد کلیدواژه محسوب نمی‌شود؛ اما اگر در پیکره پژوهش خیلی پرسامد و در پیکره مرجع کم‌بسامد باشد، آنگاه کلیدواژه به شمار می‌رود.

پس با در اختیار داشتن پیکره مناسب می‌توان در امر آموزش واژه، تهیه واژه‌نامه‌های زبان‌آموز و تهیه متون و محتوای آموزشی واژه‌محور از امکان بسامدگیری واژه‌ای و کلیدواژه‌ای استفاده کرد.

### ۳-۲ پیکره‌های برچسب‌خورده زبانی و آموزش زبان

در پیکره‌های برچسب‌خورده اطلاعات سطوح مختلف زبان برای زنجیره زبان تعیین می‌گردد. پیکره‌های برچسب‌خورده معمولاً برای آموزش ماشین و به‌منظور فعالیت‌های پردازش طبیعی زبان تولید می‌شوند؛ اما بدیهی است که زبان‌آموز، مدرسان و تهیه‌کنندگان محتوای آموزشی هم می‌توانند از این اطلاعات استفاده کنند. در زبان فارسی پیکره‌های برچسب‌خورده متعددی تهیه شده‌اند؛ پیکره وابستگی نحوی زبان فارسی، پیکره گزاره‌های معنایی زبان فارسی، پیکره گفتمانی زبان فارسی، پیکره موجودیت نامدار و پیکره مرجع‌دارها از این موارد هستند.

در ادامه، پیکره وابستگی نحوی زبان فارسی که حاوی اطلاعات صرفی و نحوی است و برای زبان‌آموزان، مدرسان و مؤلفان محتوای آموزشی بسیار مفید است، مختصراً معرفی می‌شود؛ اما در خصوص پیکره‌های دیگر که بیشتر اطلاعات تخصصی مانند نوع نقش‌های معنایی (کنشگر، کنش‌پذیر و...) در آن مشخص شده یا پیکره گفتمان و مرجع‌دار که ارتباطات تخصصی مرتبط با سطح گفتمان در آن تعیین شده است و از آن بیشتر در حد پژوهش و استفاده در محتوای آموزشی می‌توان

استفاده کرد، به اختصار توضیح خواهیم داد.

### ۱-۳-۲ پیکره وابستگی نحوی زبان فارسی

پیکره وابستگی نحوی زبان فارسی (رسولی و دیگران، ۲۰۱۳) شامل ۳۰ هزار جمله و حدود نیم‌میلیون کلمه است. تعداد برچسب‌های صرفی این پیکره ۱۷ مورد است و بیشتر این برچسب‌ها زیربخش‌هایی دارد. در این پیکره ویژگی‌های صرفی هر واژه و روابط نحوی میان واژه‌ها مشخص و به صورت دستی برچسب‌گذاری شده و در واسط کاربری تحت وب به آدرس [search.dadegan.ir](http://search.dadegan.ir) قابل دسترسی است. از آنجاکه در آموزش دستور زبان به ویژگی‌های نحوی و روابط درونی جمله اشاره می‌شود و زبان‌آموزان با اصطلاحاتی مانند فاعل، قید، پرسش‌واژه و مانند آن آشنا می‌شوند، به نظر می‌رسد که پیکره‌های نحوی هم برای خود زبان‌آموز منبع مفیدی است و هم برای مدرسان و مؤلفان و تولیدکنندگان محتواهای آموزشی سودمند است. در کنار پیکره وابستگی زبان فارسی، فرهنگ ظرفیت نحوی (رسولی و دیگران، ۲۰۱۱) هم موجود است که حدود ۵ هزار فعل زبان فارسی را فهرست کرده و در کنار هر کدام از آن‌ها ظرفیت نحوی را هم آورده است؛ به این ترتیب، تنوعات معنایی افعال را مشخص کرده است. برای مثال، می‌توان به فعل «زد» با چهار ظرفیت اشاره کرد:

□	..#زد
....	<فا،(مف)>[+/-]،(مفح)>[بر به را روی زیر]<
....	<فا ، تم>
....	<فا ، مفح>[از به]<
....	<فا>

### ۲-۳-۲ پیکره‌های برچسب‌خورده غیرنحوی زبان فارسی

پیکره‌های غیرنحوی زبان فارسی شامل پیکره گزاره‌های معنایی، گفتمانی، موجودیت‌های نامدار و مرجع‌دارها<sup>۱</sup> داده‌هایی هستند که معمولاً اطلاعات دستوری و زبان‌شناختی را مشخص می‌کنند. از این رو، حتی برچسب‌های به کاررفته در این پیکره‌ها هم می‌تواند برای زبان‌آموزان نامأنوس باشد. اما مؤلفان و پژوهشگران این حوزه می‌توانند از این اطلاعات ارزشمند در ارتباط با مهارت‌های مختلف، درس‌ها و تمرین‌هایی را طراحی کنند. مثلاً از موضوع گفتمان و روابط میان جملات و بندها می‌توان در مهارت نوشتن استفاده کرد و برای زبان‌آموز از یک طرف نحوه ایجاد متنیت در متن و به طور مشخص ایجاد رابطه منطقی میان جملات و استفاده از حروف ربط را توضیح داد؛ از سوی دیگر، برای تمرین از

۱. این پیکره‌ها در سایت [peykaregan.ir](http://peykaregan.ir) در دسترس‌اند.



