



Collocations in Persian Learner's Corpus: A Study Based on Phraseological Approach

Shahnaz Ahmadighader

PhD candidate in Teaching Persian to Non-Persian Speakers, Allameh Tabataba'i University, Tehran, Iran.

sh.ahmadi53@gmail.com

Rezamorad Sahraee *

Corresponding author, Professor in Linguistics, Allameh Tabataba'i University, Tehran, Iran.
sahraei@atu.ac.ir

Abstract:

The present study is a mixed method (quantitative-qualitative) and cross-sectional research that was conducted to evaluate how to use collocations by foreign Persian learners at six levels from introductory to advanced. The statistical population included 1112 texts from the tests taken at the Saadi Foundation (Linguistic corpus with 142,608 tokens), which were sampled by stratified random sampling method, 300 written texts with balanced volume in each Level. Various linguistic approaches were studied to analyze the sample corpus and identify collocations; Then, by examining and matching the criteria of the Phraseological approach and Farth's approach with the researcher's observations in the sample, a comprehensive definition of coincidence and criteria for identifying it in the productions of non-Iranian Persian learners were obtained. Then, the sample body was coded and analyzed in the MAXQDA software according to a specific coding system, and the research hypotheses were verified by interpreting the descriptive and inferential statistics obtained from the SPSS software. The study found that Persian learners, regardless of their proficiency level, use a similar number of collocations in their writing so the collocations are important from the very beginning of Persian learning. However, more common collocations were observed in the lower-level Persian learners' writing, and a list of them is provided in the article. While higher-level learners use a greater variety of collocations. These findings suggest that collocations are important from the beginning of language learning and should be incorporated into language instruction and assessment.

***Cite this article:** Ahmadighader, Shahnaz. Rezamorad, Sahraee. Collocations in Persian Learner's Corpu: A Study Based on Phraseological Approach. Vol. 14, No. 1 (Tome 29), April 2025,3-37.

DOI: 10.30479/jtpsol.2025.21222.1685

Received on: 24/11/2024

Accepted on: 24/01/2025



© The Author(s).

Publisher: Imam Khomeini International University

Introduction

The effective use of collocations is crucial for accurate and fluent language production (Wray, 2000). Collocations, as essential elements of pragmatic competence (Lewis, 2000), enable learners to act in diverse communication contexts effectively. The concept of collocation has been explored decades ago by linguists like Palmer (1933) and Firth (1957). Later, the phraseological approach, with its categorization of lexical combinations into terms "idioms", "collocations" and "free combinations", offers a more precise criteria for distinguishing collocations from other lexical groups (Eisenstadt, 1979; Covey, 1981; Nottinger & DeCarrico, 1992; Howarth, 1996). This approach provides a robust theoretical foundation for the current study. Jalili and Sahrai's (2012) corpus-based study, while measuring vocabulary diversity, did not specifically address collocation usage. A comprehensive corpus-based examination of Persian collocation usage in non-Iranian learner texts remains absent from the literature. While studies on English learner productions, such as those by Cowie (1992), Howarth (1996), Moon (1998), Jackendoff (1997), Altenberg (1998), Men (2018), and Lu (2016), have been more extensive, a dedicated focus on Persian collocations in non-Iranian learner texts is still needed. The primary objective of this study is to establish a comprehensive definition and precise criteria for identifying collocations in the written productions of non-native Persian learners. Additionally, the study aims to investigate the frequency and diversity of collocations employed by Persian learners at various proficiency levels. To address these objectives, the following research questions and hypotheses are proposed:

Research Question 1: Is there a statistically significant difference in the frequency of various Persian collocation types employed by learners at six distinct proficiency levels in their written productions?

Research Question 2: Is there a statistically significant difference in the use of frequent and infrequent collocations by Persian learners at six distinct language proficiency levels in their written productions?

Hypothesis 1: There is a statistically significant difference in the frequency of various Persian collocation types employed by learners at six distinct language proficiency levels: introductory, pre-intermediate, intermediate, upper-intermediate, advanced, and proficient.

Hypothesis 2: There is a statistically significant difference in the use of frequent and infrequent collocations by Persian learners at six distinct language proficiency levels.

Methodology

This study employed a mixed-methods (quantitative-qualitative) cross-sectional design. The statistical population comprised 1,112 texts from the Saadi Foundation (a corpus with 142,608 tokens). Using stratified random sampling, 300 written texts with a balanced volume in each category were sampled. To establish a comprehensive definition of collocations and clear identification criteria for identifying collocations from other word groups, a thorough review of linguistic approaches was conducted. The work of researchers such as Eisenstadt (1979, 1981), Van Roey (1990), Howarth (1996), Nesselhauf (2005), Laufer and Waldman (2011), and Men (2018) was considered. By adapting the definitions and criteria from phraseological and Firth's approaches, along with insights from the sample corpus, a robust framework was developed. The sample corpus was then input into MaxQDA software and subjected to multiple coding iterations using a coding system based on syntactic patterns. This process refined the definition and

identification criteria. The entire sample corpus was subsequently coded and analyzed according to the established criteria. Descriptive and inferential statistics, processed using SPSS software, were employed to verify the research hypotheses. To facilitate coding and analysis, lexical combinations were classified based on syntactic patterns, drawing upon the work of Hausman (1989) and Benson et al. (2010). A tailored classification of syntactic patterns was developed to suit the specific context of this research.

Results

One of the most important achievements of this research is the development of a comprehensive definition of Persian collocations and the establishment of clear criteria for recognizing various linguistic combinations. This framework, visually represented in Figures 1 and 2, facilitates the identification of collocations produced by Persian learners across different proficiency levels. The definition and criteria were derived through rigorous analysis of a substantial corpus. Collocations are defined as combinations of two or more words closely co-located in a text, relative semantic clarity, and at least one of the following characteristics: (1) limited co-occurrence range for one or more constituent elements, or (2) high frequency of co-occurrence in native speaker texts, coupled with a close syntactic relationship. Based on this definition, the following criteria were established for identifying collocations in the written work of foreign Persian learners:

- Distinction between collocations and idioms: Collocations possess relative semantic clarity.
- Distinction between collocations and free combinations: Collocations exhibit limited co-occurrence range and high frequency in native speaker texts.

Data analysis revealed that collocation usage rates do not vary significantly across six language proficiency levels, indicating their importance from the outset of language learning. Descriptive statistics indicate that in the total sample size (including 47,388 Tokens), 9,055 collocation tokens were observed. The frequency of collocations used in the written productions of different language level learners shows a gradual increase of 2% from the introductory level (14%) to the advanced level (16%), which is insignificant.

Also, to examine the frequent and infrequent collocations in the productions of Persian learners of different language levels, only well-formed collocations of the “noun-verb” were analyzed, and it was observed that 4,388 collocation tokens and 1,232 collocation types were used in the entire corpus under study. This frequency of collocation types (1232) compared to the number of collocation tokens (4388) means that on average each collocation types is repeated about 3 or 4 times in the texts under study; however, the graph below shows that the distribution of collocation types in the frequency range of 1 to 5 times (bottom of the graph) is much denser than the top of the graph. In other words, a small number of highly frequent collocation types and, on the contrary, a very large number of infrequent collocation types (with a frequency of less than five times) were observed in the sample corpus.

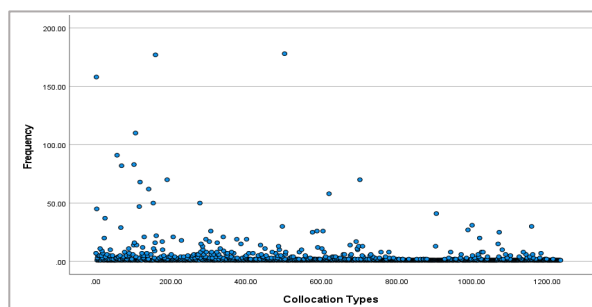


Figure: Scatter plot of the frequency of "noun-verb" collocations in Persian learners' productions

Comparative analysis of collocation diversity in learner texts indicated that each proficiency level exhibits significantly greater diversity than the preceding level, aligning with the findings of Getsaki (1999), Zhang (1993), and Men (2018).

Conclusion

The production of collocation reinforcement tasks has consistently been a focus of interest for educators. However, there remains uncertainty regarding the necessity of explicitly teaching learners about the limitations of free word combination and whether a behaviorist approach, emphasizing Audio-lingual teaching method and repetition of word clusters, is sufficient for collocation instruction. This study aims to provide a comprehensive definition and precise criteria for identifying collocations in the written productions of non-native Persian learners, and to examine the representation and diversity of collocations in the Persian productions of learners at different language levels. Through analysis of the sample corpus, significant findings were obtained. Based on these findings, we conclude that explicit instruction on the limitations of collocation selection and informing language learners about collocation patterns can contribute to the development of their vocabulary knowledge. Therefore, it is essential to explore appropriate pedagogical methods and develop curricula and instructional materials for teaching collocations. The findings of this study have implications for the development of educational resources and proficiency assessment tools for second language learners, and can also serve as a foundation for identifying various word group types in the field of second language collocation research.

Acknowledgment

The authors would like to sincerely thank the experts in this field who collaborated in assessing the reliability of the coding of the learner corpus of this study.

Conflict of Interest

The data in this study were extracted from texts produced by Persian learners in an educational context and were used without mentioning the learners' names or personal information.

Keywords: Collocation, Phraseological approach, Criteria for detecting collocations, Persian learner's corpus, Lexical diversity.



باهم‌آیی‌ها در پیکره نوشتاری فارسی آموزان خارجی: مطالعه‌ای بر پایه رویکرد عبارت‌شناختی (پژوهشی)

شهناز احمدی‌قادر

دانشجوی دکتری آموزش زبان فارسی به غیر فارسی‌زبانان، دانشگاه علامه طباطبائی، تهران، ایران.

sh.ahmadi53@gmail.com

رضامراد صحرانی*

نویسنده مسئول، استاد گروه زبان‌شناسی، دانشگاه علامه طباطبائی، تهران، ایران.

sahraei@atu.ac.ir

چکیده

مطالعه حاضر، پژوهشی آمیخته (کمی - کیفی) و مقطعی است که با هدف ارزیابی چگونگی به‌کارگیری باهم‌آیی‌ها توسط فارسی‌آموزان خارجی در شش سطح مقدماتی تا پیشرفته انجام شده است. جامعه آماری، ۱۱۱۲ متن (پیکره‌ای مشتمل بر ۱۴۲۶۰۸ مورداژه) از تولیدات نوشتاری فارسی‌آموزان بنیاد سعدی بود. نمونه‌گیری با روش تصادفی طبقه‌ای انجام شد و طی آن ۳۰۰ متن نوشتاری، با حجم مورداژه‌های متوازن در هر سطح، انتخاب شدند. برای تحلیل پیکره نمونه و شناسایی باهم‌آیی‌ها از دیگر گروه‌واژه‌ها، رویکردهای مختلف زبانی مطالعه شد؛ آن‌گاه با بررسی و متناسب‌سازی معیارهای رویکرد عبارت‌شناختی و فرض با مشاهدات پژوهش‌گر در پیکره نمونه، تعریف جامعی از باهم‌آیی و معیارهای روشنی برای شناسایی آن‌ها در تولیدات فارسی‌آموزان غیرایرانی به‌دست آمد. سپس، پیکره نمونه در نرم‌افزار مکس کیودی‌ای، کدگذاری و تحلیل شد و با تفسیر آمار توصیفی و استنباطی به‌دست آمده از نرم‌افزار اس‌پی‌اس‌اس، فرضیه‌های پژوهش مورد راستی‌آزمایی قرار گرفتند. یافته‌های پژوهش نشان داد؛ میزان به‌کارگیری باهم‌آیی در تولیدات فارسی‌آموزان در شش سطح زبانی، تفاوت معناداری ندارد و باهم‌آیی‌ها از ابتدای زبان‌آموزی دارای اهمیت هستند. همچنین، باهم‌آیی‌های پرتکرار در تولیدات فارسی‌آموزان سطوح زبانی پایین‌تر، بیشتر مشاهده شد که فهرستی از آن‌ها در مقاله آمده است. مقایسه تنوع به‌کارگیری باهم‌آیی‌ها در متون نوشتاری فارسی‌آموزان نیز نشان داد که تولیدات زبان‌آموزان هر سطح، به‌طور معنی‌داری نسبت به سطح پایین‌تر، از تنوع باهم‌آیی بیشتری برخوردارند. یافته‌های پژوهش حاضر، در تهیه و تدوین منابع آموزشی و به‌عنوان مقیاسی برای اندازه‌گیری مهارت زبان‌دوم‌آموزان سودمند است. همچنین می‌تواند مبنایی برای شناسایی انواع گروه‌واژه‌ها در حوزه مطالعات باهم‌آیی‌های زبان دوم قرار گیرد.

کلیدواژه‌ها: باهم‌آیی، رویکرد عبارت‌شناختی، معیارهای تشخیص باهم‌آیی، پیکره زبان آموز، تنوع واژگانی.

* استناد: احمدی‌قادر، شهناز. صحرایی، رضامراد. باهم‌آیی‌ها در پیکره نوشتاری فارسی‌آموزان خارجی: مطالعه‌ای بر پایه رویکرد عبارت‌شناختی، سال چهاردهم، شماره اول (پیاپی ۲۹)، بهار و تابستان ۱۴۰۴، ۳۷-۳.

شناسه دیجیتال (DOI) 10.30479/jtpsol.2025.21222.1685

تاریخ پذیرش نهایی مقاله: ۱۴۰۳/۱۱/۰۵

تاریخ دریافت مقاله: ۱۴۰۳/۰۹/۰۴

ناشر: دانشگاه بین‌المللی امام خمینی^(ه)

۱. مقدمه

افزایش مهارت در به‌کارگیری باهم‌آیی‌ها به‌طور مستقیم بر درستی^۱ و روانی^۲ مهارت‌های تولیدی زبان‌آموزان اثر می‌گذارد (Wray, 2000). بر همین اساس می‌توان گفت؛ اهمیت باهم‌آیی‌ها در توسعه زبان میانی^۳ زبان‌آموزان نیز قابل توجه است و به آن‌ها کمک می‌کند؛ تا ساختارهای زبانی دقیق‌تر و طبیعی‌تری را بیاموزند. همچنین باهم‌آیی‌ها نقش مهمی در توسعه مهارت‌های کاربردشناختی^۴ دارند (Lewis, 2000). این بدان معناست که آشنایی زبان‌آموزان با باهم‌آیی‌های زبان، موجب می‌شود، آن‌ها بتوانند در موقعیت‌های ارتباطی مختلف، به‌صورت مؤثر و مناسب عمل کنند. مفهوم باهم‌آیی، از دهه‌های پیش توسط زبان‌شناسانی مانند پالمر (Palmer, 1933) و فرث (Firth, 1957) مطرح شده است و پیروان فرث تحقیقات گسترده‌ای درباره کاربرد آن در زبان اول انجام داده‌اند. بعدها، در رویکرد عبارت‌شناختی^۵ نیز افرادی همچون آیزنشتات (Aisenstadt, 1979)، کووی (Cowie, 1981)، ناتینگر و دکاریکو (Nattinger & DeCarrico, 1992) و هاوارث (Howarth, 1996) ترکیب‌های واژگانی را بر اساس درجه ثبات آن‌ها، با نام‌های متفاوتی طبقه‌بندی کردند و در نهایت، ترکیبات باهما را به سه دسته کلی «اصطلاحات»^۶، «باهم‌آیی‌ها»^۷ و «ترکیب‌های آزاد»^۸ تقسیم کردند. همچنین در این رویکرد، معیارهایی برای تشخیص باهم‌آیی، از دیگر انواع گروه‌های واژگانی ارائه شد. همین دقت نظر، وجه تمایز این رویکرد از سایر رویکردها (از جمله رویکرد فرث) است.

