

Investigating the Correctness of Attributing the Munajat Khams 'Ashar to Imam Sajjad (PBUH) Based on Stylometry Techniques

Vol. 15, No. 6, Tome 84
pp. 137-171
January & February
2025

Isa Motaghizadeh^{1*}  & Somayyeh Modiri² 

Abstract

Advances in science and technology have made it no longer acceptable to have works with a dubious author. Stylometry is a method that uses statistical analysis to determine the author of a literary work. Author attribution methods rely heavily on writing style; assuming that each person has unique style. Author identification is used in areas such as plagiarism, criminology, and unspecified author identification. Due to the fact that many factors are involved in identifying the author of texts, a method with 100% accuracy has not been presented so far, and researchers are still trying to find a way to minimize computational errors. One of the methods that is claimed to have good accuracy is Yule's theory. In this article, Yule's theory and four other theories have been combined to compare the vocabulary richness of the Munajat Khams 'Ashar and the prayers of Al-Sahifa al-Sajjadiyya. Then, Using descriptive-analytical method and explanation of statistical datas, the correctness of the attribution of Munajat Khams 'Ashar to Imam Sajjad (PBUH) has been investigated. The results show the high accuracy of the calculations and the independence of the output of the theories to the length of the text. Also, due to the slight difference between the vocabulary richness of the Munajat Khams 'Ashar and the prayers of Al-Sahifa al-Sajjadiyya, its attribution to Imam Sajjad (PBUH) is confirmed.

Keywords: author attribution, stylometry, vocabulary richness, Munajat Khams 'Ashar, Al-Sahifa al-Sajjadiyya

Received: 11 April 2022
Received in revised form: 24 June 2022
Accepted: 9 July 2022

¹ Corresponding Author: Professor, Department of Arabic Language and Literature, Tarbiat Modares University, Tehran, Iran; ORCID ID: <http://orcid.org/0000-0002-5486-6059>;

Email: motaghizadeh@modares.ac.ir

² PhD Student, Arabic Language and Literature, Kharazmi University, Tehran, Iran.;

ORCID ID: <http://orcid.org/0000-0003-0386-5483>

1. Introduction

The issue of attributing a text to someone who did not really write it, has always been the focus of researchers. With the advancement of science in the twentieth century, the need to prove the accuracy of attributing a text to a particular author has intensified, and with the advancement of information technology, the popularity of intelligent methods of author recognition has increased. Today, to identify the author of a text, various methods are used, one of the most important methods is study the writing style.

The study of writing style is a subset of the new rhetoric. The new rhetoric aims at adding formal logic a field of reasoning, and applies whenever action is linked to rationality (Perelman, 1971). In stylistics, using text reasoning and analysis, characteristics are considered for the author's style.

A variety of methods for attribution have been proposed. There are three main approaches: lexical methods, syntactic or grammatic methods, and language-model methods, including methods based on compression (Zhao & Zobel, 2005). In this article, the lexical method will be used. One of the most practical lexical methods to achieve the author's style is the "vocabulary richness" method. Unfortunately, the output of many methods depends on the length of the text. Therefore, a method should be used that has the least dependence on the length of the text. In this paper, we have combined five theories to calculate vocabulary richness to achieve the most accurate results.

Research Question(s)

1. How accurate and reliable are the results of the five equations used in this research?
2. How much does the output of the theories depend on the length of the text?
3. What is the difference between the vocabulary richness of Munajat Khams 'Ashar and the prays of Al-Sahifa al-Sajjadiyya?

2. Literature Review

Authorship attribution (AA) is the process of attempting to identify the likely authorship of a given document, given a collection of documents whose authorship is known (Bozkurt et al., 2007). The accepted assumption behind AA is that every author writes in a distinct way; some writing characteristics cannot be manipulated by the writer's will, and therefore can be identified by an automated process (Howedi & Mohd, 2014).

One of the fundamental sub-problems of AA is the extraction of the most suitable features to represent the writing style of each author. This problem is known as "stylometry" (Howedi et al., 2020, p. 1334). stylometry is defined as those techniques that allow measure the style of an author by the identification of its features of style (stylemas). Those stylemas, also called style markers, are obtained from textual measurements normally calculated by statistical methods (Escobedo et al., 2013, Stamatatos, 2009).