نمونه‌های واقعی زبانی که در پیکره‌ها موجود هستند، نمونه آورد. بدیهی است که بهره‌برداری از این پیکره‌ها در آموزش زبان موضوعی است که می‌تواند برای پژوهش‌های آینده پیشنهاد شود. در ادامه، این پیکره‌ها به اختصار معرفی می‌شوند.

در پیکره گزاره‌های معنایی زبان فارسی لایه‌ای از اطلاعات معنایی محمول-موضوع روی پیکره برچسب‌خورده صرف و نحو دادگان زبان فارسی قرار گرفته است. در نتیجه، در هر جمله محمول و فعل مشخص می‌شود و متناسب با آن موضوع‌های معنایی و سپس افزوده‌های موجود تعیین می‌گردد.

در پیکره گفتمان زبان فارسی روابط میان جملات و اساساً روابط منطقی مطرح در متن بررسی می‌شود. نسخه (۱) این پیکره، حاوی ۳۰ هزار جمله مستقل و نیم‌میلیون واژه است و برچسب‌گذاری در آن به تعیین روابط منطقی میان بندهای درون یک جمله اختصاص یافته است. نسخه شماره (۲) هم حدود نیم‌میلیون کلمه دارد؛ اما محتوای زیرساختی آن متن است. در این پیکره برای اطلاعات مربوط به منبع خبر و نوع و جهت جملات نیز برچسب‌گذاری صورت گرفته است.

پیکره موجودیت‌های نامدار برای تعیین اسم‌های خاصی است که به افراد، مکان‌ها، سازمان‌ها، مقدارهای عددی (شامل زمان، اندازه و...)، حیوان، فن‌واژه، رویداد یا هر اسم خاص اشاره می‌کند. این برچسب‌ها را باید بر اساس بافت انتخاب کرد؛ زیرا بسیاری از اسامی در جملات مختلف به نامدارهای متفاوتی اشاره می‌کنند. مثلاً دهخدا هم اسم خاص است و هم نام سازمان است.

در پیکره مرجع‌دارهای زبان فارسی عوامل انسجام متنی مثل ارجاع، جایگزینی، انسجام واژگانی و مانند آن مشخص و برچسب‌گذاری شده است تا عوامل ایجاد متن و متنیت به این ترتیب مشخص شود.

### ۳- برهم‌کنش زبان‌شناسی پیکره‌ای و آموزش زبان

در ادامه، تعامل زبان‌شناسی پیکره‌ای با آموزش زبان را با تمرکز بر پژوهش‌های صورت‌گرفته در حوزه آموزش زبان فارسی بررسی می‌کنیم.

#### ۳-۱ کاربردهای پیکره فارسی‌آموز در آموزش زبان فارسی

همان‌طور که گفتیم پیکره زبان‌آموز شامل دست‌نوشته‌های زبان‌آموزان خارجی است که با برچسب خطاهای زبانی همراه شده‌اند. تحلیلگران خطاهای زبانی با دسترسی به انواع خطاها و با توجه به زبان اول، سطح زبانی و دیگر ویژگی‌های مرتبط زبان‌آموزان به دسته‌بندی و تحلیل خطا می‌پردازند و از پس

این تحلیل‌ها اطلاعات ارزشمند بسیاری به دست می‌آید که می‌تواند زمینه‌ساز پژوهش‌های نظری و کاربردی بعدی باشد.

دستیابی به ویژگی‌های زبان میانی یکی از سودمندی‌های پیکره زبان‌آموز است. سجادی و صحرایی (۱۳۹۷) در مقاله خود با عنوان «سلسله‌مراتب دسترسی گروه اسمی: فراگیری بندهای موصولی در زبان فارسی»، با بررسی زبان میانی فارسی‌آموزان غیرایرانی، ترتیب فراگیری انواع بندهای موصولی زبان فارسی به زبان‌آموزان خارجی را بررسی کردند. آن‌ها برای این منظور در ۴۹۳ متن نوشتاری از پیکره فارسی‌آموزان غیرایرانی بنیاد سعدی، بندهای موصولی را برچسب زده‌اند. سپس بندهای موصولی بر اساس سطح زبان‌آموزان به ۶ سطح پایه، پیش‌میانی، میانی، فوق‌میانی، پیشرفته و فوق‌پیشرفته تقسیم شده است. نتایج پژوهش نشان می‌دهد که فراگیری بندهای موصولی فاعلی و مفعولی مستقیم، از سلسله‌مراتب دسترسی کینان و کامری<sup>۱</sup> پیروی می‌کند؛ درحالی‌که در دیگر بندهای موصولی شواهدی مبنی بر عدم تطابق با این سلسله‌مراتب دیده می‌شود.

## ۲-۳ کاربردهای پیکره آموزشی زبان در آموزش زبان فارسی

پیکره‌های آموزشی به شکل‌های مختلف می‌توانند در آموزش زبان مفید باشند؛ مثل استخراج واژه‌های پایه یا پرکاربرد زبان که خود در تهیه مواد درسی کارآمد است. پیکره‌ها به‌عنوان یک ابزار کمک‌آموزشی می‌توانند در اختیار زبان‌آموزان قرار گیرند تا آن‌ها در هر سطحی که هستند به‌مثابه یک زبان‌آموز زبان‌پژوه باتوجه به اهداف زبان‌آموزی از پیکره بهره ببرند. همچنین زبان‌آموز با دراختیارداشتن پیکره مناسب، می‌تواند معنی افعال و هر واژه دیگر را در بافت، یاد بگیرد یا ساخت‌های دستوری مختلف و ترتیب واژه‌ها را در جملات پرسشی، امری و خبری باتوجه به داده طبیعی زبان بشناسد.