از این‌رو، پژوهش حاضر با مرور رویکردهای متفاوت، به باهم‌آیی زبان و با تمرکز بر معیارهای شناسایی الگوهای باهم‌آیی، از دیگر گروه‌واژه‌ها و همچنین با بررسی و کدگذاری چندین‌باره پیکره زبان‌آموز مورد مطالعه، ابتدا یک تعریف جامع برای باهم‌آیی‌های زبان فارسی ارائه می‌دهد و معیارهایی برای شناسایی آن‌ها در پیکره زبان‌آموز معرفی می‌کند. سپس، بر اساس تعریف نهایی به‌دست آمده و تعیین روال تشخیص باهم‌آیی‌ها، کل پیکره نمونه فارسی‌آموز، دوباره کدگذاری و پایگاه داده‌ها، استخراج و تحلیل و بررسی می‌شود. در گام بعدی، انواع الگوهای باهم‌آیی تولیدشده توسط فارسی‌آموزان، به‌ویژه الگوی نحوی «اسم-فعل» تحلیل و تفاوت عملکرد فارسی‌آموزان در سطوح زبانی مختلف، به لحاظ تنوع واژگانی باهم‌آیی‌ها بررسی می‌شود. بنابراین، پژوهش حاضر برای بررسی چگونگی عملکرد فارسی‌آموزان غیرایرانی در به‌کارگیری باهم‌آیی‌های زبان فارسی دو هدف را دنبال می‌کند: ۱. دستیابی به یک تعریف جامع و ارائه معیارهای دقیق شناسایی باهم‌آیی‌های زبان، در تولیدات نوشتاری فارسی‌آموزان غیرایرانی؛ ۲. بررسی بازنمایی

¹ accuracy

² fluency

³ interlanguage

⁴ pragmatic skills

⁵ phraseological approach

⁶ idioms

⁷ collocations

⁸ free combinations

ترکیبات باهم‌آیی و تنوع آن‌ها در تولیدات فارسی‌آموزان غیرایرانی سطوح مختلف زبانی. پرسش‌هایی که پژوهش حاضر به دنبال یافتن پاسخ آن‌ها است، به شرح زیر است:

پرسش اول: آیا تولیدات نوشتاری فارسی‌آموزان شش سطح مختلف، از لحاظ فراوانی کاربرد انواع باهم‌آیی‌های زبان فارسی تفاوت معناداری دارد؟

پرسش دوم: از لحاظ به‌کارگیری یکتا باهم‌آیی‌های پرتکرار و کم‌تکرار، آیا عملکرد فارسی‌آموزان در تولیدات نوشتاری در شش سطح زبانی مختلف، تفاوت معناداری دارد؟

متناظر با پرسش‌های بالا، فرضیه‌های پژوهش به ترتیب زیر مطرح شد:

فرضیه اول: از لحاظ فراوانی به‌کارگیری باهم‌آیی‌های زبان فارسی در تولیدات نوشتاری فارسی‌آموزان شش سطح زبانی مقدماتی، پیش‌میانی، میانی، فوق‌میانی، پیشرفته و ماهر، تفاوت معنی‌داری وجود دارد.

فرضیه دوم: فراوانی به‌کارگیری یکتا باهم‌آیی‌های پرتکرار و کم‌تکرار توسط فارسی‌آموزان شش سطح مختلف زبانی تفاوت معناداری دارد.

۲. چارچوب نظری

از میان مطالعاتی که تا کنون انجام شده است، دو رویکرد اصلی معرفی و بحث می‌شود که عبارتند از: ۱. رویکرد فرث و ۲. رویکرد عبارت‌شناختی. رویکرد عبارت‌شناختی عمدتاً بر معیارهای طبقه‌بندی باهم‌آیی‌ها متمرکز است که به‌طور خاص برای مطالعات باهم‌آیی در حوزه فراگیری زبان دوم مناسب است. از این‌رو، محوریت اصلی مطالعه حاضر، مبتنی بر رویکرد عبارت‌شناختی است. همچنین در راستای اهداف اصلی پژوهش، به مطالعات مبتنی بر بسامد، پیکره و همین‌طور تنوع واژگانی نیز اشاره می‌شود.

۲.۱. مفهوم باهم‌آیی در رویکرد فرث

در رویکرد فرث، باهم‌آیی در چارچوب روابط نحوی بین واژه‌ها؛ یعنی روابط خطی کلماتِ هم‌نشین^۱، تعریف می‌شود؛ فرث (Firth, 1957, p.173) معتقد است که معنا «مجموعه‌ای است از روابط متنی» که در سطوح مختلف زبان؛ اعم از آواشناسی، واج‌شناسی، دستور زبان، واژگان و ترکیبات واژگان و باهم‌آیی‌ها شکل می‌گیرد و هریک از این سطوح زبانی، جنبه‌ای خاص از معنا را آشکار می‌سازد. فرث بر این باور است که باهم‌آیی‌های زبان باید به‌عنوان یک سطح مستقل مورد مطالعه قرار گیرند. به گفته او، واژه‌ها در یک ترکیب واژگانی، بر اساس عادت‌های زبانی و به‌صورت خودکار در کنار هم قرار می‌گیرند و معنای هر واژه در این ترکیب، تحت تأثیر واژه‌های دیگر قرار دارد و معنای آن‌ها متقابلاً نسبت به یکدیگر پیش‌بینی و درک می‌شود (Firth, 1957, p.181). در واقع، در واژه‌های یک باهم‌آیی، پیش‌بینی

¹ leanier co-occurring

متقابل مطرح است. برای نمونه، در جمله فارسی «با دندان‌های تیزش آن را گاز گرفت»، دو واژه «دندان» و «گاز گرفتن» در انتظار وقوع یکدیگرند؛ حتی اگر کلمه «دندان‌های تیزش» حذف شود نیز تلویحاً معنای آن در واژه «گاز گرفتن» پنهان است (او آن را گاز گرفت یا گازش گرفت). با الهام از تعریف فرث، طرفداران او مطالعاتی را درباره ترکیب‌های واژگانی زبان روزمره مردم انجام دادند. سینکلر و هلیدی (Halliday & Sinclair) به همراه سایر زبان‌شناسانی که ترکیب‌ها را عمدتاً یک رابطه نحوی بین واژه‌های متن قلمداد می‌کردند، تعاریف دیگری را ارائه دادند و موجب گسترش این رویکرد شدند (Halliday, 1966; Hoey, 1991; Kjellmer, 1987, 1994; Lewis, 2000; Moon, 1998; Sinclair, 1966, 1987, 1991, 2004; Stubbs, 1996, 2001; inter alia). در مجموع، مطالعات پیروان فرث را از حیث روش‌شناختی، می‌توان به دو دسته تقسیم کرد: (۱) متن‌محور^۱، (۲) مبتنی بر بسامد^۲. سینکلر (Sinclair, 1991, p. 170) در مطالعات متن‌محور، باهم‌آیی را «ظهور دو یا چند واژه در فاصله کوتاهی از یکدیگر در یک متن» تعریف می‌کند. در مطالعات مبتنی بر بسامد نیز محققان، مفهوم باهم‌آیی را این‌گونه تعریف می‌کنند: «رابطه یک عنصر واژگانی با عناصری است که حضور آن‌ها در بافت (متن) فراتر از یک احتمال تصادفی است.» (Greenbaum, 1974; Hoey, 1991, p. 7; Moon, 1998; Stubbs, 2001; Sinclair et al., 2004). در این نوع مطالعات، باهم‌آیی‌های مهم از نظر کمی، با استفاده از فرمول‌های آماری شناسایی می‌شوند (Church & Hanks, 1990; Stubbs, 1995; Church & Hindle, 1990; Church et al., 1991; Stubbs, 1995; McEnery & Wilson, 1996; McEnery et al., 2006). همچنین پژوهش‌گران، یک استاندارد دستوری نیز به بسامد اضافه می‌کنند و باهم‌آیی را توالی تکرارشونده عناصر واژگانی تعریف می‌کنند که از نظر دستوری هم مرتبط هستند (Kjellmer, 1987, p. 133).

۲.۲. مفهوم باهم‌آیی در رویکرد عبارت‌شناختی

این رویکرد، عبارت‌های واژگانی را بر اساس درجه ثبات آن‌ها طبقه‌بندی می‌کند. محققان این رویکرد، معیارهایی برای تشخیص باهم‌آیی از دیگر انواع گروه‌های واژگانی ارائه داده و آن‌ها را در سه دسته کلی قرار داده‌اند که عبارتند از: «اصطلاحات»، «باهم‌آیی‌ها» و «ترکیب‌های آزاد» (Aisenstadt, 1979; Cowie, 1981; Nattinger & DeCarrico, 1992; Howarth, 1996). بر پایه این رویکرد، اگر این سه دسته گروه‌واژه را بر مبنای درجه ثبات واژگان هم‌نشین، بر روی یک پیوستار در نظر بگیریم، «اصطلاحات» و «ترکیب‌های آزاد» در دو طرف آن و «باهم‌آیی‌ها» در بین این دو قرار دارد. به بیان دیگر، در یک سوی این پیوستار ثابت‌ترین ترکیب‌ها (= اصطلاحات) و در سوی دیگر آن، آزادترین ترکیب‌ها (= ترکیب‌های آزاد) قرار دارند. در پژوهش حاضر نیز پیوستاری معادل همین پیوستار با معیارهای شناسایی که به‌دست آمد، مبنای نظام‌کدگذاری پیکره نمونه قرار گرفته است؛ به‌ویژه تمایز بین

¹ text-oriented study

² frequency-based study

باهم‌آیی‌ها و ترکیب‌های واژگانی آزاد؛ زیرا آنچه برای فارسی‌آموزان چالش‌برانگیزتر است، تمایز بین این دو است و این‌که زبان‌آموزان بتوانند «تشخیص دهند کاربران زبان در چه نقطه‌ای، عبارت‌ها را بر اساس قواعد نمی‌سازند و به‌صورت یک [ترکیب] کل آن‌ها را دستکاری می‌کنند» (Howarth, 1996, p. 31) حائز اهمیت است. کووی (Cowie, 1998, p. 5) اذعان داشت؛ ترکیب‌های عبارت‌شناختی همان «باهم‌آیی‌های محدود»^۱ هستند که جالب‌ترین و در عین حال دشوارترین مرزبندی را دارند؛ زیرا باهم‌آیی‌ها در ناحیه‌ای بین ترکیب‌های آزاد و اصطلاحات قرار دارند. محققان برای تمایز آن‌ها سه شاخص زیر را معرفی کرده‌اند:

الف. معیار شفافیت/تیرگی معنایی^۲: یعنی آیا می‌توان معنای کل ترکیب را از معنای عناصر تشکیل‌دهنده آن استنتاج کرد. این معیار برای تمایز اصطلاحات از دیگر انواع ترکیب‌ها بسیار مناسب است (Hausmann, 1989).

ب. معنای تخصصی^۳: یعنی دست‌کم یکی از عناصر تشکیل‌دهنده ترکیب باهم‌آیی‌ها باید معنای تخصصی داشته باشد؛ مثلاً واژه «چه» در ترکیب «چه بلنده!» برای نشان دادن تعجب گوینده از بلندی چیزی است و مفهوم مجازی^۴ دارد یا واژه «ایجاد» در «ایجاد فایل»، که به علوم رایانه‌ای مربوط می‌شود، مفهوم فنی^۵ دارد و یا واژه «دادن» در «گوش دادن» که فعل سبک در زبان فارسی است، مفهوم ناواژگانی^۶ دارد (Aisenstadt, 1979; Cowie, 1998; Howarth, 1998; Moon, 1998). متخصصان عبارت‌شناسی، «باهم‌آیی‌ها» را از «ترکیب‌های آزاد» با همین معیار متمایز می‌کنند.

پ. قابلیت جابه‌جایی^۷: در ترکیب‌های آزاد، قابلیت جابه‌جایی عناصر هم‌نشین محدود نمی‌شود؛ اما باهم‌آیی‌ها به لحاظ قابلیت جابه‌جایی عناصر تشکیل‌دهنده‌شان محدود می‌شوند (Aisenstadt, 1979; Cowie, 1992; Howarth, 1996, 1998). برای نمونه، ترکیب **تصمیم گرفتن** تعداد محدودی از افعال مانند **داشتن/ اتخاذ کردن** می‌توانند در این ترکیب، به جای **گرفتن** با واژه «تصمیم» هم‌نشین شوند.

هاوارث (Howarth, 1996, p.102) تلاش کرده است، با ارائه یک طبقه‌بندی پنج‌سطحی جامع، ابهام معیار قابلیت جابه‌جایی را از باهم‌آیی‌ها تا محدودی رفع کند. او در این طبقه‌بندی، باهم‌آیی «فعل – اسم» را از آزادترین قابلیت جابه‌جایی تا محدودترین قابلیت جابه‌جایی دسته‌بندی کرد. جدول (۱) نسخه ساده شده‌ای از طبقه‌بندی هاوارث (Howarth, 1996, p.102) از باهم‌آیی‌ها است.

¹ restricted collocation

² semantic transparency/opacity

³ specialised sense

⁴ figurative sense

⁵ technical sense

⁶ delexical sense

⁷ commutability/substitution

جدول ۱: چارچوبی برای مرزبندی باهم‌آیی‌ها، خلاصه طبقه‌بندی پنج‌سطحی هاوارث (Howarth, 1996, p.102)

نوع ترکیب	فعل	اسم	نمونه فارسی (از نگارندگان مطالعه حاضر)
ترکیب آزاد	-	-	کتاب/دفتر/میز/غذا/پول/دوست و ... می‌خواهم. کتاب می‌خواهم / (می‌خرم، می‌برم، پاره می‌کنم، دارم، ...)
باهم‌آیی	+	-	گوش داد/کردن، کمک احتیاج داشتن/خواستن، کردن
	-	+	جرم (خلاف/جنایت/دزدی ...) مرتکب شدن
	+	+	جلب توجه/اعتماد، در موضع قدرت/ضعف/بی‌طرفی
اصطلاح	++	++	چوب لای چرخ گذاشتن

(-) گستره باهم‌آیی وسیعی دارد؛ (+) مجموعه واژگانی باهم‌آیی او محدود است؛ (++) علاوه بر محدودیت هم‌نشینی‌ها، تیرگی معنایی هم دارد.
- واژه‌هایی که در ستون نمونه‌ها زیر آن‌ها خط کشیده شده است، گروه ترکیب هستند.

طبق این چارچوب، «کتاب می‌خواهم» یک ترکیب آزاد است؛ زیرا هم فعل و هم اسم، دارای گستره نامحدودی از کلمات هم‌نشینی هستند؛ اما «چوب لای چرخ گذاشتن» اصطلاحی است که علاوه بر تیرگی معنایی، عناصر آن گستره محدودی از واژه‌های هم‌نشینی دارند. همچنین «گوش دادن» باهم‌آیی است؛ علاوه بر شفافیت معنایی، دست‌کم یکی از عناصر آن دو معیار گستره باهم‌آیی محدود را دارد. جدول زیر، خلاصه تعاریف از باهم‌آیی و معیارهای مرزبندی گروه‌واژه‌ها توسط محققان این رویکرد است که از سه معیار فوق استفاده شده است. این جدول، برگرفته از منگ (Men, 2018, p.26) است که تعریف او نیز در ردیف آخر افزوده شده است.