Some researchers have used a combination of some lexical richness functions to achieve better results, namely: K proposed by Yule (1944), R proposed by Honore (1979), W proposed by Brunet (1978), S proposed by Sichel (1975), and D proposed by Simpson (1949) which are defined as follows (Stamatatos et al., 2000):

$$K = \frac{10^4 (\sum_{i=1}^{\infty} i^2 v_i - N)}{N^2}$$

$$R = \frac{100 \log N}{(1 - (\frac{V_1}{V}))}$$

$$W = N^{V^{-\alpha}}$$

$$S = \frac{V_2}{V}$$

$$D = \sum_{i=1}^V v_i \frac{i(i-1)}{N(N-1)}$$

where:

V_i : is the number of words used exactly i times

N: Total number of words

V: Number of non-repetitive words

α : usually is fixed at 0.17

The final output for calculating vocabulary richness is obtained by combining these five equations.

Since the series of narrators and the document of Munajat Khams 'Ashar is not mentioned completely in the available sources, attributing it to Imam Sajjad (PBUH) needs to be proved, so in this research, using stylometry techniques, it is examined.

3. Methodology

In the present article, the correctness of attributing Munajat Khams 'Ashar to Imam Sajjad (PBUH) is examined by sampling the prays of Al-Sahifa al-Sajjadiyya and comparing his vocabulary richness with the Munajat Khams 'Ashar. Since, according to the claim, the output of the theories is not dependent on the length of the text, two statistical populations are selected: the first consists of prays which 80 words have been selected, and the second consists of prayers With different number of words; Therefore, in addition to comparing the vocabulary richness of the samples, the dependence of the equations on the length of the text will also be examined. Also, From Munajat Khams 'Ashar, we chose the first, fifth, tenth and fifteenth prays as samples.

4. Results

The results show that:

1. The accuracy of the calculations is very high and therefore the output of the theories is reliable.
2. The output of the theories was not dependent on the length of the text and did not increase in proportion to the increase in the number of words.
3. There is not much difference between the vocabulary richness of Munajat Khams 'Ashar and the prays of Al-Sahifa al-Sajjadiyya in both statistical populations; Therefore, the correctness of attributing the Munajat Khams 'Ashar to Imam Sajjad (PBUH) - from the perspective of stylometry techniques - is proved.

[DOI: 10.29252/LRR.15.6.5]

[DOR: 20.1001.1.23223081.1401.0.0.175.7]

[Downloaded from lrr.modares.ac.ir on 2024-11-28]



دوماهنامه بین‌المللی

د ۱۵، ش ۶ (پیاپی ۸۴)، بهمن و اسفند ۱۴۰۳، صص ۱۳۷-۱۷۱

مقاله پژوهشی

<http://dorl.net/dor/20.1001.1.23223081.1401.0.0.175.7>

بررسی درستی استناد مناجات خمس عشر به امام سجاد (ع) بر اساس تکنیک‌های سبک‌سنجی

عیسی متقی‌زاده^{۱*}، سمیه مدیری^۲

۱. استاد گروه زبان و ادبیات عربی، دانشگاه تربیت مدرس، تهران، ایران.

۲. دانشجوی دکتری زبان و ادبیات عربی، دانشگاه خوارزمی، تهران، ایران.