برای دستیابی به اطلاعات زبانی درون پیکره‌های آموزشی می‌توان از ابزارهای مختلفی چون *انت‌کانک*<sup>۲</sup>، *لنکس‌باکس*<sup>۳</sup>، *ورد/اسمیت*<sup>۴</sup>، *اسکچ/نچین*<sup>۵</sup> و مانند آن استفاده کرد (میرزائی، ۱۳۹۶). در ادامه، به برخی از فعالیت‌های آموزشی مرتبط با پیکره آموزشی زبان اشاره می‌شود.

استخراج واژه‌های کاربردی و استخراج واژه‌های پایه یکی از پرسابقه‌ترین و مهم‌ترین کاربردهای پیکره‌های آموزشی است. صحرایی و دیگران (۱۳۹۷)، ذوالفقارکندری و دیگران (۱۳۹۸) برای استخراج

1. E. L. Keenan & B. Comrie

2. AntConc

3. Lancsbox

4. WordSmith

5. Sketch Engine

فهرست‌های واژه‌ای از پیکره‌های زبانی استفاده کرده‌اند. نکته مهم این است که در آموزش باید از پیکره‌های کارویژه و مرتبط استفاده کرد؛ در غیراین صورت، پیکره باید بسیار حجیم باشد تا بتواند واژه‌های زبان را نمایندگی کند.

دسترسی به محتوای واژگانی تنها سودمندی پیکره‌ها نیست، بلکه دسترسی به توالی‌های پرتکرار هم بسیار ارزشمند است. میرزائی (۱۴۰۰) برای بررسی واژه‌های محتوایی، توالی‌های دو، سه و چهارواژه‌ای واژه‌های محتوایی را بررسی کرده است. هدف این پژوهش تأکید بر اهمیت آموزش زبان پیکره‌یار و سنجش میزان کاربردی بودن پیکره‌ها در زبان‌آموزی است. در این پژوهش توالی‌های پیکره همشهری، توالی‌های یک پیکره کوچک از متون داستانی و مجموعه‌داده‌ای از متون ورزشی باهم مقایسه شد. برای این منظور ابتدا از ۵۰ فایل پیکره همشهری، پرتکرارترین توالی‌های دوواژه‌ای با استفاده از نرم‌افزار انت‌کانک و سربرگ چندپشته استخراج شد.

در جدول (۱) این توالی‌ها ذکر شده‌اند. پس‌از آن، پرتکرارترین خوشه‌های سه‌واژه‌ای و پربسامدترین خوشه‌های چهارواژه‌ای نیز استخراج شد. نتایج نشان داد پرتکرارترین خوشه‌های گزارش شده همگی در موضوع خبر بودند و این یعنی نوع پیکره بر محتوای واژگانی آن اثر دارد. بدیهی است محتوای واژگانی برگرفته از پیکره همشهری و رشته‌های واژگانی مستخرج از آن برای زبان‌آموزی با اهداف عمومی کافی و مناسب نیست؛ ولی می‌تواند محتوای آموزشی مناسبی را برای حوزه‌های مختلف خبرنویسی و مانند آن در اختیار قرار دهد.

جدول (۳). خوشه‌های دو، سه و چهارواژه‌ای در منتخبی از پیکره همشهری

واژه‌های محتوایی دوواژه‌ای	فراوانی	واژه‌های محتوایی سه‌واژه‌ای	فراوانی	واژه‌های محتوایی چهارواژه‌ای	فراوانی
جمهوری اسلامی	۷۰۶	مجلس شورای اسلامی	۴۲۹	مجمع تشخیص مصلحت نظام	۱۱۲
خبرنگار همشهری	۶۲۲	جمهوری اسلامی ایران	۳۶۷	دوره مجلس شورای اسلامی	۱۰۱
شورای اسلامی	۵۱۶	گزارش خبرنگار همشهری	۳۰۰	ششمین دوره مجلس شورای	۹۸
گزارش خبرنگار	۴۷۱	خبرگزاری جمهوری اسلامی	۱۸۴	انتخابات ششمین دوره مجلس	۷۶
مجلس شورای	۴۳۵	گروه علمی فرهنگی	۱۷۵	تیم ملی فوتبال ایران	۷۱

در مرحله بعد، برای بررسی بیشتر، پیکره‌های داستانی بررسی شد. نتایج نشان داد، پرتکرارترین خوشه‌های دوواژه‌ای این پیکره مانند «خورشید خانم»، «خرس کوچولو»، «روزی روزگاری» و توالی سه واژه‌ای مانند «گوهر شب چراغ»، «گفت نگران نباش» و چهارواژه‌ای مانند «یکی بود یکی نبود» همگی

در گونه<sup>۱</sup> روایتگری و داستان است. نتیجه این جست‌وجو با جست‌وجوی خوشه‌های چندواژه‌ای در پیکره همشهری که همگی با موضوع خبر بود کاملاً متفاوت و از نوع داستانی بود. بنابراین، انطباق محتوای واژگانی پیکره‌ها با موضوع آن مورد تأیید قرار گرفت.