جدول ۲: تعاریف باهم‌آیی و معیارهای مرزبندی اتخاذ شده توسط محققان رویکرد عبارت‌شناختی

معیار مرزبندی	تعریف باهم‌آیی‌ها	محقق
شفافیت معنایی؛ قابلیت جابجایی	«باهم‌آیی‌ها ترکیب دو یا چند واژه با یکدیگر هستند؛ به‌طوری‌که هریک در یکی از معانی معمول و غیراصطلاحی خود به‌کار رفته باشد؛ از الگوهای ساختاری خاص خود پیروی کند و قابلیت جابجایی آن‌ها، نه تنها به‌واسطه ظرفیت دستوری و معنایی (مانند اجزای ترکیبات به اصطلاح آزاد)، بلکه بر اساس کاربرد هم، محدود باشد.»	آیزنشتات (Aisenstadt, 1979, p. 71)
شفافیت معنایی؛ قابلیت جابجایی	«باهم‌آیی‌ها نوعی ترکیب واژگانی متشکل از دو یا چند واژه هستند که از نظر معنایی، غیراصطلاحی هستند، از الگوهای ساختاری معین پیروی می‌کنند؛ نه تنها از نظر معنایی، بلکه از نظر کاربرد نیز قابلیت جابجایی محدود دارند؛ متعلق به حوزه هم‌نشینی هستند.»	آیزنشتات (Aisenstadt, 1981, p. 54)

قابلیت جابجایی	«باهم‌آیی، پدیده زبانی است که در آن یک عنصر واژگانی معین، هم‌نشینی با واژه‌ای را به جای واژه مترادف آن ترجیح می‌دهد، به دلیل این که محدودیت‌ها در سطح نحوی یا معنایی مفهومی ^۱ نیستند؛ بلکه در سطح کاربردی هستند.»	ون روی (Van Roey, 1990, p. 46)
مفهوم تخصصی یک عنصر؛ قابلیت جابجایی (قابلیت هم‌نشینی)؛ شفافیت معنایی (به لحاظ معنایی هدفمند)	«باهم‌آیی ترکیبی است که در آن یک جزء در معنای تحت‌اللفظی ^۲ خود استفاده می‌شود؛ ولی جزء دیگر آن در معنای تخصصی ^۳ خود به کار می‌رود. معنای تخصصی یک عنصر می‌تواند مجازی ^۴ ، ناواژگانی ^۵ یا به نوعی فنی ^۶ باشد و عامل مهمی برای قابلیت هم‌نشینی ^۷ محدود در عنصر دیگر است. با این حال، این ترکیب‌ها کاملاً هدف‌مند هستند.»	هاوارث (Howarth, 1996, p. 47)
مفهوم تخصصی یک عنصر؛ قابلیت جابجایی	«باهم‌آیی ترکیبی است که در آن حداقل یک عنصر، معنای غیرتحت‌اللفظی داشته باشد و دیگری معنای تحت‌اللفظی داشته باشد و در آن‌ها قابلیت جابجایی به‌طور دلخواه محدود شده است؛ یعنی جابجایی تا حدودی امکان‌پذیر است.»	نسلهاف (Nesselhauf, 2005, p. 25)
شفافیت معنایی؛ قابلیت جابجایی	«باهم‌آیی‌ها ترکیبات واژگانی معمولی هستند که با هم‌نشینی محدود عناصر و شفافیت معنایی نسبی مشخص می‌شوند»	لاوفر و والدمن (Laufer & Waldman, 2011, p. 648)
شفافیت معنایی؛ قابلیت جابجایی	«باهم‌آیی‌ها ترکیبی از دو یا چند واژه هستند که با دامنه محدودی از هم‌نشینی، حداقل در یکی از کلمات تشکیل‌دهنده آن‌ه‌ا با شفافیت نسبی در معنا مشخص می‌شوند.»	منگ (Men, 2018, p. 30)

مطابق جدول ۲، با استفاده از معیار شفافیت معنایی می‌توان باهم‌آیی‌ها را از اصطلاحات متمایز کرد؛ اما دو معیار «مفهوم تخصصی» و «قابلیت جابه‌جایی» اگرچه به شناسایی محدودیت‌ها در باهم‌آیی‌ها کمک می‌کنند؛ اما کم و بیش در سطح شهودی عمل می‌کنند. تشخیص این که آیا مفهوم یک عنصر تخصصی است یا نه، همیشه آسان نیست و روایی جهان‌شمول و فراگیر ندارند (Hausmann, 1998, p. 170). اعمال این معیار در محدود ساختن باهم‌آیی‌ها به نوبه خود، تعداد زیادی از باهم‌آیی‌های واقعی را حذف می‌کند (Men, 2018)؛ برای نمونه، در ترکیب‌های «ارتکاب جرم» و «غذا درست کردن» هیچ‌یک از عناصر سازنده آن مفهوم تخصصی ندارند، اما به لحاظ بسامد هم‌نشینی، باهم‌آیی محسوب می‌شوند. معیار قابلیت جابه‌جایی نیز یک معیار مبهم است؛ زیرا از یک‌سو، عناصر تشکیل‌دهنده ترکیب‌های باهم‌آیی نیز می‌توانند با مجموعه نسبتاً وسیعی از هم‌نشین‌ها همراه باشد و

¹ conceptual meaning

² literal meaning

³ specialised sense

⁴ figurative

⁵ delexical

⁶ technical

⁷ collocability

محدودیت جابه‌جایی نداشته باشند. برای نمونه، واژه «تصمیم» می‌تواند با افعال مختلف مانند «تصمیم گرفتن»، «به تعویق انداختن تصمیم»، «انتقاد کردن از تصمیم»، «توضیح دادن تصمیم»، «تغییر دادن تصمیم»، «تجدید نظر کردن در تصمیم» همراه باشد (Nesselhauf, 2005, p. 27). بنابراین، هنوز هم نمی‌توان به یکی از تعاریف ارائه شده اتکا کرد و همچنان به یک تعریف روشن از باهم‌آیی نیاز داریم که در بخش بعدی در این باره بحث می‌شود.

۳.۲. تعریف باهم‌آیی در مطالعه حاضر

پژوهش حاضر، برای شناسایی و تمایزبخشی باهم‌آیی‌ها از اصطلاحات و ترکیبات آزاد در تولیدات فارسی‌آموزان از معیارهای رویکرد عبارت‌شناختی استفاده کرده است، اما همانطور که اشاره شد دو معیار «مفهوم تخصصی» و «قابلیت جابه‌جایی» همچنان ابهاماتی دارد و برای ارائه تعریفی جامع، کافی نیستند.

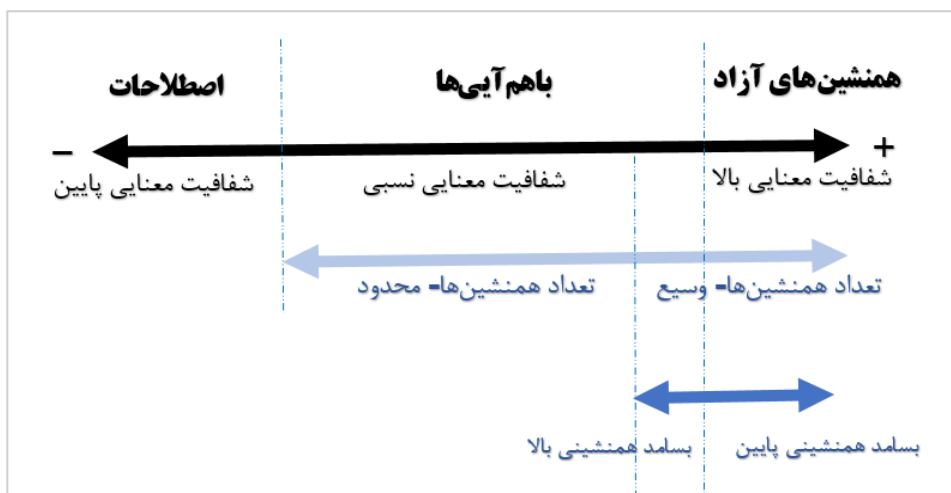
بنا به گفته کوی (Cowie, 1981) و هاوارث (Howarth, 1996)، تعریف کلی برای باهم‌آیی‌ها این است که هریک از دو عنصر یک ترکیب، دارای مجموعه محدودی از واژه‌های هم‌نشین هستند. این مجموعه محدود، تعداد واژه‌هایی که می‌توانند با یک واژه هم‌نشین باشند، هستند که «دامنه باهم‌آیی»^۱ نامیده می‌شود (Cowie, 2003; Nesselhauf, 1974, p. 20; Leech, 1981). برای نمونه، در زبان فارسی، واژه‌های «کارگر»، «رادیو»، «چرخ گوشت» و غیره باهم‌آیندهای فعل «کار کردن» هستند. این موارد و بسیاری واژه‌های دیگر که باهم‌آیندهای فعل «کار کردن» هستند، دامنه باهم‌آیی فعل «کار کردن» را تشکیل می‌دهند. بنابراین، با در نظر گرفتن «دامنه باهم‌آیی» به عنوان یک معیار دیگر برای تمایز باهم‌آیی‌ها و ترکیب‌های آزاد در رویکرد عبارت‌شناختی، می‌توان کمبودهای معیارهای پیشین را جبران کرد؛ زیرا عناصر موجود در باهم‌آیی‌ها دارای شمار محدودی از واژه‌های هم‌نشین هستند. مطابق این معیار، در نمونه «ارتکاب جرم»، «ارتکاب» دارای طیف محدودی از اسامی باهم‌آیند است؛ مانند «جنایت»، «خلافکاری»، «قتل»، «پس «ارتکاب جرم» یک باهم‌آیی تمام‌عیار است. همچنین، بر اساس این معیار، ترکیب‌های واژگانی که هر دو عنصر آن‌ها دارای دامنه وسیعی از واژه‌های هم‌نشین باشند، ترکیب‌های واژگانی آزاد محسوب می‌شوند. برای نمونه؛ ترکیب «غذا خواستن» که برای آن فعل «خواستن» می‌تواند با اسم‌های زیادی مانند «هم‌سفر»، «کتاب»، «آرامش» و غیره هم‌نشین شود و اسم «غذا» نیز می‌تواند با فعل‌های زیادی مانند «خوردن»، «خریدن»، «پختن» و غیره هم‌نشین شود، یک ترکیب آزاد است. پژوهش‌گران مطالعه حاضر، پس از بررسی تعاریف متعدد در دو رویکرد عبارت‌شناختی و فرث، و چندین بار کدگذاری ترکیب‌های مختلف برگرفته از تولیدات فارسی‌آموزان، پیوستاری را طراحی کرده‌اند که مطابق آن می‌توان گروه‌واژه‌های باهم‌آیی را با دقت از دیگر گروه‌واژه‌ها بازشناخت. همچنین، برآیند دقت عمیق در تعریف‌های

¹ collocational range

ارائه شده از ترکیب‌های باهم‌آیی و کدگذاری‌های چندین‌باره، داده‌های موجود در پیکره فارسی‌آموز، ما را به تعریف زیر از «باهم‌آیی‌ها» رساند:

«باهم‌آیی‌ها ترکیبی از دو یا چند واژه هستند که نزدیک به هم در یک متن هم‌نشین شده، شفافیت معنایی نسبی دارند و دست‌کم یکی از این دو ویژگی را دارا هستند: (۱) حداقل در یکی از عناصر تشکیل‌دهنده، گستره هم‌نشین‌ها محدود است و یا (۲) با شرط داشتن رابطه نحوی نزدیک، با بسامد بالا در تولیدات بومی‌زبانان هم‌نشین می‌شوند».

تعریف بالا نسبت به تعاریف دیگر، معیارهای روشن‌تری برای تشخیص اصطلاحات، باهم‌آیی‌ها و هم‌نشین‌های آزاد از همدیگر به دست می‌دهد. لازم به ذکر است که روش ارزیابی معیار «بسامد» بر اساس تولیدات بومی‌زبانان، با روش‌های مختلفی امکان‌پذیر است، برای نمونه، می‌توان پیکره بومی‌زبانان را بررسی کرد و یا مانند پژوهش حاضر، از شم‌زبانی چند گویشور بومی زبان کمک گرفت؛ تا بسامد ترکیب‌های مد نظر تعیین شود و پایایی آن سنجیده شود؛ انتخاب یکی از این روش‌ها برای تعیین بسامد ترکیب‌ها در روش‌شناسی پژوهش، بسته به امکانات و محدودیت‌های پژوهش‌گران انجام می‌شود. نمودار (۱) پیوستار انواع گروه‌واژه‌ها بر اساس معیارهای تشخیص باهم‌آیی است که در تعریف بالا ارائه شده است. به‌کارگیری عملی این فرایند برای تشخیص باهم‌آیی‌ها در یک پیکره نیز مستلزم رعایت سلسله مراتبی است که در نمودار فرایند تشخیص نوع گروه‌واژه‌ها در بخش ۴-۵، شکل (۲) به روشنی ترسیم شده است.



شکل (۱): پیوستار انواع گروه‌واژه‌ها بر اساس معیارهای تشخیص باهم‌آیی

بر اساس تعریف ارائه شده و پیوستار شکل (۱)، گروه‌واژه‌هایی که به لحاظ معنایی تیره هستند (مانند سردآوردن، دست از پا درازتر، دلم‌گرفت) در پیوستار شکل (۱) در سمت «اصطلاحات» قرار می‌گیرند؛ گروه‌واژه‌هایی که شفافیت معنایی نسبی داشته و هر دو عنصر در آن‌ها طیف وسیعی از واژه‌های هم‌نشین دارند. به‌عنوان «هم‌نشین‌های آزاد» طبقه‌بندی می‌شوند و در سمت راست پیوستار قرار می‌گیرند (مانند به باغ می‌روم، کار می‌خواهم)؛ و گروه واژه‌هایی که علاوه بر داشتن شفافیت معنایی نسبی، در آن‌ها یکی از عناصر و یا هر دو عنصر دارای گستره محدودی از واژه‌های هم‌نشین هستند (مانند درس/روزنامه/کتاب/داستان - خواندن و درس - خواندن/پس‌دادن/حاضرکردن/بلدبودن)، و یا ترکیباتی که با ارتباط نزدیک دستوری، بسامد بالایی در متون بومی‌زبانان دارد (مانند سم دادن، کلاس داشتن، کار داشتن) به‌عنوان «باهم‌آیی» دسته‌بندی می‌شوند. تعریف ارائه شده از باهم‌آیی و پیوستار شکل (۱) طی سلسله‌مراتبی که در بخش ۴-۵ شکل (۲) ارائه شده است، با کدگذاری کل پیکره نمونه (با حجم ۴۷۳۸۸ موردواژه^۱) آزمایش و اعتبار آن تأیید شد.

اکنون که به یک تعریف جامع از باهم‌آیی دست یافتیم، به بررسی انواع باهم‌آیی در چارچوب الگوی نحوی می‌پردازیم و دسته‌بندی انواع باهم‌آیی‌ها به لحاظ الگوی نحوی در نظام کدگذاری را مشخص می‌کنیم.

۲.۴. انواع باهم‌آیی‌ها بر اساس الگوی نحوی

در پژوهش حاضر، ترکیب‌های واژگانی طبقه‌بندی شده توسط هاسمن (Hausmann, 1989, p. 1010) و بنسون و همکاران (Benson et al., 2010) مدّ نظر است و با توجه به حوزه مطالعاتی این پژوهش با اعمال تغییرات لازم، طبقه‌بندی جدول (۳) شکل گرفت که در کدگذاری و استخراج داده‌ها مورد استفاده قرار گرفت. باهم‌آیی‌های «اسم-فعل» به صورت ترکیبات «اسم - فعل سبک/ ناواژگانی» و «اسم - فعل سنگین/واژگانی» مشخص شد. در زبان فارسی، فعل‌های فعل سبک یا ناواژگانی می‌توانند با صفت نیز هم‌نشین شوند، مانند: **ویران کردن، عزیز کردن**؛ بنابراین ترکیب «صفت - فعل سبک/ناواژگانی» نیز وارد جدول شد.