تاریخ پذیرش: ۱۴۰۱/۰۴/۱۸

تاریخ دریافت: ۱۴۰۱/۰۱/۲۲

چکیده

با پیشرفت علم و تکنولوژی، دیگر وجود آثار با نویسنده مشکوک، پذیرفتنی نیست. سبک‌سنجی روشی است که با استفاده از تجزیه و تحلیل آماری، نویسنده آثار ادبی را تعیین می‌کند. روش‌های انتساب نویسنده، عمدتاً متکی بر شناسایی نویسنده بر اساس سبک نوشتاری هستند؛ با این پیش‌فرض که سبک نوشتاری هر شخص، دارای ویژگی‌های منحصر به فردی است. تشخیص هوشمند نویسنده در زمینه‌هایی مانند: سرقت ادبی، جرم‌شناسی و تعیین نویسنده نامشخص، کاربرد دارد. به دلیل آنکه عوامل بسیاری در تشخیص نویسنده متون، دخیل هستند، تاکنون روشی با دقت ۱۰۰ درصد ارائه نشده است و پژوهشگران همچنان در تلاش هستند روشی را ارائه کنند که خطای محاسباتی را به حداقل ممکن برسانند. یکی از روش‌هایی که ادعا شده از دقت خوبی برخوردار است نظریه یول است. در جستار حاضر، نظریه یول و چهار نظریه دیگر در زمینه غنای واژگانی، ترکیب شده‌اند و با بهره‌گیری از روش توصیفی - تحلیلی و تبیین داده‌های آماری، مقایسه‌ای میان غنای واژگانی مناجات خمس عشر و دعاهای صحیفه سجادیه انجام شده تا درستی انتساب این مناجات به امام سجاد (ع) مورد بررسی قرار گیرد. نتایج تحقیق، بیانگر دقت بالای محاسبات انجام شده و نیز عدم وابستگی خروجی نظریه‌ها به طول متن است. همچنین به دلیل عدم وجود اختلاف قابل ملاحظه، میان غنای واژگانی مناجات خمس عشر و دعاهای صحیفه سجادیه، انتساب این مناجات به امام سجاد (ع) تأیید می‌شود.

واژه‌های کلیدی: انتساب نویسنده، سبک‌سنجی، غنای واژگانی، مناجات خمس عشر، صحیفه سجادیه.

E-mail: motaghizadeh@modares.ac.ir

* نویسنده مسئول مقاله:

۱. مقدمه

در طول تاریخ بشریت، همواره متونی وجود داشته‌اند که با انگیزه‌های مختلف، به نویسندگانی غیرواقعی نسبت داده می‌شده‌اند و محققان نیز همواره در تلاش بوده‌اند که با روش‌های مختلف و با بهره‌گیری از علم زمان خود، درستی یا نادرستی این نسبت‌ها را اثبات کنند. با پیشرفت علم در قرن بیستم، احساس نیاز به اثبات درستی استناد یک متن به نویسنده مشخص، شدت یافته و با رشد چشمگیر فناوری اطلاعات، اقبال به روش‌های هوشمند تشخیص نویسنده نیز رو به فزونی رفته است. امروزه، جهت تشخیص نویسنده یک متن، از روش‌های مختلفی استفاده می‌شود که یکی از مهم‌ترین آن‌ها بررسی سبک نوشتاری^۱ است.

مطالعه سبک نوشتاری، زیرمجموعه بلاغت جدید^۲ به شمار می‌رود. هدف بلاغت جدید آن است که شاخه‌ای از استدلال^۳ را به منطق^۴ بیفزاید و هر جا که کار با عقلانیت^۵، پیوند خورده باشد کاربرد دارد (Perelman, 1971, p.148). در سبک‌شناسی نیز با استفاده از استدلال و تحلیل متن، ویژگی‌هایی برای سبک نویسنده، در نظر گرفته می‌شود.

واژه «سبک‌شناسی»^۶ نخستین بار در قرن نوزدهم ظاهر شد و در اوایل قرن بیستم به‌طور گسترده، مورد استفاده قرار گرفت. سبک‌شناسی مبتنی بر بلاغت، هم‌زمان با پیدایش زبان‌شناسی جدید^۷ پدید آمده (Fowler, 1987, p. 187; Bagirzade, 2019, p.41) و تلاشی برای پیوند دادن ادبیات و زبان‌شناسی است.

نخستین پژوهش‌های کمی^۸ مربوط به شناسایی نویسنده در سال ۱۸۷۷ توسط مندنهال^۹ هوشناس اهل اوهایو صورت گرفت که پیشنهاد کرد که از توزیع طول کلمات^{۱۰} به‌عنوان ویژگی وابسته به نویسنده استفاده شود (مندنهال، ۱۸۸۷؛ مرادی و بحرانی، ۲۰۱۵، ص. ۸۴). به‌طور کلی، سه رویکرد اصلی برای تشخیص نویسنده یک متن وجود دارد: روش واژگانی^{۱۱}، روش نحوی یا دستوری^{۱۲} و روش مدل زبانی^{۱۳} (Zhao & Zobel, 2005, p. 175). در این تحقیق از روش واژگانی استفاده خواهد شد.