تا اینجا مشخص شد که اگرچه پیکره‌ها برای آموزش واژه‌های محتوایی بسیار مهم هستند، هر پیکره‌ای با هر موضوعی نمی‌تواند برای زبان‌آموزی با اهداف عمومی مناسب باشد؛ اما برای زبان‌آموزی با اهداف ویژه بسیار کارآمد است. میرزائی (۱۴۰۰) برای اطمینان از مقایسه دو پیکره همشهری و داستانی و نتایج به دست آمده، یک مجموعه داده ورزشی از خبرگزاری ایران گردآوری کرد. توالی پرتکرار واژه‌های دوتایی همچون «تیم ملی»، «فدراسیون فوتبال» و «جام جهانی»، سه تایی همچون «تیم ملی بسکتبال» و «لیگ قهرمانان اروپا» و چهارواژه‌ای مانند «کادر فنی تیم ملی» و «وزیر ورزش و جوانان» همگی ورزشی بودند. در نتیجه، تأثیر موضوع داده‌های پیکره بر محتوای واژگانی آن‌ها بار دیگر تأیید شد؛ البته مشخص شد که در تهیه و تدوین محتوای آموزشی برای اهداف ویژه و برای تهیه متون تخصصی و محتوای واژگانی مرتبط می‌توان از توالی‌های واژه‌ای استفاده کرد. ایده تهیه درس‌نامه‌ها و محتواهای آموزشی برای آموزش زبان تخصصی از نتایج این پژوهش است و چشم‌انداز روشنی از شیوه کار در تهیه محتواهای آموزشی در این حوزه به دست می‌دهد.

سودمندی دیگر پیکره‌های آموزشی در ارتباط با واژه‌های نقشی و توالی‌های واژه‌ای مرتبط با آن‌ها قابل طرح است. میرزائی (۱۳۹۸) به بررسی ویژگی‌های پیکره آموزشی زبان در آموزش واژه با تأکید بر واژه‌های نقشی پرداخته است. بررسی پیکره‌های زبانی نشان می‌دهد که همواره و در هر پیکره‌ای، واژه‌های نقشی جزء پربسامدترین واژه‌ها هستند. همچنین در میان نقشی‌ها، حروف اضافه بالاترین رتبه را دارند. در پژوهش مورد اشاره، توالی‌های واژه‌ای شامل پنج حرف اضافه «در، به، از، با و برای» در ۵۰ فایل از پیکره همشهری بررسی شد. تعداد این حروف در ۵۰ فایل همشهری مطابق با گزارش جدول (۴) است.

جدول (۴). فراوانی حروف اضافه بسیط و پربسامد

حروف اضافه	در	به	از	با	برای
فراوانی	۷۱۴۶۷	۴۹۴۶۹	۳۳۸۳۸	۱۷۲۲۱	۸۵۵۱

بررسی داده‌ها و توالی واژه‌ای در این پژوهش نشان داد حرف اضافه «در» می‌تواند در معنای زمانی

(مانند «در سال»، «در طول» یا «در پایان») یا در معنای مکانی (مانند «در کشور» یا «در جامعه») باشد. همچنین این حرف اضافه می‌تواند در ترکیب‌هایی مانند «در مورد»، «در زمینه»، «در جهت»، «در پی» و مانند آن ظاهر شود. علاوه بر آن، برخی از ترکیب‌های «دری» در نقش قیدی یا حرف ربطی ظاهر شده‌اند و قابل جایگزینی با حروف ربط هستند. بدیهی است این نتایج در آموزش دستور چه در روش‌های معنما محور و چه به صورت مستقیم می‌تواند به عنوان الگو مورد توجه قرار گیرد.

عمده توالی‌های «به» ای، حروف اضافه ترکیبی و قابل جایگزینی با حرف اضافه بسیط هستند؛ مانند «به عنوان»، «به صورت»، «به دلیل»، «به طور»، «به منظور». در جدول (۵)، دوازده توالی پرتکرار حاوی حرف اضافه «به» قابل مشاهده است.

جدول (۵). خوشه‌های دوازده‌ای حاوی حرف اضافه «به» در پیکره همشهری

ترکیبات	فراوانی	ترکیبات	فراوانی
به عنوان	۱۴۲۲	به منظور	۳۸۸
به گزارش	۱۱۳۱	به گفته	۳۴۱
به صورت	۵۵۸	به همین	۳۳۴
به دلیل	۵۲۸	به نام	۳۰۹
به نظر	۴۷۹	به خاطر	۲۴۷
به طور	۶۶۸	به خود	۲۳۵

بررسی خوشه‌های واژه‌ای «از»ی نشان داد که این حرف، عمدتاً به عنوان مبدأ زمانی و مکانی به کار می‌رود. همین‌طور برای انتخاب عضوی از یک مجموعه نیز استفاده می‌شود؛ مانند «از بین» و «از میان»، «از اعضای». همچنین برخی خوشه‌های واژه‌ای «از»ی در معنای حرف اضافه مرکب به کار می‌روند؛ مانند «از جمله»، «از نظر»، «از لحاظ». برخی از این حروف مرکب قابل جایگزینی با حروف ساده هستند. مثلاً «توسط» می‌تواند جایگزین «از طریق» شود.

بررسی خوشه‌های واژه‌ای «با»ی نیز در جست‌وجوی پیکره همشهری نشان داد که توالی‌های واژه‌ای الگوی منظمی دارند. عمده کلمه‌هایی که همراه با حرف اضافه «با» هستند در معنای فعلی هستند و نقش محمولی می‌گیرند؛ مانند خوشه‌های دوتایی «با اعلام»، «با شرکت»، «با اجرای»، «با استفاده»، «با تلاش» و «با تأکید».