جدول (۳): طبقه‌بندی باهم‌آیی‌های معیار در مطالعه حاضر

نمونه	الگوی نحوی باهم‌آیی‌ها	
تأثیر می‌گذارد، شناسایی کرد، لطف کرد، خواهش کرد	اسم - فعل سبک	اسم-فعل
در کارخانه کار می‌کند، دانشجو تکلیف انجام دادند.	اسم (فاعل/مفعول/متمم/ بخشی از فعل مرکب) - (اسم-فعل سبک)	
• درخواست را رد کنید،	اسم (فاعل/مفعول/متمم/ بخشی از فعل مرکب) - فعل واژگانی	

¹ token

• زنبورها ویزویز می‌کنند، باران می‌بارد، باد می‌وزد		
• دیوار ساخت، کد را شناسایی کردند		
• غذا پختن، درس خواندن		
عزیز کرد، ویران ساخت	صفت-فعل سبک	
معلم محبوب، لباس بلند، سلام گرم، کتاب خواندنی	اسم- صفت	
خواب شب، جوی آب، شاخه گل	اسم-اسم	
بسیار جذاب، به شدت ترسناک، خیلی باهوش	قید- صفت	
عمیقاً تحت تأثیر قرار گرفت، شدیداً برخورد کرد	قید- فعل	

در مطالعه حاضر، علاوه بر انواع الگوهای نحوی ترکیبات باهم‌آیی، مقایسه به کارگیری یکتا باهم‌آیی‌های پرتکرار و کم‌تکرار نیز توسط فارسی‌آموزان مورد مطالعه قرار گرفت که برای تجزیه و تحلیل آن لازم است به تنوع واژگانی و معیارهای سنجش آن نیز اشاره شود.

۲.۵. تنوع واژگانی

جلیلی و صحرائی (Jalili, Sahraee, 2022) در پژوهش خود با اشاره به تعریف دالر و خو (Daller & Xue, 2007, p. 150)، مبنی بر این که «بسنندگی زبانی^۱ افراد در زبان اول و زبان خارجی، شدیداً به اندازه^۲ و عمق دایره واژگانی آنان مربوط است و غنای واژگانی نتیجه همین دانش واژگانی زیربنایی است که دربرگیرنده جنبه تنوع واژگانی (واژه‌های فعالی که در متون تولیدی نویسنده به کار رفته است)، ظرافت واژگانی (تعداد واژه‌های کم‌بسامد)، یا تراکم واژگانی (نسبت واژه‌های محتوایی^۳ و نقشی^۴) است». تنوع واژگانی را یکی از مؤلفه‌های غنای واژگانی در نظر گرفتند که اندازه‌گیری آن در متون تولیدی فارسی‌آموزان به ما کمک می‌کند تا دانش واژگانی فارسی‌آموزان را بسنجیم. سنجش تنوع واژگانی، به‌طور سنتی، با تقسیم کردن تعداد موردواژه‌ها بر یکتاواژه‌ها^۵ به دست می‌آید که به روش تی‌تی‌آر^۶ (نسبت تایپ- توکن) موسوم است. جلیلی (Jalili, 2022) پس از مقایسه معیارها و فرمول‌های مربوط به سنجش غنای واژگانی معیار «شاخص گیراد^۷» را که به‌عنوان «جذر تی‌تی‌آر» (با فرمول $G = \text{types}/\text{vtoken}$) شناخته می‌شود، به‌عنوان جایگزین روش تی‌تی‌آر انتخاب کرد. هرچه این نسبت

¹ language proficiency

² size

³ content words

⁴ function words

⁵ types

⁶ type-token ratio (TTR)

⁷ guiraud's index

بالا تر باشد، تنوع واژگانی بیشتر است. بنابراین، شمارش موردواژه‌ها و یکتاواژه‌ها، مبنای کار قرار دارد. مطابق تعریف میرازئی (Mirzaei, 2017, p. 24)، در پژوهش پیکره‌ای، هر واژه نحوی^۱ یک موردواژه است. بنابراین، موردواژه‌ها همه موجودی واژه‌ای پیکره است با کل امکانات وقوع آن‌ها و به هر تعداد که در پیکره ذکر شده است (Ibid, p. 27). مؤلفه دیگر، محاسبه شاخص تنوع واژگانی یکتاواژه است. یکتاواژه یا کلمات یکتا، همه موجودی واژه‌ای پیکره است؛ اما تنها فقط یک‌بار وقوع آن‌ها شمارش می‌شود (Ibid). در مطالعه حاضر، هر گروه‌واژه هم‌نشین که به‌عنوان باهم‌آیی در متن تشخیص داده شده‌اند، یک «موردواژه باهم‌آیی» محسوب می‌شود؛ یعنی همه موجودی باهم‌آیی‌های پیکره با کل امکانات وقوع آن‌ها و به هر تعداد که در پیکره ذکر شده است، تعداد موردواژه‌های باهم‌آیی پیکره را تشکیل می‌دهند. همچنین با شمارش یک‌بار وقوع همه صورت واژه‌های باهم‌آیی پیکره، تعداد یکتاواژه‌های باهم‌آیی پیکره به دست می‌آید که از این پس به آن‌ها با عنوان «یکتاباهم‌آیی» یا «باهم‌آیی‌های یکتا» اشاره می‌شود.

۳. پیشینه پژوهش

در دهه‌های اخیر، شاهد حجم زیادی از مطالعات در مورد یادگیری باهم‌آیی‌های زبان دوم بوده‌ایم؛ اما بیشتر داده‌ها از طریق اجرای آزمون‌های چندگزینه‌ای یا تکالیف ترجمه جمع‌آوری می‌شده که تعمیم‌پذیری نتایج آن‌ها سؤال‌برانگیز است (Howarth, 1996; Granger, 1998; Lorenz, 1999). پژوهش‌هایی که در خصوص تولیدات فارسی‌آموزان غیرایرانی انجام شده است نیز از همین روش‌ها بهره برده‌اند؛ از جمله پژوهش جزایری (Jazayeri, 2019). او در مطالعه خود تأثیر آموزش باهم‌آیی‌های واژگانی را در تقویت مهارت صحبت کردن فارسی‌آموزان سطح میانی، به کمک درس‌نامه و آزمون محقق‌ساخته ارزیابی کرده است. او با تحلیل آماری توصیفی داده‌ها نتیجه گرفت که بین نمرات آزمون باهم‌آیی شرکت‌کنندگان و نمرات آزمون صاحب‌رابطه مستقیم وجود دارد. در نهایت، او در مطالعه خود گنجاندن آموزش باهم‌آیی‌ها در برنامه درسی برای تقویت همه مهارت‌های زبانی تأکید داشته است.

پناهی (Panahi, 2002) نیز با هدف طرح و توصیف مفهوم باهم‌آیی در زبان فارسی و توجه به نقش ترکیبات باهم‌آیی در حوزه زبان‌شناسی کاربردی؛ مانند تألیف فرهنگ‌های آموزشی تحقیق کرده است. او در پژوهش خود به این نتیجه رسید که تبیین و توضیح باهم‌آیی بر اساس معناشناسی موجه‌تر است. از این رو، در حوزه‌های نظری، پس از مروری بر آرا و ملاک‌ها و طبقه‌بندی برخی زبان‌شناسان غربی مانند فرث (Firth, 1957)، پالمر (Palmer, 1933) و دیگر محققان، به الگویی در تألیف فرهنگ ترکیبات باهم‌آیی اشاره می‌کند.

¹ syntactic word

در نهایت، تعریف و توصیف معنایی و صرفی باهم‌آیی‌های زبان فارسی و همچنین پیشنهادی برای تدوین فرهنگ ترکیبات باهم‌آیند ارائه می‌کند.

مطالعه جلیلی و صحرایی (Jalili & Sahraee, 2022) نیز پژوهشی پیکره - محور با هدف سنجش تنوع واژگانی است که در آن تولیدات فارسی‌آموزان غیرایرانی چینی‌زبان و عربی‌زبان بر پایه شاخص گیراد سنجیده شده است؛ اما به کاربرد باهم‌آیی‌ها در تولیدات فارسی‌آموزان نپرداخته است.

براساس جستجوی نگارندگان، مطالعات پیکره‌محور در حوزه بررسی چگونگی به‌کارگیری باهم‌آیی‌ها که به‌طور مشخص بر روی باهم‌آیی‌های زبان فارسی در متون فارسی‌آموزان غیرایرانی تمرکز کرده باشد، مشاهده نشد. در حالی که این مطالعات بر روی تولیدات انگلیسی‌آموزان بیشتر انجام شده است؛ مانند پژوهش مینگ (Men, 2018) که مبتنی بر پیکره بوده است و به بررسی رابطه بین رشد دانش واژگانی و تولید باهم‌آیی‌ها توسط انگلیسی‌آموزان چینی پرداخت و نتیجه گرفت که افزایش دانش واژگانی با بروز خطا در به‌کارگیری باهم‌آیی‌ها هم‌بستگی مثبت دارند؛ و در حوزه‌های معنایی خاص افعال در ترکیب‌های باهم‌آیی «فعل - اسم» که توسط همه سطوح یادگیرنده استفاده می‌شوند، به‌طور فزاینده‌ای در تولیدات زبان‌آموزان سطوح بالا با خطا همراه است. همچنین، لو (Lu, 2017) با مقایسه پیکره تولیدات زبان‌دوم‌آموزان و پیکره بومی‌زبانان، دریافت که زبان‌دوم‌آموزان کمتر از بومی‌های آن زبان از باهم‌آیی‌ها استفاده می‌کنند و ساختارها و الگوهای باهم‌آیی‌های آنان با آن بومی‌های زبان، متفاوت و متأثر از فرهنگ و زبان اول آن‌ها است. بسیاری مطالعات پیکره‌محور در حوزه باهم‌آیی‌ها نیز توسط محققان غیرایرانی انجام شده است که عمدتاً پیکره انگلیسی‌زبانان را تحلیل کرده‌اند و به لحاظ روش‌شناسی، با مطالعه حاضر مشابهت دارد؛ برای نمونه، کووی (Cowie, 1992) با تحقیق بر روی متون خبری و سرمقاله‌ها در روزنامه‌ها، هاوارث (Howarth, 1996) در تحقیقات درباره متون دانشگاهی و مون (Moon 1998) با بررسی نوشته‌های عمومی در زبان انگلیسی، نتیجه گرفتند که ترکیب «فعل-اسم» بسیار پرکاربرد است.

۴. روش‌شناسی پژوهش

پژوهش حاضر، مطالعه‌ای است پیکره‌بنیان و مقطعی^۱ که طی آن، پیکره‌ای از تولیدات نوشتاری فارسی‌آموزان متعدد که در سطوح یادگیری مختلفی قرار دارند، با روش آمیخته (کمی و کیفی) تجزیه و تحلیل می‌شود. هدف اصلی پژوهش، بررسی چگونگی عملکرد فارسی‌آموزان غیرایرانی در به‌کارگیری باهم‌آیی‌های زبان فارسی است. از این‌رو، ابتدا برای دستیابی به یک تعریف جامع و ارائه معیارهای دقیق شناسایی باهم‌آیی‌ها در تولیدات نوشتاری فارسی‌آموزان غیرایرانی، آرای محققان پیشین، مطالعه و بررسی شد و با تحلیل پیکره نمونه به فرایند دقیقی برای

¹ Cross-sectional study

کدگذاری دست یافتیم؛ سپس به ترتیبی که شرح آن در این بخش ارائه می‌شود، در نهایت پایگاه داده‌های ترکیبات باهم‌آیی استخراج شد؛ تا بتوانیم تنوع واژگانی فارسی‌آموزان در به‌کارگیری باهم‌آیی‌ها را در سطوح مختلف زبانی بسنجیم.

۴.۱. جامعه آماری

داده‌های این پژوهش، از ۱۱۱۲ متن از تولیدات نوشتاری فارسی‌آموزان در پیکره زبان‌آموز بنیاد سعدی استخراج شده است که همگی تحت آموزش فارسی عمومی بوده‌اند. این تولیدات، مشتمل بر آزمون‌های بسندگی، آزمون‌های تعیین سطح و آزمون پایانی دوره‌های دانش‌افزایی بوده‌اند که همگی در شرایط آزمون‌ی مشابهی تولید شده‌اند. تولیدات نوشتاری فارسی‌آموزان در شش سطح زبانی مقدماتی، پیش‌میانی، میانی، فوق میانی، پیشرفته و ماهر است که مبنای استاندارد تعیین سطح زبان‌آموزان بر اساس کتاب «استاندارد مرجع آموزش زبان فارسی در جهان» (Sahraee & Marsous, 2016) بوده است. لازم به ذکر است که زبان‌آموزان، زبان فارسی را به‌عنوان «زبان خارجی» در محیط خارج از ایران آموخته‌اند؛ اصطلاح «زبان خارجی» در مقابل اصطلاح «زبان دوم» مطرح می‌شود. یادگیری زبان، به‌عنوان زبان خارجی، بدان معناست که زبان مقصد جزو زبان روزمره یادگیرنده نبوده و زبان‌آموز خارج از محیط گویشوران بومی، آن را آموخته است. همچنین یادگیری یک زبان به‌عنوان زبان دوم، یعنی زبان مقصد، یکی از زبان‌های رایج جامعه زبان‌آموز است و او در محیط زبانی مقصد قرار دارد.

۴.۲. ابزارهای تجزیه و تحلیل داده‌ها

- کدگذاری پیکره مورد بررسی، به‌صورت دستی و تجزیه و تحلیل و محاسبات آماری توسط پژوهش‌گران مطالعه حاضر و با استفاده از نرم‌افزارها و محاسبه‌گرهای رایانه‌ای انجام شد که عبارتند از:
- نرم‌افزار برخط «محاسبه‌گر اندازه نمونه داده‌ها»^۱ برای تعیین تعداد حجم نمونه قابل قبول؛
 - نرم‌افزار برخط «تولید اعداد تصادفی دیجیتالی»^۲، برای انتخاب تصادفی طبقه‌ای پیکره نمونه؛
 - نرم‌افزار تحلیل متن «مکس کیودی‌ای»^۳ نسخه ۲۰۲۰، برای کدگذاری، پردازش و شمارش باهم‌آیی‌ها و تولید فهرست داده‌ها به‌طور نظام‌مند و تحلیل محتوا در بخش کیفی پژوهش؛
 - نرم‌افزار تحلیل آماری اسپ‌اس‌اس^۴ نسخه ۲۷ برای انجام تحلیل‌های آماری؛
 - نرم‌افزار ماکروسافت اکسل^۵ برای ایجاد پایگاه داده‌های پژوهش.

^۱ sample size calculator (<https://goodcalculators.com/sample-size-calculator/>)

^۲ sigital random number generator (<https://stattrek.com/statistics/random-number-generator#table>)

^۳ MAXQDA2020

^۴ SPSS statistics

^۵ microsoft Excel

۳.۴. نمونه‌گیری

از آنجایی که تعداد دقیق جمعیت آماری مشخص است (۱۱۱۲ متن)، این امکان وجود دارد که به کمک فرمول‌های تعیین اندازه حجم نمونه فرمول کوکران^۱ و جدول کرجسی و مورگان^۲ حجم نمونه را تعیین کرد (Hafeznia, 2010; Habibi & Sarabadani, 2022). با استفاده از یکی از نرم‌افزارهای برخط «محاسبه‌گر اندازه نمونه داده‌ها» با سطح اطمینان^۳ ۹۵٪، اندازه نمونه ۲۸۶ متن برآورد شد. بر اساس جدول نمونه‌گیری مورگان (Morgan, 1970) نیز برای جامعه ۱۱۰۰ نفر ۲۸۵ نفر نمونه تعیین شده است. بنابراین، با در نظر گرفتن احتمال وجود «داده‌های از دست رفته»^۴ تعداد متن‌های نمونه، ۳۰۰ متن از میان کل جامعه آماری تعیین شد. برای نمونه‌گیری، از روش «نمونه‌گیری تصادفی لایه‌ای یا طبقه‌ای»^۵ استفاده شد و از هر شش سطح زبانی (شش طبقه)، داده‌ها به‌طور متوازن، با تعداد موردواژه‌های برابر به کمک نرم‌افزار «تولید اعداد تصادفی دیجیتالی»^۶، گزینش شدند. انتخاب تصادفی ساده متن‌ها برای هر شش سطح زبانی به‌طور جداگانه انجام شد. تعداد متن‌های هر طبقه، به‌گونه‌ای تعیین شد که به‌طور متوسط در هر سطح ۷۹۰۰ مورد واژه قرار گرفت (جدول ۴). در نهایت، ۳۰۰ متن (مشمول بر ۴۷۳۸۸ موردواژه) به‌طور تصادفی طبقه‌ای انتخاب شدند که متعلق به فارسی‌آموزان بزرگسال، از ۱۵ کشور اسپانیا، آلمان، اکراین، بلاروس، پرتغال، ترکیه، چین، روسیه، عراق، فرانسه، قزاقستان، کره جنوبی، لهستان، مجارستان و مکزیک بودند؛ همچنین در اطلاعات پیکره فارسی‌آموز، ملیت شماری از نویسندگان متن‌ها «نامشخص» درج شده بود. با توجه به این‌که موضوع ملیت، به‌طور مستقیم در نتیجه پژوهش حاضر، اثرگذار نیست و جزو پرسش‌های پژوهش نیست. تولیداتی که با ملیت نامشخص به‌طور تصادفی انتخاب شده بودند نیز مورد تجزیه و تحلیل قرار گرفتند.