یکی از کاربردی‌ترین روش‌های واژگانی جهت دستیابی به سبک نویسنده، روش «غنای واژگانی»^{۱۴} است. روش‌های مختلفی برای محاسبه غنای واژگانی وجود دارد؛ از جمله: محاسبه نسبت افعال به صفات، کلماتی که یک بار تکرار می‌شوند، معیار K، معیار R و ... متأسفانه، خروجی بسیاری از روش‌های سنجش غنای واژگانی، وابسته به طول متن^{۱۵} بوده و بنابراین با

کوتاه شدن و طولانی شدن متن، غنای واژگانی نیز افزایش یا کاهش می‌یابد؛ لذا باید از روشی استفاده کرد که کمترین وابستگی را به طول متن داشته باشد.

در جستار حاضر، جهت دستیابی به دقیق‌ترین نتایج، پنج معیار مختلف برای محاسبه غنای واژگانی را ترکیب کرده و با پاسخ به سؤالات زیر، درستی یا نادرستی انتساب مناجات خمس عشر به امام سجاد (ع) را بررسی خواهیم کرد:

۱- نتایج معادلات پنج‌گانه به‌کار گرفته شده در این تحقیق، از نظر دقت محاسباتی، تا چه میزان، قابل استناد است؟

۲- خروجی نظریه‌ها تا چه میزان، وابسته به طول متن مورد بررسی است؟

۳- میزان اختلاف مناجات خمس عشر با دعاهاى صحیفه سجادیه، از نظر غنای واژگانی چقدر است؟

۲. پیشینه پژوهش

با توجه به گسترش آثار نویسندگان نامعلوم و یا مشکوک، پژوهشگران تلاش کرده‌اند راه‌هایی جهت تشخیص نویسنده ارائه دهند. معرفی نظریه‌ها در این خصوص، غالباً در پژوهش‌های به زبان لاتین انجام شده است؛ از جمله:

بوزکورت^{۱۱} و همکاران (2007) از مجموعه ویژگی‌های مختلف، مانند: کلمات تابع، مجموعه ویژگی‌های سبک‌سنجی و همچنین طبقه‌بندی‌کننده‌هایی مانند: طبقه‌بند بیز با چگالی گاوسی، ماشین‌های بردار پشتیبان، هیستوگرام، روش k و ... استفاده کرده و بهترین نتایج را با طبقه‌بندی‌کننده «گاوسی» و «ماشین بردار پشتیبان» به‌دست آوردند.

اولیوریا جونیور^{۱۲} و همکاران (2013) سه نوع مدل فشرده‌سازی مختلف برای انتساب نویسنده (BZip، GZip و PPM) را مقایسه کرده و نشان دادند که عملکرد نسبتاً مشابهی دارند.

هویدی^{۱۳} و همکاران (2020) برای تشخیص نویسنده از طبقه‌بند KNN همراه با ویژگی‌های سیکی استفاده کردند و با تطبیق بر متون کوتاه تاریخی عربی به نتایج بسیار دقیق (۹۰٪) دست یافتند.

ربابه^{۱۹} و همکاران (2016) از دو روش SF و BOW و ویژگی‌های متفاوت آن‌ها برای تشخیص نویسنده در توییت‌های عربی استفاده کردند و نتیجه گرفتند که ترکیب تمام این ویژگی‌ها، به بهترین نتایج منجر می‌شود.

تورلا و کاسادا^{۲۰} (2013) هفت معادلهٔ مختلف برای محاسبهٔ غنای واژگانی را از نظر وابستگی آن‌ها به طول متن مقایسه کردند و به این نتیجه رسیدند که خروجی چهار مورد از آن‌ها وابسته به طول متن نیست.

در زبان فارسی و عربی نیز پژوهش‌هایی - عمدتاً به صورت تطبیقی - انجام شده است؛ از جمله: متقی زاده و همکاران (2022) نامه‌های نهج البلاغه را از منظر دو فاکتور مطرح در سبک‌شناسی آماری (رویکرد انفعالی و غنای واژگانی) مورد بررسی قرار داده و به این نتیجه رسیده‌اند که اختلافات موجود در سبک نامه‌ها، به دلیل تأثیر شرایط اجتماعی بر شیوهٔ انتخاب کلمات است. وطن‌خواه و همکاران (2021) با نمونه‌گیری از آثار ملک الشعرای بهار و کدگذاری آن‌ها، به مدلی دست یافتند که بیانگر دو بلوک «بافت و محیط زمینه‌ای» و «سبک‌شناسی» است و به این نتیجه رسیده‌اند که وی در نظریهٔ خود، معتقد به تأثیر بافت و محیط زمینه‌ای بر سبک بوده، هرچند صریحاً به این مطلب، اشاره نکرده است.