همچنین بررسی خوشه‌هایی دوتایی حاوی حرف اضافه «برای» نشان داد که این حرف اضافه نیز مانند «با» بیشتر در همراهی با محمول‌ها اسمی قرار می‌گیرد؛ مانند «برای شرکت»، «برای ایجاد»، «برای

حضور»، «برای اجرای»، «برای کاهش» و «برای رفع». در فهرست گزارش شده پنج ترکیب دیگر هم وجود دارد («برای من»، «برای مثال»، «برای نمونه» و «برای این که») که آن‌ها نیز الگوی معنادار دیگری را به دست داده‌اند. مثلاً در «برای من» معنای مالکیت مطرح شده است.

به این ترتیب، در کنار ارائه فهرست‌های واژگانی، دسترسی به چندپشته‌ها و خوشه‌های واژه‌ای پرتکرار زبان از دیگر امکانات مؤثر و مناسبی است که پیکره‌های زبانی و ابزارهای وابسته به آن می‌تواند در اختیار زبان‌آموزان، مدرسان و مؤلفان محتوای آموزشی قرار دهد.

### ۳-۳ کاربردهای پیکره برچسب‌خورده زبان فارسی در آموزش زبان فارسی

پیکره‌های برچسب‌خورده معمولاً با هدف یادگیری ماشینی تهیه می‌شوند؛ اما مشخص است که در آموزش زبان هم بسیار اهمیت دارند. برچسب‌های این پیکره‌ها در سطوح مختلف صرف، نحو، معنا و مانند آن مطرح هستند. در زبان فارسی یکی از این پیکره‌ها پیکره دادگان وابستگی (رسولی و دیگران، ۲۰۱۳) است.

میرزائی (۱۴۰۰) با توجه به اهمیت حروف اضافه و با استفاده از پیکره مذکور تلاش کرده است با بررسی کارکردهای نقشی و معنایی سه حرف اضافه «از»، «به» و «برای» الگوی مناسبی را برای آموزش زبان فارسی به دست دهد. بررسی‌های این پژوهش و پایش پیکره نحوی نشان داد در زبان فارسی حرف اضافه «از» در جایگاه افزوده یا متمم در مفاهیم مختلف «مکان»، «زمان»، «مبدأ»، «مسیر»، «منشأ»، «ابزار» و «علت» نمایان می‌شود. افزوده‌ها و متمم‌های «به» ای در مفاهیم مختلفی مانند «ابزار»، «زمان»، «مکان»، «سبب»، «علت»، «شرط»، «مقدار»، «محرک»، «بهره‌ور» و «مقصد» ظاهر می‌شوند و حرف اضافه «برای» هم در این پیکره در مفاهیم «مکان»، «زمان»، «علت»، «منظور»، «بهره‌ور»، «مقدار» مطرح شده بود. در این بررسی ارائه این دسته‌بندی‌ها با نمونه‌های کافی تفاوت‌های معنایی حروف اضافه را به خوبی نشان داده است و این اطلاعات به خصوص در آموزش زبان می‌تواند مورد توجه مؤلفان محتواهای آموزشی باشد.

حروف اضافه در ساخت واژه‌های زبان فارسی به‌ویژه در ساخت فعل‌های مرکب هم به کار می‌روند. بررسی پیکره وابستگی نحوی زبان فارسی نشان داد حروف اضافه «از»، «به» و «در» در ساختمان افعال مرکب، به شرحی که در جدول (۶) قابل مشاهده است، به کار می‌رود.

جدول (۶). فعل‌های مرکب با حرف اضافه «از»، «به» و «در»

«از»	«به»	«در»
از پا افتاد	به چشم آمد	در بر داشت
از چشم افتاد	به وجود آمد	در نظر آورد
از قلم افتاد	به دنیا آمد	در نظر گرفت
از کار افتاد	به دست آمد	در پی داشت
از شکل افتاد	به نظر آمد	در میان گذاشت
از بین برد	به جان آمد	در جریان گذاشت
از میان برد	به تنگ آمد	در پیش گرفت
از پیش برد	به یاد داشت	در بر گرفت
از یاد برد	به اجرا گذاشت	در امان ماند
از بین رفت	به کار برد	در میان نهاد
از دست رفت	به نظر رسید	-----

طبق جدول، حرف اضافه «از» با فعل‌های «افتاد»، «برد» و «رفت» فعل مرکب می‌سازد. فعل‌های «افتاد»، «برد» و «رفت» در شکل بسیط، متمم حرف‌اضافه‌ای «از»ی دارند و در اینجا که حرف اضافه جزئی از ساختمان فعل است، معنای استعاری به دست آمده است. مثلاً «از» در «از چشم افتاد» معنای استعاری دارد.

در ستون دوم جدول بالا فعل‌هایی که در ساختمان آن‌ها حرف اضافه «به» وجود دارد، گزارش شده است. فعل‌های بسیط «آمد»، «آورد»، «برد» و «رسید» نیز دارای متمم حرف‌اضافه‌ای «به»ای هستند و اینجا در ترکیب «به» با این افعال معنی استعاری داریم. یعنی در «به چشم آمد» دیدن یا به‌نظررسیدن مدنظر است و معنای حرکت مکانی مطرح نیست. در همراهی حرف اضافه «در» با فعل‌های «داشت»، «آورد»، «گذاشت»، «گرفت»، «ماند» و «نهاد» نیز این‌گونه است. تعریف و تدریس معنای غیراستعاری این افعال، سپس تمرکز بر فعل‌های حرف‌اضافه‌ای و تمرکز بر معناهای استعاری حاصل، خود می‌تواند در آموزش دستور مبنای عمل قرار گیرد.