جدول (۴): نمونه‌گیری از پیکره فارسی‌آموز به تفکیک سطوح زبانی بر اساس موردواژه‌های هر طبقه

سطح	تعداد متن‌های کل جامعه آماری	تعداد موردواژه‌های کل جامعه آماری	تعداد متن‌های نمونه	تعداد موردواژه‌های نمونه
سطح مقدماتی	۱۶۹	۱۳۸۸۰	۹۲	۷۸۵۲
سطح پیش میانی	۱۶۸	۲۰۸۰۸	۵۳	۷۹۴۹
سطح میانی	۲۷۹	۴۱۰۵۶	۴۷	۷۹۰۱
سطح فوق میانی	۲۵۰	۱۵۹۸۲	۳۷	۷۹۴۱
سطح پیشرفته	۱۶۷	۳۴۹۰۰	۳۴	۷۹۳۹
سطح فوق پیشرفته	۷۹	۱۵۹۸۲	۳۷	۷۸۰۶
کل	۱۱۱۲	۱۴۲۶۰۸	۳۰۰	۴۷۳۸۸

¹ cochrane

² krejcie and morgan

³ confidence Level (α)

⁴ missing data

⁵ stratified random sampling

⁶ digital random number generator

۴.۴. کدگذاری به وسیله نرم‌افزار مکس کیودی‌ای

پیکره فارسی‌آموزان بنیاد سعدی، برچسب‌گذاری مقوله دستوری^۱ ندارد. همچنین واژه‌های باهم‌آیی همیشه در موقعیت توالی خطی، دقیقاً در مجاورت هم و بی‌واسطه قرار ندارند؛ در نتیجه، امکان کدگذاری ماشینی وجود نداشت. بنابراین، تمام ۳۰۰ متن پیکره نمونه وارد نرم‌افزار مکس کیودی‌ای شدند و به‌طور دستی توسط پژوهشگران کدگذاری شدند. نظام کدگذاری در این نرم‌افزار بر اساس الگوهای نحوی باهم‌آیی‌ها مطابق با جدول (۳) در بخش ۲-۴ در نرم‌افزار تعریف شد. فرایند کدگذاری نیز دقیقاً مطابق با نمودار «پیوستار انواع گروه‌واژه‌ها بر اساس معیارهای تشخیص باهم‌آیی» در شکل (۱) در بخش ۲-۳ ترسیم و اجرا شد که نمودار فرایند تشخیص باهم‌آیی‌ها و اجرای کدگذاری در شکل (۲) آمده است.

فرایند کدگذاری از وقت‌گیرترین و مهم‌ترین مراحل این پژوهش بود؛ زیرا هم به‌منظور پایش تعریف و معیارهای شناسایی باهم‌آیی‌ها بود که چندین بار تکرار شد و هم طی آن، تحلیل متن پیکره صورت می‌گرفت؛ تا مبنای بخش کمی پژوهش باشد. دو نمونه از کدگذاری را در ادامه ملاحظه می‌کنید.

نمونه ۱: می‌خواهد داستان خود را برای ما حرف بزند. (کد متن: ۱۹۵۶)

حرف بزند؛ کد: باهم‌آیی/اسم- فعل سبک/ درست

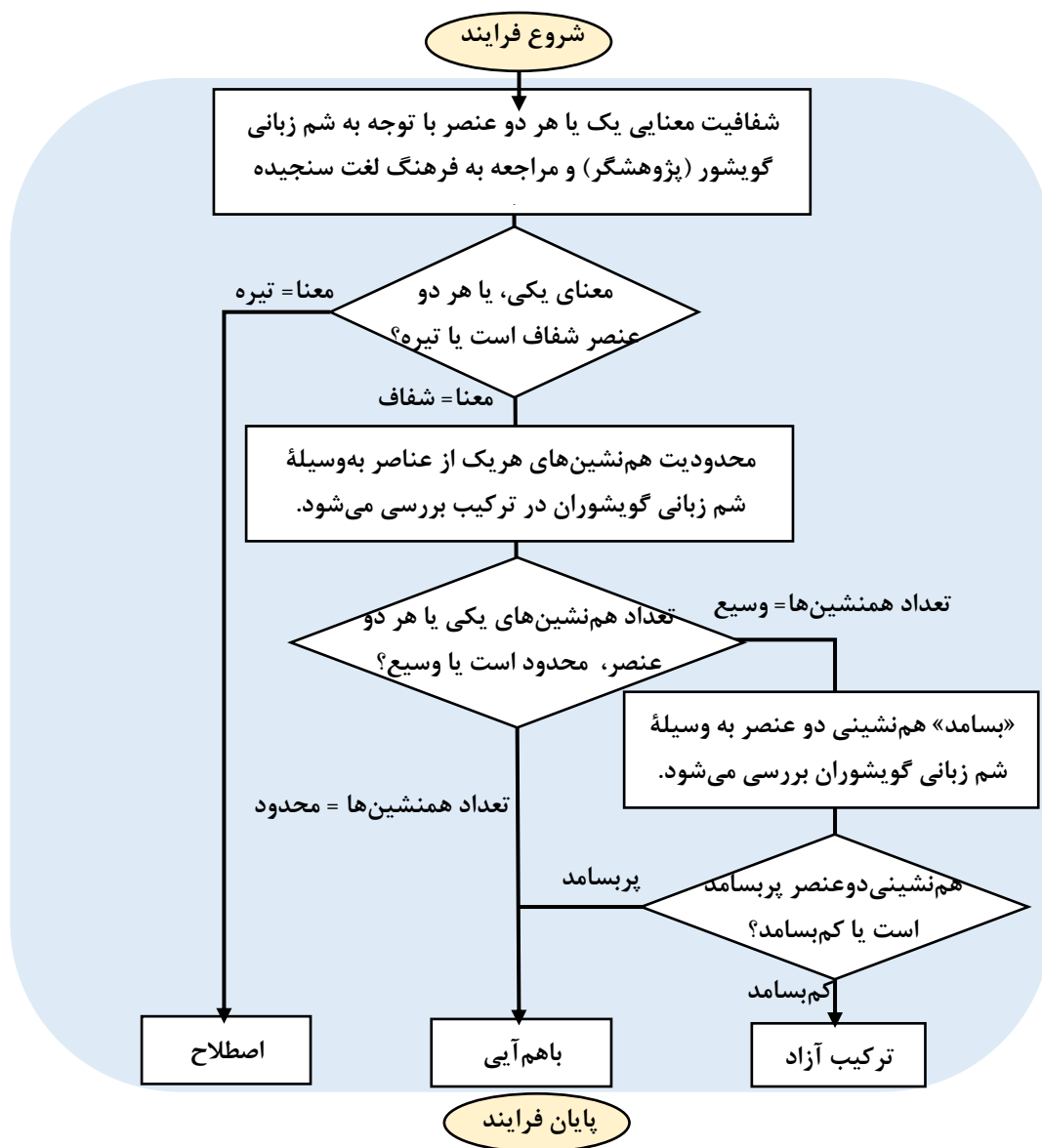
داستان - حرف بزند؛ کد: باهم‌آیی/ اسم - فعل واژگانی/ خطا در هم‌نشین (داستان «بگوید»)

نمونه ۲: بیشتر همدیگر را گوش کند. (کد متن: ۴۱۷۹)

گوش کند؛ کد: باهم‌آیی/اسم-فعل سبک/ درست

همدیگر را گوش کند؛ کد: باهم‌آیی/ اسم - (اسم-فعل سبک)/ خطا در گره («حرف» همدیگر را گوش کند)

^۱ POS= Part-of-speech tagging



شکل (۲): نمودار فرایند تشخیص باهم‌آیی‌ها، مطابق با «پیوستار انواع گروه‌واژه‌ها بر اساس معیارهای تشخیص باهم‌آیی»

راهنمای شکل: شکل بیضی (شروع و پایان فرایند) مربع (ورودی و خروجی، بررسی و انجام عملیات) لوزی (زمان تصمیم‌گیری)

۴.۵. سنجش پایایی کدگذاری

برای سنجش پایایی کدگذاری، ۲۰ درصد از پیکره مورد تحلیل به طور تصادفی انتخاب شد و در اختیار کدگذار دوم قرار گرفت. پس از اتمام کدگذاری در نرم‌افزار تحلیل متن، ضریب توافق بین دو کدگذار توسط نرم‌افزار مکس کیودی‌ای محاسبه و ضریب توافق کاپای^۱ ۹۵ صدم حاصل شد. به این ترتیب، پایایی کدگذاری پیکره مورد بررسی با اطمینان ۹۸٪ تایید می‌شود.

۴.۶. ایجاد پایگاه داده‌ها و آماده‌سازی برای تجزیه و تحلیل

داده‌ها از نرم‌افزار تحلیل محتوا در قالب فایل اکسل استخراج شد. در این فایل هر موردواژه (هر الگوی باهم‌آیی) در یک ردیف اکسل به همراه اطلاعات آن؛ از جمله اطلاعات متن و اطلاعات فارسی‌آموز تولیدکننده مانند سطح زبانی، ملیت و کد اختصاص یافته به آن، مثل نوع گروه‌واژه و درستی یا خطا بودن آن، قرار دارد. برای محاسبه فراوانی یکتاباهم‌آیی‌ها پایگاه داده‌ها به‌طور دستی پالایش شد. یکتاباهم‌آیی‌ها نرمال‌سازی^۲؛ یعنی «تبدیل متن به یک فرمت استاندارد» (Mirzaei, 2017, p. 22) و کدگذاری شدند. برای نرمال‌سازی، همه موردواژه‌های باهم‌آیی به‌صورت بن‌واژه^۳؛ یعنی «صورت انتزاعی واحد میان همه تصریفات آن واژه» (Ibid, p. 26) ثبت شدند.

۵. ارائه و واکاوی داده‌ها

در تجزیه و تحلیل پژوهش پیش‌رو، یک بُعد کمی وجود دارد که با کمک محاسبات آماری (آمار توصیفی در این بخش و آمار استنباطی در بخش بعدی) ارائه می‌شود. بُعد کیفی پژوهش نیز با ارائه استدلال‌ها بر اساس مشاهدات و نتایج حاصل مورد بحث و بررسی قرار می‌گیرد. به‌همین منظور، برای استخراج داده‌های مورد نظر پیکره نمونه، بر اساس معیارهای تمایزبخشی بین باهم‌آیی و گروه‌واژه‌های دیگر کدگذاری و داده‌ها دسته‌بندی شدند و خروجی حاصل پالایش شد. در این بخش، به تحلیل داده‌ها بر اساس جدول‌های آمار توصیفی و نمودارها می‌پردازیم.

۵.۱. آمار توصیفی باهم‌آیی‌های تولید شده توسط فارسی‌آموزان به تفکیک گروه‌های سطح

زبانی

آمار توصیفی جدول (۵) و نمودار شکل (۳) گویای آن است که در کل حجم نمونه (شامل ۴۷۳۸۸ موردواژه)، ۹۰۵۵ موردواژه باهم‌آیی مشاهده شده است. فراوانی موردواژه‌های باهم‌آیی به کار رفته در تولیدات نوشتاری گروه‌های

^۱ kappa coefficient

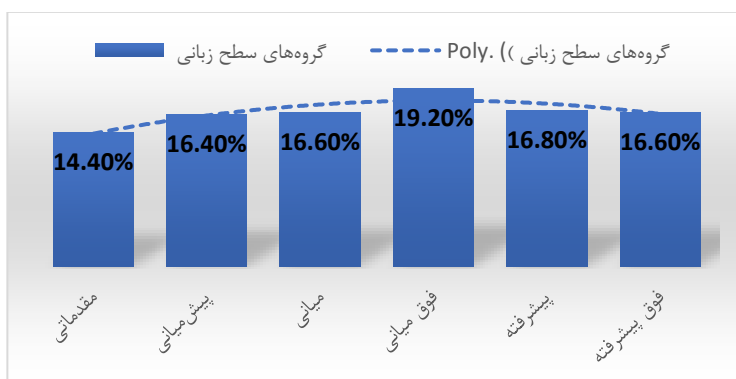
^۲ normalization

^۳ lemma

مختلف سطح زبانی، از سطح مقدماتی (۱۴٪) تا سطح فوق پیشرفته (۱۶٪) افزایش تدریجی ۲ درصدی را نشان می‌دهد.

جدول (۵): فراوانی کل باهم‌آیی‌های تولید شده توسط فارسی‌آموزان به تفکیک گروه‌های سطح زبانی

درصد	فراوانی موردواژه‌ها	
۱۴،۴۰٪	۱۳۰۱	مقدماتی
۱۶،۴۰٪	۱۴۸۷	پیش‌میانی
۱۶،۶۰٪	۱۵۰۱	میانی
۱۹،۲۰٪	۱۷۳۹	فوق میانی
۱۶،۸۰٪	۱۵۲۱	پیشرفته
۱۶،۶۰٪	۱۵۰۷	فوق پیشرفته
۱۰۰٪	۹۰۵۵	کل

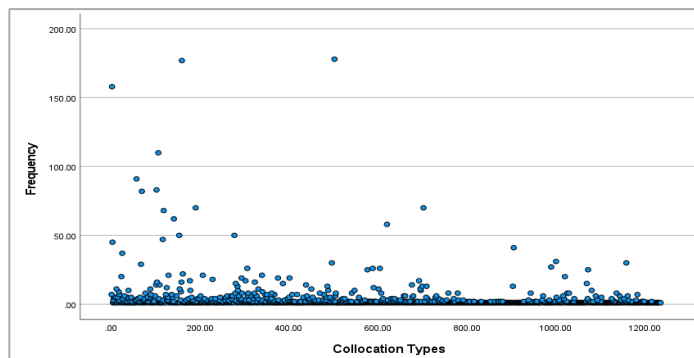


شکل (۳): نمودار فراوانی کل باهم‌آیی‌های تولید شده توسط فارسی‌آموزان به تفکیک گروه‌های سطح زبانی

۵.۲. آمار توصیفی موردواژه‌های باهم‌آیی و یکتا باهم‌آیی‌های الگوی نحوی «اسم-فعل»

تحلیل این بخش از مطالعه، مستلزم ارائه آمار توصیفی دقیق‌تر برحسب موردواژه‌های باهم‌آیی و یکتا باهم‌آیی‌ها بود که با توجه به مشاهدات اولیه پایگاه داده‌ها، بر پربسامدترین الگوی نحوی؛ یعنی الگوی نحوی «اسم-فعل» تمرکز شد. بنابراین، پس از مرتب‌سازی داده‌ها و پالایش مجدد آن‌ها، حذف مقادیر مبهم، مشاهده شد که ۴۳۸۸ موردواژه باهم‌آیی و ۱۲۳۲ یکتا باهم‌آیی الگوی نحوی «اسم-فعل» در کل پیکره مورد بررسی، به کار رفته است. این میزان فراوانی یکتا باهم‌آیی (۱۲۳۲)، به نسبت موردواژه‌های باهم‌آیی (۴۳۸۸)، بدان معناست که به‌طور متوسط هر یک از باهم‌آیی‌های یکتا در حدود ۳ یا ۴ بار در متون مورد بررسی تکرار شده است؛ اما نمودار شکل (۴) گویای آن است که پراکنش یکتا باهم‌آیی‌ها در بخش فراوانی ۱ تا ۵ بار (پایین نمودار) بسیار متراکم‌تر از بالای

نمودار است. به عبارت دیگر، شمار اندکی از یکتاباهم‌آیی‌های پرتکرار و در مقابل، شمار بسیار زیادی یکتاباهم‌آیی‌های کم‌تکرار (با فراوانی کمتر از پنج بار) در پیکره نمونه مشاهده شده است.