مسجدی و همکاران (2021) به واکاوی نرم‌افزارهای رایج متن‌کاوی به عنوان ابزارهایی جهت استفاده از داده‌های کمی در جهت تولید اطلاعات با کیفیت و تحلیل‌پذیر پرداخته و نقاط ضعف و قوت آن‌ها را ارزیابی کرده‌اند.

زنگویی و نعمتی شمس‌آباد (2013) ویژگی‌های مختلف متن‌ها را در چهار دستهٔ ویژگی‌های لغوی، نحوی، خاص و ساختاری قرار داده و به تشخیص هوشمند نوشته‌های ۵۰ نویسندهٔ سایبری - کاربران سایت آمازون - پرداخته‌اند.

مدیری (2020) در پایان‌نامهٔ خود، براساس نظریهٔ «یول» و همکارانش و نظریهٔ «جانسون»، غنای واژگانی نهج البلاغه و صحیفهٔ سجادیه، را مقایسه کرده و نتیجه گرفته که تفاوت غنای واژگانی میان آن‌ها به دلیل تأثیر شرایط اجتماعی و مراعات حال مخاطب بر شیوهٔ انتخاب کلمات توسط امام علی (ع) و امام سجاد (ع) بوده است.

امیدوار و امید علی (2015) جهت بررسی اعتبار نسبت دیوان شعر به امام علی (ع)، طبق معادلهٔ یول، میان ۷۰۰۰ کلمه از نهج البلاغه و اشعار دیوان، مقایسه انجام داده و به این نتیجه

رسیدند که بین دو متن، تفاوت‌های زیادی وجود دارد و لذا نظر استاد حسن‌زاده آملی در مورد عدم صحت انتساب این دیوان به امام علی (ع) تأیید می‌شود.

یا حقی و ایزانلو (۲۰۰۶) با استفاده از شیوه «کیوسام» برای تشخیص نویسنده یکی از قصاید فارسی که به دو شاعر (سنایی و جبلی) انتساب یافته، کارایی این شیوه آماری را محک زده‌اند. در زمینه سبک‌شناسی مناجات خمس عشر نیز، پژوهش‌های مستقلی انجام شده است؛ از جمله:

لطفی (۲۰۱۲) در پایان‌نامه خود، سبک مناجات خمس عشر را در سطوح آوایی، نحوی، دلالی و همچنین از نظر انسجام مطالب، مورد بررسی قرار داده است. چراغی‌وش و همکاران (۲۰۱۳) سبک مناجات‌التائین را بر مبنای رویکرد ساختارگرایی، مورد تحلیل قرار داده و به این نتیجه رسیده‌اند که سطوح آوایی، نحوی و دلالتی، حول محور کلی متن (ارتباط با خدا) می‌چرخند. مشهدی‌نوش‌آبادی (۲۰۱۵) با بررسی بسامد واژگان مناجات خمس عشر و مقایسه آن با دعای عرفه و مناجات ابوحمزه ثمالی به این نتیجه رسیده است که این مناجات‌ها از نظر بسامد واژگانی، منطبق هستند و لذا صحت انتساب مناجات خمس عشر به امام سجاد (ع) اطمینان‌بخش است. بنابراین تاکنون، پژوهشی مستقل، جهت بررسی صحت استناد مناجات خمس عشر به امام سجاد (ع) با استفاده از تکنیک‌های سبک‌سنجی، انجام نشده است.

۳. چارچوب نظری

۳-۱. سبک‌شناسی آماری^{۲۱}

متون ادبی معمولاً با استفاده از ویژگی‌های زبانی^{۲۲} مشخص از یکدیگر تمیز داده می‌شوند و این ویژگی‌های زبانی زمانی که میزان تکرار بالایی داشته باشند، به ویژگی‌های سبکی^{۲۳} تبدیل می‌شوند. این نوع مطالعه، «سبک‌شناسی آماری» نامیده می‌شود و یکی از حوزه‌های مطالعات زبان‌شناسی سبک^{۲۴} معاصر است (مصلوح، ۱۹۹۲، ص. ۳۴).