حروف اضافه آن‌چنان‌که گفته شد می‌تواند در جایگاه متمم قرار بگیرد. بررسی افعال همراه با متمم‌های حرف‌اضافه‌ای‌شان برای آموزش و ارائه الگوهای آموزشی مناسب هستند. در همین زمینه، صحرائی و میرزائی (۱۴۰۱) با هدف ایجاد فهرست‌های دستوری مناسب در فارسی‌آموزی به دسته‌بندی معنایی فعل‌های ساده با ظرفیت مفعول حرف‌اضافه «به»ای تلاش کرده‌اند در پژوهشی پیکره‌بنیاد به

دسته‌بندی افعال ساده دارای متمم پردازند تا بر اساس آن زمینه مناسبی را در آموزش افعال ساده و تنوعات معنایی آن‌ها ایجاد کنند. در نهایت، هفت دسته معنایی فعلی شامل «مفهوم نمود و وجه»، «معنای لمس کردن یا اتصال و تماس با سطوح»، «معنای عرضه کردن و انتقال مالکیت مادی و معنوی»، «معنای اظهار و بیان داشتن»، «معنای حرکت و جابه‌جایی»، «معنای تناول» و «معنای افعال مرتبط با حواس پنج‌گانه» گزارش شده است. بدیهی است دستیابی به این معانی در آموزش زبان و تدریس فعل به‌عنوان هسته اصلی جمله مفید خواهد بود. همین‌طور چنین پژوهش‌هایی می‌تواند به‌صورت نظام‌مند و با برنامه‌ریزی قبلی در تولید محتوای آموزشی دریچه‌ای جدید باز کند.

#### ۴- بحث و نتیجه‌گیری

در این مقاله پیکره زبان‌آموز و پیکره آموزشی زبان معرفی و سودمندی‌های این نوع پیکره‌ها در فارسی‌آموزی بیان شد. بررسی‌ها نشان داد استفاده از پیکره‌های برجسب‌خورده می‌تواند در آموزش دستور و آموزش قالب‌های نحوی و معنایی کارگشا باشد. بررسی‌ها نشان داد پیکره نحوی زبان فارسی الگوهای مناسبی را برای آموزش حرف اضافه، همچنین فعل‌های مرکب حرف اضافه‌ای در اختیار قرار می‌دهد. استفاده از پیکره‌های آموزشی زبان در تهیه محتوای درسی برای واژه و آموزش واژه در بافت مفید است. هر دوی این پیکره‌ها به‌عنوان ابزار کمک آموزشی می‌تواند در اختیار زبان‌آموز باشد و او را در یادگیری زبان یاری برساند.

همچنین نشان دادیم که پیکره‌های زبان‌آموز و خطاهای برجسب‌خورده آن امکاناتی هستند که در تولید منابع درسی به کار می‌آیند و حتی می‌توانند سیر تکوین یادگیری مباحث زبانی را نشان دهند. می‌توان گفت انواع پیکره‌های زبانی شامل پیکره زبان‌آموز، پیکره آموزشی زبان و پیکره‌های برجسب‌خورده همگی در آموزش زبان به شکل‌های مختلف مفید و راهگشا هستند. تهیه پیکره‌های آموزشی مناسب، با توجه به اهداف زبان‌آموزان می‌تواند به‌عنوان چشم‌انداز پژوهش‌های آتی دیده شود. همچنین بهره‌گیری هرچه بیشتر از پیکره‌های برجسب‌خورده زبان فارسی در ارائه الگوهای آموزشی می‌تواند از مطالعات پیش‌رو باشد. برجسب‌گذاری تحلیل خطای پیکره‌های فارسی‌آموز و سپس بهره‌برداری از آن‌ها با توجه به سطح زبان‌آموز، زبان اول، اهداف زبان‌آموزی و مانند آن، از موضوع‌های دیگری است که برای مطالعات آینده پیشنهاد می‌شود.

#### منابع

بی‌جن‌خان، محمود؛ نصری، عباس؛ جلالی، شهره (۱۳۹۳). نقش واژگان بسامدی در ارزیابی مهارت واژگانی