شکل ۴: نمودار پراکندگی فراوانی یکتاباهم‌آیی‌های الگوی «اسم-فعل» در تولیدات فارسی‌آموزان

با مراجعه به پایگاه داده‌ها نیز مشاهده می‌شود که ۸۱۶ یکتاباهم‌آیی فقط یک‌بار توسط کل زبان‌آموزان به کار برده شده است؛ ۱۶۰ یکتاباهم‌آیی، فقط دو موردواژه؛ ۶۲ یکتاباهم‌آیی، سه موردواژه؛ ۴۰ یکتاباهم‌آیی، چهار موردواژه؛ و از ۲۷ یکتاباهم‌آیی، پنج موردواژه در کل پایگاه داده‌ها مشاهده شده است. از سوی دیگر، شماری از یکتاباهم‌آیی‌ها هستند که با تعداد موردواژه‌های بسیار بالا (بین ۱۰ تا ۱۷۸ موردواژه از هریک) به کار رفته است. با مشاهده پراکنش یکتاباهم‌آیی‌ها بر حسب فراوانی آن‌ها در نمودار شکل ۵ و با توجه به آمار توصیفی که ارائه شد، برای تفسیر این عدم توازن شدید، کل یکتاباهم‌آیی‌ها بر حسب فراوانی آن‌ها؛ یعنی تعداد موردواژه‌هایشان به سه دسته تقسیم شدند: «یکتاواژه‌های کم‌تکرار» که پنج بار یا کمتر (≤ 5) به کار رفته‌اند، «یکتاباهم‌آیی‌های پرتکرار» که بیش از ده بار (≥ 10) استفاده شده‌اند و یکتاباهم‌آیی‌هایی که بین پنج و ده بار ($5-10$) به کار رفته‌اند، به عنوان مواردی که در حاشیه دو گروه مشاهده شده‌اند، در نظر گرفته شد. مینگ (Meng, 2018) نیز در پژوهش خود به همین عدم توازن در پراکندگی فراوانی مواجه شده و با تقسیم‌بندی مشابه داده‌های خود را تحلیل کرد. لازم به ذکر است که بنا به تعریف بایبر و همکاران (Biber, et al., 1999)، ترکیبی باهم‌آیی محسوب می‌شد که دست‌کم ده بار در هریک میلیون «موردواژه» تکرار شده باشد و به طور هم‌زمان دست‌کم در پنج متن مختلف ثبت شده باشد. با وجود این، در مطالعه حاضر یکتاباهم‌آیی‌هایی که فقط یک‌بار در کل پیکره مورد بررسی مشاهده شده بودند (۸۱۶ موردواژه) به عنوان داده‌های از دست رفته از پایگاه داده‌ها حذف نشدند و در فرایند تجزیه و تحلیل قرار گرفتند؛ زیرا، اولاً، پیکره مورد بررسی تولیدات فارسی‌آموزان خارجی است و با پیکره بومی‌زبانان فرق دارد؛ یعنی در تعریف تشخیص باهم‌آیی‌ها در تولیدات فارسی‌آموزان، فقط یکی از معیارها «بسامد» است و آن هم در صورتی باهم‌آیی در نظر گرفته می‌شود که آن ترکیب در متون فارسی‌زبانان پربسامد بوده باشد؛ در این

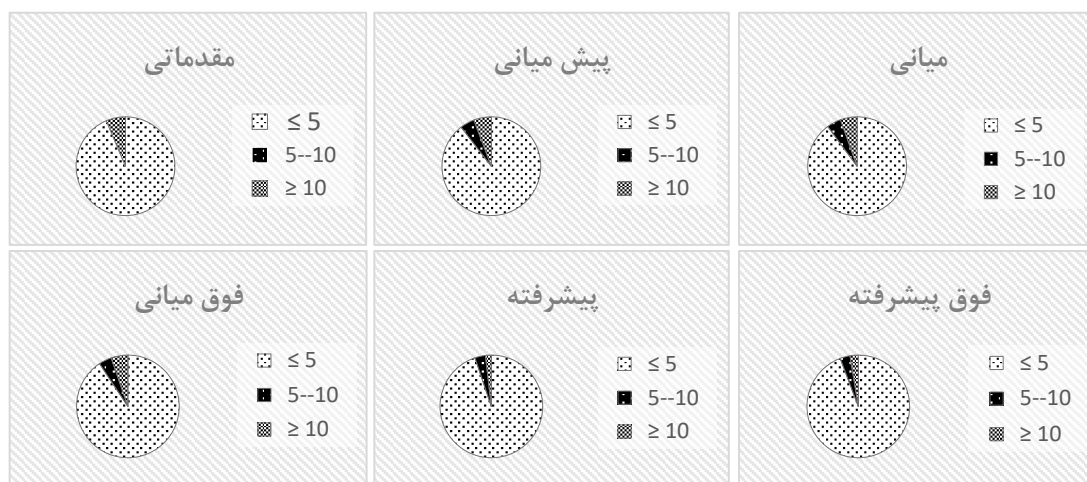
پژوهش پربسامد بودن ترکیب مورد نظر در متون فارسی‌زبانان، با شم زبانی گویشوران بومی و کدگذاران سنجیده شد که در فرایند ترسیم شده (نمودار ۲) در چندین مرحله، قبل از تصمیم‌گیری و تعیین نوع ترکیب به ارزیابی گویشوران زبان بومی اشاره شده است؛ دوماً، فرایند تشخیص باهم‌آیی‌ها با دقت بسیار زیادی انجام شده و این تعداد خطا در کدگذاری ممکن نیست.

تا این مرحله از پژوهش، باهم‌آیی‌های درست و خطا در آمار توصیفی، مورد بررسی قرار گرفتند؛ تا زبان میانی و فرایند رشد واژگانی فارسی‌آموزان نیز در نظر گرفته شود. در ادامه، یکتاباهم‌آیی‌های خوش‌ساخت استخراج شدند؛ تا قضاوت رشد تنوع واژگانی در به‌کارگیری باهم‌آیی‌ها به‌درستی صورت گیرد. در همین راستا، آمار توصیفی جدول (۶) به تفکیک باهم‌آیی‌های پرتکرار و کم‌تکرار در شش سطح زبانی ارائه شده است. با مقایسه درصد یکتاباهم‌آیی‌ها که در دو گروه کم‌تکرار و پرتکرار، فقط توسط سطح مقدماتی تولید شده است، درمی‌یابیم که زبان‌آموزان سطح مقدماتی درصد زیادی (۹۳،۶۴٪) یکتاباهم‌آیی کم‌تکرار تولید کرده‌اند و درصد ناچیزی (۶،۹۴٪) یکتاباهم‌آیی پرتکرار تولید کرده‌اند.

جدول ۶: آمار توصیفی مقایسه درصد به‌کارگیری یکتاباهم‌آیی‌ها در دو گروه پرتکرار و کم‌تکرار برای هر سطح زبانی

درصد	مجموع	پرتکرار		۵-۱۰		کم تکرار		دسته‌بندی موردواژه‌ها
		≤۱۰	فرآوانی	درصد	فرآوانی	≤۵	فرآوانی	(Tokens) فرآوانی نوع باهم‌آیی‌ها در گروه‌های سطح زبانی (Types)
٪۱۰۰	۱۷۳	٪۶.۳۵	۱۱	٪۰.۰۰	۰	٪۹۱.۶۱	۱۶۲	مقدماتی
٪۱۰۰	۲۰۱	٪۵.۴۷	۱۱	٪۴.۴۸	۹	٪۹۰.۰۵	۱۸۱	پیش‌میانی
٪۱۰۰	۲۳۷	٪۵.۴۸	۱۳	٪۴.۶۴	۱۱	٪۸۹.۸۷	۲۱۳	میانی
٪۱۰۰	۳۲۲	٪۵.۲۷	۱۷	٪۴.۰۴	۱۳	٪۹۰.۶۸	۲۹۲	فوق میانی
٪۱۰۰	۲۷۱	٪۱.۸۴	۵	٪۳.۳۲	۹	٪۹۴.۸۳	۲۵۷	پیشرفته
٪۱۰۰	۳۴۹	٪۲.۵۷	۹	٪۳.۱۵	۱۱	٪۹۴.۲۷	۳۲۹	ماهر/فوق پیشرفته

همین اطلاعات آماری در سطوح زبانی دیگر نیز قابل مشاهده است و هرچه به سمت سطح ماهر می‌رویم، فاصله فرآوانی به‌کارگیری یکتاباهم‌آیی‌های کم‌تکرار با پرتکرار بیشتر می‌شود؛ به‌طوری که در سطح فوق پیشرفته ۹۴٪ کم‌تکرار در مقابل ۲٪ پرتکرار مشاهده شده است. نمودارهای شکل (۵) نیز این فاصله را به خوبی نمایش می‌دهند.



شکل ۵: نمودار مقایسه درصد به کارگیری یکتاباهم‌آیی‌ها در دو گروه پرتکرار و کم‌تکرار برای هر سطح زبانی

در مقایسه بین سطوح زبانی به کارگیری یکتاباهم‌آیی‌های پرتکرار (≥ 10) توسط فارسی‌آموزان (ن.ک. جدول ۶، خط‌چین‌های کم‌رنگ‌تر) می‌توان گفت؛ به‌طور کلی در هر شش سطح، فارسی‌آموزان تعداد محدودی از باهم‌آیی‌های پرتکرار استفاده کرده‌اند؛ اما زبان‌آموزان سطوح مقدماتی، پیش میانی و میانی و فوق میانی (۶٪ و ۵٪) بیش از زبان‌آموزان ماهر (۱٪ و ۲٪) از یکتاباهم‌آیی‌های پرتکرار استفاده کرده‌اند. در مقابل، ستون یکتاباهم‌آیی‌های کم‌تکرار (≤ 5)، نشان می‌دهد که شمار قابل توجهی یکتاباهم‌آیی کم‌تکرار در تولیدات نوشتاری فارسی‌آموزان مشاهده شده است. همچنین درصد یکتاباهم‌آیی‌های کم‌تکرار، از سطح مقدماتی تا ماهر، به تدریج افزایش می‌یابد؛ در سطح ماهر یکتاباهم‌آیی‌های کم‌تکرار (۲۲،۹۴٪) نسبت به سطح مقدماتی (۱۱،۳۰٪) افزایش قابل قبولی داشته است.

۶. تفسیر داده‌ها و بررسی فرضیه‌های پژوهش

مشاهدات و نتایج به‌دست آمده از پیکره نمونه که در بخش پیشین ارائه شد، به کمک آمار استنباطی حاصل از محاسباتی که با نرم‌افزار SPSS انجام شد، مورد بررسی قرار گرفت و فرضیه‌های پژوهش، یک‌به‌یک راستی‌آزمایی شد؛ تا بتوان مشاهدات و نتایج به دست آمده از پیکره نمونه را به جامعه آماری مورد مطالعه تعمیم داد که گزارش آن در ادامه ارائه می‌شود.

۶.۱. بررسی فرضیه اول

فرضیه نخست مطالعه حاضر این بود که «از لحاظ فراوانی به کارگیری باهم‌آیی‌های زبان فارسی، در تولیدات نوشتاری فارسی‌آموزان شش گروه سطح زبانی مقدماتی، پیش میانی، میانی، فوق میانی، پیشرفته و ماهر، تفاوت

معنی‌داری وجود دارد» که تفسیر داده‌ها بر اساس آمار توصیفی در بخش ۵ مورد بررسی قرار گرفت. برای راستی‌آزمایی آن لازم است ابتدا تعیین کرد، توزیع داده‌ها نرمال بوده است یا خیر، برای تعیین نرمال بودن توزیع داده‌ها، از آماره شاپیرو-ویلک^۱ استفاده شد که نتیجه آن نشان می‌دهد ($sig = 0.001 < 0.05$) توزیع داده‌های پژوهش، نرمال نیست. بنابراین، متغیر کاربرد باهم‌آیی‌ها در تولیدات نوشتاری فارسی‌آموزان شش گروه مستقل، به‌وسیله آزمون ناپارامتریک کروسکال-والیس^۲ مقایسه شد؛ نتیجه آن ($sig. = 0.416 > 0.05$) گویای آن است که تفاوت میزان کاربرد باهم‌آیی‌ها توسط فارسی‌آموزان شش سطح زبانی مقدماتی تا ماهر معنی‌دار نیست و فرضیه رد می‌شود؛ تفاوت فراوانی به‌کارگیری باهم‌آیی‌ها در سطوح زبانی مختلف، چندان قابل توجه نیست و تقریباً در تمام سطوح زبانی به یک میزان از باهم‌آیی‌ها استفاده می‌کنند. می‌توان نتیجه گرفت، باهم‌آیی‌های زبان از همان ابتدا نقش مؤثری در تولیدات نوشتاری زبان‌آموزان دارد.

۲.۶. بررسی فرضیه دوم

فرضیه دوم، ناظر بر مقایسه فراوانی یکتا باهم‌آیی‌های پرتکرار و کم‌تکرار در تولیدات فارسی‌آموزان سطوح مختلف زبانی است که بدین ترتیب عنوان شده بود: «فراوانی به‌کارگیری یکتا باهم‌آیی‌های پرتکرار و کم‌تکرار در تولیدات فارسی‌آموزان شش سطح مختلف زبانی تفاوت معناداری دارد». برای راستی‌آزمایی فرضیه دوم پژوهش، نرمال بودن توزیع فراوانی باهم‌آیی‌های الگوی نحوی «اسم-فعل» با استفاده از آزمون شاپیرو-ویلک ($sig = 0.369 > 0.05$) محاسبه شد و نشان داد؛ توزیع داده‌ها نرمال است؛ بنابراین، برای سنجش معنی‌داری تفاوت، از آزمون پارامتریک تحلیل واریانس^۳ استفاده شد. نتیجه آزمون تحلیل واریانس ($F = 39.08, sig. = 0.000 < 0.05$) با ۹۹ درصد اطمینان، تفاوت بین یکتا باهم‌آیی‌های پرتکرار در سطوح زبانی مختلف را معنی‌دار نشان داد. آزمون تعقیبی توکی^۴ با مقایسه جفتی گروه‌های سطوح زبانی نیز نشان داد که بین سطوح مقدماتی، پیش‌میان و میانی، هیچ تفاوت معنی‌داری وجود ندارد و محل این معنی‌داری در بین چهار سطح اول (مقدماتی، پیش‌میان، میانی و فوق‌میان) با سطوح پیشرفته و ماهر است ($sig.$ کوچکتر از ۰.۰۵)؛ یعنی فارسی‌آموزان با سطوح زبانی پایین‌تر، به‌طور معنی‌داری بیشتر از سطح پیشرفته و ماهر از یکتا باهم‌آیی‌های پرتکرار استفاده کرده‌اند.