- فارسی‌آموزان. پژوهش‌نامه آموزش زبان فارسی به غیرفارسی‌زبانان، ۳(۲) (۷ پیاپی)، ۲۵-۴۵.
- بکنوف، استم؛ میرزایی حصاریان، محمدباقر؛ گل‌پور، لیلیا (۱۳۹۸). خطاهای نوشتاری فارسی‌آموزان قزاق: مطالعه‌ای برپایه انگاره کوردر. پژوهش‌های زبانی، ۱۰(۲)، ۴۵-۶۵.
- جهانگردی، کیومرث؛ عاصی، مصطفی؛ افراشی، آزیتا؛ وکیلی‌فرد، امیررضا (۱۳۹۵). واژه در کتاب‌های آموزش زبان فارسی به غیرفارسی‌زبانان: پژوهشی پیکره‌بنیاد. پژوهش‌نامه آموزش زبان فارسی به غیرفارسی‌زبانان، ۵(۲) (پیاپی ۱۲)، ۳-۲۶.
- ذوالفقار کندری، زهره؛ موسوی میانگاه، طیبه؛ روشن، بلقیس (۱۳۹۸). مقایسه روش‌های کلاسیک و روش‌های مبتنی بر اندازه‌های آماری پیکره‌بنیاد در استخراج خودکار واژه‌های پایه علوم پزشکی به روش بسامدی. پژوهش‌نامه آموزش زبان فارسی به غیرفارسی‌زبانان، ۸(۱) (۱۷ پیاپی)، ۲۲۷-۲۴۴.
- رضاپور، ابراهیم؛ موجانی، ابوالفضل (۱۴۰۰). تأثیر جنسیت در کاربرد افزوده‌های وجهی در نوشتار فارسی‌آموزان خارجی سطح پیشرفته از منظر دستور نقش‌گرایی. پژوهش‌های زبانی، ۱۲(۱)، ۱۲۱-۱۴۳.
- سجادی، شهره السادات؛ صحرائی، رضامراد (۱۳۹۷). سلسله‌مراتب دسترسی گروه اسمی: فراگیری بندهای موصولی در زبان فارسی. پژوهش‌های زبانی، ۹(۱)، ۲۱-۳۸.
- صحرائی، رضامراد؛ طالبی، مروارید؛ مجیری، امیرحسین (۱۳۹۶). مقایسه واژه‌های پایه زبان فارسی در شش پژوهش. پژوهش‌نامه آموزش زبان فارسی به غیرفارسی‌زبانان، ۶(۱)، ۱۱۵-۱۳۴.
- صحرائی، رضامراد؛ مجیری فروشان، امیرحسین؛ طالبی، مروارید (۱۳۹۷). واژه‌های پایه زبان فارسی بر متون مطبوعاتی. زبان‌پژوهی، ۱۱(۳۳)، ۳۵۳-۳۷۸.
- صحرائی، رضامراد؛ میرزائی، سمیرا (۱۴۰۱). طبقه‌بندی معنایی افعال ساده دارای متمم حرف‌افزافه «به»‌ای: یک بررسی پیکره‌بنیاد. مطالعات زبان‌ها و گویش‌های غرب ایران، ۱۰(۳)، ۲۳-۴۳.
- صحرائی، رضامراد (۱۳۹۹). زبان فارسی را چگونه آموزش دهیم. تهران: نشر نویسه پارسی.
- صفری، سعید (۱۳۹۱). طراحی و ایجاد پیکره تولیدی زبان‌آموز فارسی. پایان‌نامه کارشناسی‌ارشد آموزش زبان فارسی به غیرفارسی‌زبانان، دانشگاه علامه طباطبائی.
- صفری، سعید (۱۳۹۵). پیکره زبان‌آموز: مبانی، روش‌شناسی، الگوی طراحی و تولید. در مجموعه مقالات دومین همایش ملی زبان‌شناسی پیکره‌ای، (۹۳-۱۲۴).
- میرزائی، آزاده (۱۳۹۶). آشنایی با زبان‌شناسی پیکره‌ای. تهران: دانشگاه علامه طباطبائی.
- میرزائی، آزاده (۱۳۹۹). فرهنگ توصیفی زبان‌شناسی پیکره‌ای. تهران: نشر علمی.
- میرزائی، سمیرا (۱۳۹۸). آموزش واژه‌های نقشی به فارسی‌آموزان خارجی با بهره‌گیری از پیکره زبانی. مجموعه مقالات سومین همایش ملی زبان‌شناسی پیکره‌ای، (۲۵۳-۲۶۹).

میرزائی، سمیرا (۱۴۰۰). آموزش بیکره‌بنیاد واژه به فارسی‌آموزان خارجی با تمرکز بر حروف اضافه. پایان‌نامه کارشناسی ارشد آموزش زبان فارسی به غیرفارسی‌زبانان، دانشگاه علامه طباطبائی.

نویدی، امین؛ عامری، حیات؛ ابوالحسنی چیمه، زهرا (۱۴۰۰). مقایسه واژه‌های آموزشی کتاب‌های آموزش زبان فارسی با واژه‌های پربسامد، پژوهش‌نامه آموزش زبان فارسی به غیرفارسی‌زبانان، ۱۰(۲)، ۲۱۳-۲۳۵.

## Reference

- Bakenov, A., Mirzaei Hassarian, M. B., & GolPour, L. (2020). Writing errors analysis of Kazakh Persian language learners based on Corder model. *Journal of Language Research, 10*(2), 45-65. (In Persian)  
<https://doi.org/10.22059/jolr.2020.285069.666528>
- Bijankhan, M., Nasri, A., & Jalai, Sh. (2014). The role of frequency lexicon in assessing lexical. *Journal of Teaching Persian to Speakers of Other Languages, 3*(2 (7)), 25-45. (In Persian)
- Boontam, P., & Phoocharoensil, S. (2018). *Effectiveness of English preposition learning through data-driven learning (DDL)* (Doctoral dissertation, Thammasat University).
- Boulton, A. (2009). Testing the limits of data-driven learning: Language proficiency and training. *ReCALL, 21*(1), 37-54. doi:10.1017/S0958344009000068
- Gablasova, D., Brezina, V., & McEnery, T. (2017). Collocations in corpus-based language learning research: Identifying, comparing, and interpreting the evidence. *Language Learning, 67*(1), 155-179. <https://doi.org/10.1111/lang.12225>
- Gavioli, L., & Aston, G. (2001). Enriching reality: Language corpora in language pedagogy. *ELT Journal, 55*(3), 238-246. <http://dx.doi.org/10.1093/elt/55.3.238>
- Jahangardi, K., Assi, M., Afrashi, A., & Vakilifard, A. (2016). Vocabulary in the textbooks of teaching Persian to non-Persian speakers: A corpus-based study. *Proficiency of Persian Language Learners Journal of Teaching Persian to Speakers of Other Languages, 5*(2 (12)), 3-26. (In Persian)
- Johansson, S. (2008). Contrastive analysis and learner language: A corpus-based approach. *University of Oslo, 9*.
- James, C. (1998) *Errors in language learning and use: Exploring error analysis*. Longman, Essex.
- Johns, T. (1991). From printout to handout: Grammar and vocabulary teaching in the context of data-driven learning. *English Language Research Journal, 4*, 27-45.
- Lin, M. H. (2021). Effects of data-driven learning on college students of different grammar proficiencies: A preliminary empirical assessment in EFL classes. *Sage Open, 11*(3). <https://doi.org/10.1177/21582440211029936>
- Mirzaei, A. (2017). *An introduction to corpus linguistics*. Tehran: Allameh Tabataba'i University. (In Persian)
- Mirzaei, A. (2020). *Descriptive dictionary of corpus linguistics*. Tehran: Nashre Elmi. (In Persian)
- Mirzaei, S. (2019) Teaching functional words to non-Persian language learners using a pedagogical corpus. *The Proceeding of the Third National Corpus Linguistics Conference* (pp. 253-269). (In Persian)

- Mirzaei, S. (2021) *Corpus-based vocabulary teaching to non-native Persian language learners by focusing on preposition*. M.A. in teaching Persian to non-Persian speakers, Allameh Tabataba'i University. (In Persian)
- Moon, R. (2010). What can a corpus tell us about lexis. In M. McCarthy, & A. O'Keeffe (Eds.), *The Routledge Handbook of Corpus Linguistics* (pp. 197-211). Abingdon: Routledge.
- Mushtaq, M., Mahmood, M., Kamran, M., Ismail, M., Ahmad, M., & Adeel, H. (2019). A corpus-based analysis of EFL learners' errors in written composition at intermediate level. *Indian Journal of Natural Sciences*, 9(52), 16842-16852.
- Navidi A., Ameri, H., & Abolhassani Chime, Z. (2021). Comparison of vocabulary in the textbooks of teaching Persian to non-Persian speakers with Persian basic words. *Journal of Teaching Persian to Speakers of Other Languages*, 10 (2) 213-235. (In Persian) <https://doi.org/10.30479/jtpsol.2022.16550.1566>
- Paker, T., & Özcan, Y. E. (2017). The effectiveness of using corpus-based materials in vocabulary teaching. *International Journal of Language Academy*, 5(14), 62-81. <http://dx.doi.org/10.18033/ijla.3494>
- Rasooli, M. S., Moloodi, A. S., Kouhestani, M., & Minaei-Bidgoli, B. (2011) A syntactic valency lexicon for Persian verbs: The first steps towards Persian dependency treebank. In *5th Language and Technology Conference (LTC): Human Language Technologies as a Challenge for Computer Science and Linguistics* (pp. 227-231).
- Rasooli, M. S., Kouhestani, M., & Moloodi, A. S. (2013). Development of a Persian syntactic dependency treebank. In *Proceedings of The 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL HLT)* (pp. 306-314). Atlanta, Georgia. Association for Computational Linguistics.
- Rezapour, E., & Mouchani, A. (2021). The effect of gender in the use of the modal adjuncts in the writings of advanced foreign learners of Persian language based on functional grammar. *Language Research*, 12 (1), 121-143. (In Persian) <https://doi.org/10.22059/jolr.2021.314508.666671>
- Safari, S. (2012). *Design and development of a learner corpus for Persian*. M.A. in teaching Persian to non-Persian speakers, Allameh Tabataba'i University. (In Persian)
- Safari, S. (2016). Learner corpora: An overview of research methodology, design and construction. *The Proceeding of the Second National Corpus Linguistics Conference*, 93-124. (In Persian)
- Sahraei, R. M. (2020). *How to teach Persian*. Tehran: Neveeseh Parci Publications. (In Persian)
- Sahraei, R. M., & Mirzaei, S. (2022). Classifying the simple verbs with the complement of the preposition /be/: A corpus-based research. *Research in Western Iranian Language and Dialects*, 10(3) 23-43. (In Persian) <https://doi.org/10.22126/jlw.2022.7181.1601>
- Sahraei, R. M., Mojiri Foroushani, A. H., & Talebi, M. (2020). Persian basic words based on texts by press. *Zabanpazhuhi*, 11(33) 353-378. (In Persian) <https://doi.org/10.22051/jlr.2019.19564.1520>

- Sahraei, R. M., Talebi, M., & Mojiri Foroushani, A. H. (2017). Basic words in Farsi: A comparison of six studies. *Journal of Teaching Persian to Speakers of Other Languages*, 6(1), 115-134. (In Persian)
- Sajjadi, sh., & Sahraei, R. M. (2018). Noun phrase accessibility: acquisition of relative clauses in Persian language. *Language research*, 9(1) 21-38 (In Persian) <https://doi.org/10.22059/jolr.2018.66648>
- Sánchez, A. M. (2013). A corpus-based analysis of errors in adult EFL writings. *Revista Nebrija de Lingüística Aplicada*, 13.
- Vyatkin, N. (2016). Data-driven learning for beginners: The case of German verb-preposition collocations. *ReCALL: Journal of Eurocall*, 28(2), 207-226. <https://psycnet.apa.org/doi/10.1017/S0958344015000269>
- Zulfaqar Kandari, Z., Mousavi Miangah, T., & Roshan, B. (2018). Comparison of classical methods and methods based on the statistical sizes of the body in the automatic extraction of the basic words of medical sciences by the frequency method. *Proficiency of Persian Language Learners Journal of Teaching Persian to Speakers of Other Languages*, 8(1 (17)), 227-244. (In Persian) <https://doi.org/10.30479/jtpsol.2019.9257.1393>