نتیجه آزمون تحلیل واریانس ($F = 4335.28, sig. = 0.000 < 0.05$) با ۹۹ درصد اطمینان، تفاوت بین یکتا باهم‌آیی‌های کم‌تکرار (≤ 5) در سطوح زبانی مختلف را نیز معنی‌دار نشان داد. آزمون تعقیبی توکی با مقایسه جفتی گروه‌های سطوح زبانی نشان داد که این معناداری در بین تمام سطوح زبانی وجود دارد؛ یعنی فارسی‌آموزان

^۱ shapiro-wilk

^۲ kruskal-wallis test

^۳ ANOVA

^۴ tukey

هر سطح زبانی نسبت به فارسی‌آموزان سطح پایین‌تر از خود، از یکتا‌باهم‌های کم‌تکرار بیشتری استفاده می‌کنند که این به افزایش تنوع واژگانی آن‌ها مربوط است. برای تأیید این یافته محاسبه تنوع باهم‌آیی‌های فارسی‌آموزان از معیار «شاخص گیراد» استفاده شد که در جدول (۷) نشان می‌دهد تنوع باهم‌آیی‌های به کار رفته در هر گروه سطح زبانی از سطوح مقدماتی تا ماهر به تدریج افزایش پیدا کرده است.

جدول ۷: مقایسه تنوع باهم‌آیی در تولیدات فارسی‌آموزان سطوح مختلف زبانی بر حسب شاخص گیراد

سطح زبانی	مقدماتی	پیش‌میانی	میانی	فوق میانی	پیشرفته	فوق پیشرفته
یکتا‌باهم‌آیی‌های Types	۱۷۳	۲۰۱	۲۳۷	۳۲۲	۲۷۱	۳۴۹
موردواژه‌های باهم‌آیی Tokens	۴۲۴	۵۸۹	۶۱۹	۸۱۵	۶۱۵	۶۳۲
تنوع باهم‌آیی‌های «سم-فعل» بر حسب شاخص گیراد	۸،۴۰	۸،۲۸	۹،۵۲	۱۱،۲۷	۱۰،۹۲	۱۳،۸۸

۶. نتیجه‌گیری و پیشنهادهای آموزشی و پژوهشی

در مرحله اول پژوهش، به یک تعریف روشن و معیارهای تمایزبخشی بین باهم‌آیی و گروه‌واژه‌های دیگر در تولیدات فارسی‌آموزان غیرایرانی دست یافتیم که منجر به ترسیم پیوستار و نموداری (ر.ک. شکل ۱ و ۲) برای فرایند تشخیص باهم‌آیی‌ها در متون پیکره نمونه شد. پیکره مورد بررسی بر اساس آن کدگذاری و داده‌ها دسته‌بندی شدند و خروجی حاصل پالایش شد. در بخش پژوهش کمی با تجزیه و تحلیل داده‌ها یافته‌هایی به دست آمد که در ادامه بحث و جمع‌بندی می‌شود.

۱. در پژوهش حاضر، مطابق تعریف ارائه شده (ر.ک. بخش ۲-۴)، معیارهای تمایزبخشی باهم‌آیی‌ها از اصطلاحات و ترکیبات آزاد در متون نوشتاری فارسی‌آموزان خارجی به ترتیب زیر مشخص شد.

- وجه تمایز باهم‌آیی‌ها و اصطلاحات: باهم‌آیی‌ها شفافیت معنایی نسبی دارند.

- وجه تمایز باهم‌آیی‌ها با ترکیبات آزاد، در وهله اول: باهم‌آیی‌ها گستره همنشین‌های محدودی دارند؛ در وهله دوم: باهم‌آیی‌ها در متون بومی‌زبانان پربسامد هستند.

پناهی (Panahi, 2002) نیز در مطالعه خود تعریف باهم‌آیی‌ها و ملاک‌های شناسایی آن‌ها را بر اساس رویکرد معناشناسی ارائه کرده است؛ اما ملاک‌های بازشناسی باهم‌آیی‌ها از دیگر ترکیب‌ها جزو اهداف پژوهش او نبوده است و به روشنی پژوهش حاضر نیاورده است. همچنین در بخشی از مطالعه خود به ملاک‌های طبقه‌بندی انواع باهم‌آیی‌ها بر اساس الگوهای نحوی پرداخته است و آن‌ها را در دسته‌بندی کلی‌تر به ترکیبات فعلی و غیرفعلی تقسیم کرده است که با مطالعه حاضر مشابهت دارد.

۲. با رد فرضیه اول، مبنی بر متفاوت بودن فراوانی باهم‌آیی‌ها، در متون نوشتاری سطوح مختلف زبانی با یکدیگر می‌توان نتیجه گرفت، فارسی‌آموزان خارجی تقریباً در تمام سطوح زبانی به یک میزان از باهم‌آیی‌ها استفاده

می‌کنند؛ یعنی باهم‌آیی‌های زبان از همان ابتدا نقش مؤثری در تولیدات نوشتاری زبان‌آموزان دارد و این نکته باید در تهیه و تدوین مواد آموزشی از همان شروع دوره فارسی‌آموزی مورد توجه ویژه قرار گیرد.

۳. از میان ۱۰ الگوی نحوی که کدگذاران در تولیدات فارسی‌آموزان (شامل ۴۷۳۸۸ موردواژه) جستجو کردند، حدود نیمی از باهم‌آیی‌ها (۱۲۳۲ مورد باهم‌آیی) الگوی نحوی «اسم-فعل» بود و در رتبه‌های بعدی الگوهای نحوی «اسم-اسم» و «اسم-صفت» قرار داشتند که همسو با بخشی از مطالعات مینگ (Men 2018) و لو (Lu, 2017) است.

۴. همان‌طور که پیش‌بینی می‌شد، یکتاباهم‌آیی‌های پرتکرار توسط فارسی‌آموزان سطوح مقدماتی تا فوق‌میان‌بیشتر از سطوح پیشرفته و ماهر به‌کار رفته بودند. بنابراین، شناسایی یکتاباهم‌آیی‌های پرتکرار، به‌منظور بهره‌برداری در متون آموزشی حائز اهمیت است. از این‌رو، پژوهش‌گران با مشاهده پایگاه داده‌ها فهرست یکتاباهم‌آیی‌های «اسم - فعل» که بیش از ۲۰ بار در پیکره مورد بررسی مشاهده شده است را استخراج و در جدول زیر معرفی کردند. همان‌طور که در جدول زیر مشاهده می‌کنید، در پرتکرارترین ترکیبات باهم‌آیی «اسم-فعل»، حدوداً ۸۰٪ (۲۴ یکتاباهم‌آیی از میان ۳۴ یکتاباهم‌آیی)، از فعل سبک استفاده شده است.

جدول ۸: فهرست پرتکرار باهم‌آیی‌های یکتا از الگوی «اسم-فعل» که بیست بار یا بیشتر تکرار شده‌اند

یکتاباهم‌آیی	فراوانی	الگوی نحوی	یکتاباهم‌آیی	فراوانی	الگوی نحوی
۱. دوست داشتن (کد ۵۰۲)	۱۷۸	اسم-فعل سبک	۱۸. احتیاج داشتن (کد ۲۵)	۳۷	اسم-فعل واژگانی
۲. زندگی کردن (کد ۱۵۹)	۱۷۷	اسم-فعل سبک	۱۹. کار-انجام دادن (کد ۱۰۰۰)	۳۱	اسم- (اسم-فعل سبک)
۳. فکر کردن (کد ۲)	۱۵۸	اسم-فعل سبک	۲۰. درست کردن (کد ۴۹۶)	۳۰	اسم-فعل سبک
۴. درس خواندن (کد ۱۰۶)	۱۱۰	اسم-فعل واژگانی	۲۱. نیاز داشتن (کد ۱۱۵۸)	۳۰	اسم-فعل واژگانی
۵. غذا خوردن (کد ۵۷)	۹۱	اسم-فعل واژگانی	۲۲. استراحت کردن (کد ۶۷)	۲۹	اسم-فعل سبک
۶. وجود داشتن (کد ۱۰۲)	۸۳	اسم-فعل واژگانی	۲۳. غذا- درست کردن (کد ۹۸۹)	۲۷	اسم- (اسم-فعل سبک)
۷. استفاده کردن (کد ۶۹)	۸۲	اسم-فعل سبک	۲۴. تشکیل شدن (کد ۳۰۶)	۲۶	اسم-فعل سبک
۸. یادگرفتن (کد ۱۹۰)	۷۰	اسم-فعل سبک	۲۵. سعی کردن (کد ۵۸۸)	۲۶	اسم-فعل سبک
۹. کمک کردن (کد ۷۰۲)	۷۰	اسم-فعل سبک	۲۶. شروع شدن (کد ۶۰۴)	۲۶	اسم-فعل سبک
۱۰. انجام دادن (کد ۱۱۸)	۶۸	اسم-فعل سبک	۲۷. وقت داشتن (کد ۵۷۶)	۲۵	اسم-فعل واژگانی
۱۱. کار کردن (کد ۱۴۱)	۶۲	اسم-فعل سبک	۲۸. مشکل داشتن (کد ۱۰۷۲)	۲۵	اسم-فعل واژگانی

۱۲. صحبت کردن (کد ۶۲۰)	۵۸	اسم-فعل سبک	۲۹. غذا پختن (کد ۱۶۱)	۲۲	اسم-فعل واژگانی
۱۳. بازی کردن (کد ۱۵۳)	۵۰	اسم-فعل سبک	۳۰. اوقات را گذراندن (کد ۱۲۹)	۲۱	اسم-فعل واژگانی
۱۴. پیدا کردن (کد ۲۷۷)	۵۰	اسم-فعل سبک	۳۱. علاقه داشتن (کد ۲۰۶)	۲۱	اسم-فعل واژگانی
۱۵. انتخاب کردن (کد ۱۱۶)	۴۷	اسم-فعل سبک	۳۲. تماشا کردن (کد ۳۳۹)	۲۱	اسم-فعل سبک
۱۶. کتاب خواندن (کد ۳)	۴۵	اسم-فعل واژگانی	۳۳. احترام گذاشتن (کد ۲۳)	۲۰	اسم-فعل سبک
۱۷. خوابگاه- زندگی کردن (کد ۹۰۵)	۴۱	اسم- (اسم-فعل سبک)	۳۴. گوش دادن (کد ۱۰۲۰)	۲۰	اسم-فعل سبک

۱. فارسی‌آموزان ماهرتر باهم‌آیی‌های متنوع‌تری نسبت به فارسی‌آموزان با سطح مهارت کمتر تولید می‌کنند. نتیجه به‌دست آمده با یافته‌های گیتساک (Gitsaki, 1999) و ژانگ (Zhang, 1993) و مینگ (Men 2018) مطابقت دارد. این یافته طبیعی است؛ زیرا هرچه زبان‌آموزان بیشتر آموزش ببینند، ترکیب‌های بیشتری می‌آموزند؛ اما آنچه قابل توجه است، این است که در مقایسه جفتی سطوح زبانی در نمونه‌پیکره مورد بررسی، تنوع واژگانی هر سطح به‌طور معناداری از سطح پایین‌تر خود بیشتر بود. بر اساس این یافته می‌توان گفت؛ پس از مطالعات دقیق‌تر، تسلط بر باهم‌آیی‌ها می‌تواند در حوزه سنجش تعیین سطح سودمند باشد؛ البته تسلط بر باهم‌آیی‌های زبان به‌صورت روانی و دقت در گفتار نمود پیدا می‌کند و با این دو شاخص در آزمون‌های گفتار سنجیده می‌شود، اما معیاری برای تعیین سطح مهارت فارسی‌آموزان در نوشتار نیز می‌تواند باشد.

۲. مدرسان زبان فارسی همواره به تولید و اجرای فعالیت‌های تقویت واژه و باهم‌آیی‌ها توجه داشته‌اند؛ اما بی‌گمان در ذهن آن‌ها این پرسش‌ها گذشته‌است: آیا ضرورتی دارد زبان‌آموزان تشخیص دهنده عناصر همیشه نمی‌توانند آزادانه ترکیب شوند؟ آیا آموزش ترکیبات باهم‌آیی به روش مستقیم ضروری است؟ یعنی آیا باید مطابق رویکرد رفتارگرایی^۱ در روش تدریس گفتاری - شنیداری^۲ در کلاس با تمرین و تکرار خوشه‌واژه‌ها به آموزش باهم‌آیی‌ها بپردازیم؟ برای پاسخ‌گویی به این پرسش‌ها، با نگاهی کلی به آموزش واژه، می‌توان گفت محققان براین باورند که خواندن گسترده برای وسعت دادن به دانش واژگانی بسیار کارساز است و ذهن زبان‌آموزان هنگام خواندن، واژه‌ها را صرفاً با هدف درک مطلب، برای فراگیری انتخاب می‌کند؛ بنابراین، در این فرایند ممکن است بسیاری از واژه‌ها و ترکیب‌های جدید هنگام خواندن نادیده گرفته شوند و یا حتی ممکن است، فارسی‌آموزان تصور کنند که معنای واژه را درست حدس زده‌اند و از آن عبور کنند؛ در حالی که ممکن است چنین نباشد. از طرفی دیگر، توقف بر درک واژه‌های ناآشنا، پردازش کلی متن و درک

^۱ behaviorism

^۲ audio-lingual method

مطلب را مختل می‌کند (Laufer, 2005). همچنین «در هر برنامه آموزشی واژه، وقتی که هدف از آموزش، دستیابی به عمیق‌ترین سطح دانش واژگانی باشد، آموزش مستقیم واژه در متن، نسبت به یادگیری تصادفی ارجح‌تر است و رها کردن زبان‌آموزان در موقعیت یادگیری تصادفی، تأثیر چندانی در تثبیت و ماندگاری معنی واژه در پی نخواهد داشت» (Sahraei & Ahmadighader, 2016)؛ این یافته با یافته‌های اشمیت و سنبل (Sonbul & Schmitt, 2010) همسو است. در خصوص آموزش ترکیب‌های باهم‌آیی نیز با توجه به آماری که در بخش یافته‌های پژوهش حاضر ارائه شد، می‌توان نتیجه گرفت؛ آموزش صریح محدودیت‌های موجود در گزینش واژه‌های هم‌نشین و آگاهی دادن به زبان‌آموزان در رشد دانش واژگانی آن‌ها کمک‌کننده است. بنابراین، در خصوص آموزش باهم‌آیی‌ها در کلاس درس، طراحی و اجرای فعالیت‌هایی که توجه مستقیم زبان‌آموزان را به ترکیب‌ها و نوع ارتباط اجزای آن‌ها با یکدیگر جلب کند، لازم است؛ از جمله این فعالیت‌ها می‌توان به فعالیت‌های جورکردن واژه‌های باهم‌آیند، دسته‌بندی واژه‌ها و طبقه‌بندی واژه‌های متعلق به یک گستره باهم‌آیی و ایجاد شبکه واژگانی اشاره کرد. همچنین، تلاش برای رسیدن به یک روش آموزشی مناسب در راستای تهیه برنامه درسی و تولید محتوای آموزشی برای باهم‌آیی‌ها، یکی دیگر از پیشنهادات مهم آموزشی نگارندگان این پژوهش است. به‌طور کلی، ترتیب و توالی آموزش باهم‌آیی‌ها در منابع آموزشی، از اهمیت ویژه‌ای برخوردار است که می‌توان به این گونه عمل کرد؛ اول آن که پیشنهاد می‌شود، از همان ابتدای آموزش زبان فارسی، آموزش باهم‌آیی‌ها در منابع جدی گرفته شود. دوم؛ آن که در سطوح ابتدایی فارسی‌آموزی، آموزش باهم‌آیی‌ها با الگوی نحوی «اسم – فعل سبک» مطابق فهرست باهم‌آیی‌های پرتکرار که ارائه شده است، در اولویت قرار گیرد و با دسته‌بندی و قالب مشخصی گنجانده شوند؛ تا فارسی‌آموزان سریع‌تر بر پرکاربردترین الگوی نحوی باهم‌آیی‌ها؛ یعنی «اسم – فعل سبک» تسلط یابند و بتوانند زودتر متون کم‌اشتباه قابل قبولی تولید کنند؛ و سوم آن که در سطوح پیش‌میان و بالاتر، زمانی که زبان‌آموز در مسیر رشد دانش واژگانی قرار می‌گیرد، برای آموزش باهم‌آیی‌های الگوی نحوی «اسم – فعل واژگانی» وقت آموزشی بیشتری اختصاص دهند؛ زیرا آنچه در رشد تنوع باهم‌آیی‌ها مهم است، یادگیری و کاربرد درست الگوی نحوی «اسم – فعل واژگانی» است.

از مهم‌ترین نقاط قوت مطالعه حاضر می‌توان به این اشاره کرد که پژوهش‌گران مطالعه حاضر، حین تحلیل پیکره نمونه و مطالعات آرای محققان پیشین، به الگویی دقیق برای شناسایی ترکیبات باهم‌آیی‌ها دست یافتند که می‌تواند مطالعه بر روی باهم‌آیی‌ها را برای پژوهش‌گران دیگر آسان‌تر کند. همچنین نگارندگان، پیکره نمونه را به‌طور دستی و با دقت و صرف زمان چندین ماهه کدگذاری کردند و این امر را به کدگذاری ماشینی نسپردند؛ تا سطح دقت نتیجه‌گیری افزایش یابد. باوجود این، مطالعه پیش رو دارای محدودیت‌هایی نیز هست. سه مورد از مهم‌ترین محدودیت‌ها عبارتند از: (۱) نبود پیکره فارسی‌آموزی که کد دستوری خورده باشد؛ (۲) نبود نرم‌افزاری مطمئن با سطح دقت بالا برای برچسب‌زنی ماشینی؛ وجود چنین پیکره و نرم‌افزاری که در دو محدودیت اول اشاره شد موجب می‌شود پژوهشگران این حوزه به جای تجزیه و تحلیل پیکره نمونه، از کل پیکره توسط نرم‌افزار داده‌های مورد نظر استخراج کنند و تنها با

پالایش کل خروجی، داده‌های اصلی خود برای تجزیه و تحلیل را به‌دست آورند. نمونه چنین نرم‌افزارهایی برای زبان انگلیسی موجود است که موجب می‌شود در مدت زمان کمتری پیکره‌های گسترده‌ای را با ۹۵ درصد اطمینان تجزیه و تحلیل کنند. (۳) نبود امکان مقایسه پیکره فارسی آموز با پیکره بومیان زبان فارسی؛ چنانچه در پژوهش‌های آتی چنین مقایسه‌ای صورت پذیرد، قطعاً به یافته‌های قابل توجهی در این حوزه مطالعاتی دست خواهیم یافت.

فهرست منابع:

- پناهی، ثریا. (۱۳۸۱). فرایند باهم آیی و ترکیبات باهم آیند در زبان فارسی. نامه فرهنگستان. دوره پنجم، شماره ۳.
- جزایری، زهرا. (۱۳۹۸). تأثیر آموزش باهم آیی‌های واژگانی بر تقویت مهارت صحبت کردن فارسی آموزان سطح میانی. (پایان نامه کارشناسی ارشد)، دانشگاه علامه طباطبائی، تهران، ایران.
- جلیلی، اکبر. (۱۴۰۱). تحلیلی بر دایره‌واژگانی فعال فارسی آموزان غیرایرانی در گفتار و نوشتار. (رساله دکتری)، دانشگاه علامه طباطبائی، تهران، ایران.
- جلیلی، اکبر. صحرانی، رضامراد. (۱۴۰۱). تنوع واژگانی در گفتار فارسی آموزان چینی زبان و عربی زبان: پژوهشی بر پایه شاخص گیراد. پژوهشنامه آموزش زبان فارسی به غیرفارسی زبانان. قزوین: دانشگاه بین‌المللی امام خمینی (ره). / آماده انتشار.
- حافظنیا، محمدرضا. (۱۳۸۹). مقدمه‌ای بر روش تحقیق در علوم انسانی. تهران: انتشارات سمت.
- حبیبی، آرش؛ سرآبادانی، مونا. (۱۴۰۱). آموزش کاربردی اس‌پی‌اس/اس. تهران: انتشارات نارون.
- صحرانی، رضامراد. احمدی‌قادر، شهناز (۱۳۹۴) آموزش مستقیم واژه در متن: مقایسه تأثیر دو رویکرد یادگیری مستقیم و تصادفی در یادگیری واژه. فصل‌نامه علمی زبان پژوهی. تهران: دانشگاه الزهراء. دوره ۷، شماره ۱۷، ص ۷۷-۱۰۲.
- صحرانی، رضامراد. مرصوص، فائزه (۱۳۹۵) استاندارد مرجع آموزش زبان فارسی در جهان. تهران: انتشارات دانشگاه علامه طباطبائی.
- صفوی، کوروش. (۱۳۹۵). درآمدی بر معنی‌شناسی. تهران: انتشارات سوره مهر (وابسته به حوزه هنری).
- میرزائی، آزاده. (۱۳۹۶). آشنایی با زیانشناسی پیکره‌ای. تهران: انتشارات دانشگاه علامه طباطبائی.

References:

- Aisenstadt, E. (1979). Collocability restrictions in dictionaries. In R. R. K. Hartmann (Ed.), *Dictionaries and their users: Papers from the 1978 B.A.A.L. seminar on lexicography* (pp. 71-74). Exeter: University of Exeter.
- Aisenstadt, E. (1981). Restricted collocations in English lexicology and lexicography. *ITL: Review of Applied Linguistics*, 53, 53-61.
- Benson, M., Benson, E., & Ilson, R. (2010). *The BBI combinatory dictionary of English: Your guide to collocations and grammar* (3rd ed.). Amsterdam: Benjamins.
- Biber, D., Johansson, S., Leech, G., et al. (1999). *Longman grammar of spoken and written English*. Harlow: Longman.
- Church, K., & Hanks, P. (1990). Word association norms, mutual information, and lexicography. *Computational Linguistics*, 16, 22-29.
- Church, K., & Hindle, D. (1990). Collocational constraints and corpus-based linguistics. In *Working Notes of the AAAI Symposium: Text-Based Intelligent Systems*.

- Church, K., Gale, W., Hanks, P., et al.** (1991). Using statistics in lexical analysis. In U. Zernik (Ed.), *Lexical acquisition: Exploring on-line resources to build a lexicon* (pp. 115–164). Hillsdale: Erlbaum.
- Cowie A. P.** (1992). Multiword lexical units and communicative language teaching. In P. Arnaud, H. Bejoint (Eds.), *Vocabulary and applied linguistics* (pp. 1–12). London: Macmillan.
- Cowie, A. P.** (1981). The treatment of collocations and idioms in learners' dictionaries. *Applied Linguistics*, 2(3), 223–235.
- Cowie, A. P.** (1998). *Phraseology: Theory, analysis and applications*. Oxford: Oxford University Press.
- Daller, H. & Xue, H.** (2007). Lexical richness and the oral proficiency of Chinese EFL students; A comparison of different measures. In Daller, H., Milton, J. & Treffers-Daller, J. (Eds), *Modelling and assessing vocabulary knowledge* (pp. 150-164). Cambridge University Press.
- Firth, J. R.** (1957). *Papers in linguistics 1934–1951*. London: Oxford University Press.
- Gitsaki, C.** (1999). *Second language lexical acquisition: A study of the development of collocational knowledge*. San Francisco: International Scholars Publications.
- Granger, S.** (1998). Prefabricated patterns in advanced EFL writing: Collocations and formulae. In A. P. Cowie (Ed.), *Phraseology: Theory, analysis, and applications* (pp. 145–160). Oxford: Oxford University Press.
- Greenbaum, S.** (1974). Some verb-intensifier collocations in American and British English. *American Speech*, 49(1, 2), 79–89.
- Habibi, A., Sarabadani, M.** (2022). SPSS Practical Training. Tehran: Naroon Publications. [In Persian]
- Hafeznia, M.** (2010). *An Introduction to Research Methods in the Humanities*. Tehran: SAMT Publications. [In Persian]
- Halliday, M. A. K.** (1966). Lexis as a linguistic level. In C. E. Bazell, J. C. Catford, M. A. K. Halliday, et al. (Eds.), *In memory of J. R. Firth* (pp. 148–162). London: Longman.
- Hausmann, F. J.** (1989). Le dictionnaire de collocations. In F. J. Hausmann, O. Reichmann, H. E. Wiegand, et al. (Eds.), *Wörterbücher: ein internationales Handbuch zur Lexicographie. Dictionaries. Dictionnaires* (pp. 1010–1019). Berlin: De Gruyter.
- Hoey, M.** (1991). *Patterns of lexis in Text*. Oxford: Oxford University Press.
- Howarth, P.** (1996). *Phraseology in English academic writing: Some mplications for language learning and dictionary making*. Tübingen: Niemeyer.
- Howarth, P.** (1998). The phraseology of learners' academic writing. In A. P. Cowie (Ed.), *Phraseology: Theory, analysis and applications* (pp. 161–186). Oxford: Oxford University Press.
- Jalili, A.** (2022). An analysis of the active vocabulary of non-Iranian Persian learners in speech and writing. (PhD. Dissertation). Allameh Tabataba'i university, Tehran, Iran. [In Persian]
- Jalili, A. Sahraei, R.M** (2022). Lexical Diversity in the Speech of Chinese and Arab Persian Language Learners: A Research Based on Guiraud's index. *Journal on Teaching Persian to Speakers of Other Languages (JTPSOL)*. Qazvin: Imam Khomeini International University. / Ready for Publication. [In Persian]

- Jazayeri, Z.** (2019). *The effect of teaching lexical collocations on improve intermediate persian learners speaking skill* (Master's thesis), Allameh Tabatabaei University, Tehran, Iran. [In Persian]
- Jones, S., & Sinclair, J.** (1974). English lexical collocations: A study in computational linguistics. *Cahiers de Lexicologie*, 23(2), 15–61.
- Kjellmer, G.** (1987). Aspects of English collocations. In W. Meijs (Ed.), *Corpus Linguistics and Beyond: Proceedings of the Seventh International Conference on English Language Research on Computerized Corpora* (pp. 133–140). Amsterdam: Rodopi.
- Kjellmer, G.** (1994). *A dictionary of English collocations: Based on the Brown corpus*. Oxford: Clarendon Press.
- Krejcie, R. V., & Morgan, D. W.** (1970). Determining sample size for research activities. *Educational and psychological measurement*, 30(3), 607-610.
- Krejcie, R.V., & Morgan, D.W.** (1970). Determining Sample Size for Research Activities. *Educational and Psychological Measurement*.
- Laufer, B.** (2005). Focus on Form in Second Language Vocabulary Learning. *EUROSLA Yearbook* 5:223-250
- Laufer, B., & Waldman, T.** (2011). Verb-noun collocations in second language writing: A corpus analysis of learners' English. *Language Learning*, 61(2), 647–672.
- Leech, G.** (1974). *Semantics*. Harmondsworth: Penguin.
- Lewis, M.** (2000). *Teaching collocation: Further developments in the lexical approach*. London: Language Teaching Publications.
- Lorenz, G.** (1999). *Adjective intensification—Learners versus native speakers: A corpus study of argumentative writing*. Amsterdam: Rodopi.
- Lu, Y.** (2017). *A Corpus Study of Collocation in Chinese Learner English*. New York: Routledge.
- McEnery, T., & Wilson, A.** (1996). *Corpus linguistics*. Edinburgh: Edinburgh University Press.
- McEnery, T., Xiao, R., & Tono, Y.** (2006). *Corpus-based language studies: An advanced resource book*. London: Routledge.
- Meara, P.** (1984). The study of lexis in interlanguage. In A. Davies, C. Criper, & A. P. R. HOWATT (Eds.), *Interlanguage* (pp. 225–235). Edinburgh: Edinburgh University Press.
- Men, H. Y.** (2018). *Vocabulary Increase and Collocation Learning, a Corpus-Based Cross-sectional Study of Chinese Learners of English*. Singapore: Springer Nature
- Mirzaei, A.** (2017). *An introduction to corpus linguistics*. Tehran: Allameh Tabataba'i University Press. [In Persian]
- Moon, R.** (1998). *Fixed expressions and idioms in English: A corpus-based approach*. Oxford: Clarendon Press.
- Nattinger, J. R. and DeCarrico, J. S.** (1992). *Lexical Phrases and Language Teaching*. Oxford: Oxford University Press.
- Nesselhauf, N.** (2003). The use of collocations by advanced learners of English and some implications for teaching. *Applied Linguistics*, 24(2), 223–242.
- Nesselhauf, N.** (2005). *Collocations in a learner corpus*. Amsterdam: Benjamins.
- Palmer, H. E.** (1933). *Second interim report on English collocations*. Tokyo: Kaitakusha.
- Panahi, S.** (2002). The process of collocations and collocational compounds in Persian. *Nameh-ye Farhangestan*, Volume 5, Issue 3. Tehran [In Persian]

- Partington, A.** (1998). *Patterns and meanings: Using corpora for English language research and teaching*. Amsterdam: Benjamins.
- Safavi, K.** (2016). *An Introduction to Semantics*. Tehran: Surah Mehr Publications (affiliated with the Arts Department). [In Persian]
- Sahraei, R.M. Ahmaighader, Sh.** (2016). Direct Teaching of Vocabulary in Context: the Comparison of Effect of Direct and Incidental Teaching in Learning Vocabulary. *Journal of Language Research (ZABANPAZHUI)*. Tehran: Alzahra University. Volume 7, Issue 17 - Serial Number 17, Pages 77-102. [In Persian]
- Sahraei, R.M. Marsous, F.** (2016). *Persian Teaching Standard Reference (PTSR)*. Tehran: Allameh Tabatabaei University Press. [In Persian]
- Sinclair, J.** (1966). Beginning the study of lexis. In C. E. Bazell, J. C. Catford, & M. A. K. Halliday, et al. (Eds.), *In memory of J. R. Firth* (pp. 410–430). London: Longman.
- Sinclair, J.** (1987). Collocation: A progress report. In R. Steele & T. Threadgold (Eds.), *Language topics: Essays in honour of Michael Halliday* (Vol. 2, pp. 319–331). Amsterdam: Benjamins.
- Sinclair, J.** (1991). *Corpus, concordance, collocation*. Oxford: Oxford University Press.
- Sinclair, J.** (2004). *Trust the text*. London: Routledge.
- Sinclair, J., Jones, S., & Daley, R.** (2004). *English collocation studies: The OSTI report*. London: Continuum.
- Sonbul, S. & Schmitt, N.** (2010). Direct teaching of vocabulary after reading: Is It Worth the Effort? *ELT Journal* Volume, 64(3), 253-260.
- Stubbs, M.** (1995). Collocations and semantic profiles: On the cause of the trouble with quantitative study. *Functions of Language*, 2(1), 23–55.
- Stubbs, M.** (1996). *Text and corpus analysis: Computer-assisted studies of language and culture*. Oxford: Blackwell.
- Stubbs, M.** (2001). *Words and phrases: Corpus studies of lexical semantics*. Oxford: Blackwell.
- Van Roey, J.** (1990). *French-English contrastive lexicology: An introduction*. Louvain-la-Neuve: Peeters.
- Wray, A.** (2002). *Formulaic language and the lexicon*. Cambridge: Cambridge University Press.
- Zhang, X.** (1993). *English collocations and their effect on the writing of native and non-native college freshmen*. Indiana: Indiana University of Pennsylvania.