سبک‌شناسی آماری، بر پایه دستیابی به ویژگی‌های سبک‌شناسی یک اثر ادبی با استفاده از کمیت^{۲۵} و ارزش‌های عددی، استوار است و برای تحقق این هدف، عناصر لغوی اثر ادبی را شمرده و یا بر طول کلمات و جمله‌ها و یا ارتباط بین صفت‌ها، اسم‌ها و فعل‌ها تمرکز می‌کند (بلیث، ۱۹۸۹، ص. ۳۷). امروزه، استفاده از روش‌های آماری در تحلیل سبک، به‌عنوان یک روش قابل اعتماد،

نزد پژوهشگران، پذیرفته شده و ثابت شده است (Biber, 1995). استفاده از آمار در تحلیل سبک اشخاص، به وجود معیاری برای مقایسه علمی و منطقی متون و در نتیجه، دور شدن از قضاوت‌های شخصی و دخالت عوامل بیگانه در تصمیم‌گیری نهایی منجر می‌شود.

۳-۲. انتساب نویسنده^{۲۶}

انتساب نویسنده‌گی روشی برای شناسایی نویسنده اثری است که مشخص نیست چه کسی آن را نوشته است (Bozkurt et al., 2007, p.1). پیش‌فرض این روش آن است که هر نویسنده‌ای سبک منحصر به فردی دارد که به‌طور ارادی، دستخوش تغییرات گسترده نمی‌شود (Howedi & Mohd, 2014, p. 48) و لذا با استفاده از روش‌های آماری و اندازه‌گیری برخی ویژگی‌های متنی، می‌توان میان آثار نویسندگان مختلف، تمایز قائل شد (Stamatatos, 2009, p.538). در این روش، با بررسی آثار هر شخص و دستیابی به ویژگی‌های مشخص سبک نگارش وی، درستی یا نادرستی انتساب یک اثر به او، مورد بررسی قرار می‌گیرد.

۳-۲-۱. کاربرد روش‌های تشخیص نویسنده

انتساب نویسنده در زمینه‌های مختلفی کاربرد دارد؛ از جمله: سرقت ادبی^{۲۷} (Zhao & Zobel, 2005, p.175)، ایمیل‌های ناشناس یا تهدیدآمیز (نویسندگان فضای مجازی)، حل معضلات تاریخی در تعیین نویسنده نامشخص یا مورد مناقشه یک متن (Bozkurt et al., 2007, p.1). اخیراً، کاربردهای عملی برای انتساب نویسنده‌گی در زمینه‌هایی مانند قانون کیفری و قانون کپی رایت برای ردیابی نویسندگان، افزایش یافته است. علاوه بر این، سبک نگارش را می‌توان با استفاده از عوامل درون یک متن یا با مقایسه دو متن توسط یک نویسنده تحلیل کرد (Howedi et al., 2020, p. 1334-1335). در این تحقیق، از روش دوم، یعنی مقایسه دو اثر یک نویسنده، جهت تشخیص درستی انتساب مناجات خمس عشر به امام سجاد (ع) استفاده شده است.

۳-۲-۲. انتساب نویسنده‌گی و طبقه‌بندی متن^{۲۸}

انتساب نویسنده‌گی، یک نوع طبقه‌بندی متن است؛ با این تفاوت که در آن، سبک نوشتاری نیز به اندازه محتوای متن، اهمیت دارد، اما در طبقه‌بندی متن، تنها محتوا مهم است (Howedi et al., 2020, p. 1334). همچنین در انتساب نویسنده، مجموعه ویژگی‌ها، به اندازه طبقه‌بندی متن،

مشخص و از پیش تعیین شده نیست؛ بنابراین، این تفاوت‌ها کار انتساب مؤلف را دشوارتر می‌کند (Bozkurt et al., 2007, p.1).

محققان، تکنیک‌های طبقه‌بندی متن را ارتقا داده و آن‌ها را در انتساب نویسندگی به همراه مسائلی همچون شناسایی موضوع^{۲۹}، شناسایی زبان^{۳۰} و کشف ژانر^{۳۱}، به کار گرفته و مورد ارزیابی قرار می‌دهند (Stamatatos, 2009, p.561). به عبارت دیگر، انتساب نویسندگی در قیاس با طبقه‌بندی متن، گامی رو به جلو بوده و به دلیل اینکه ویژگی‌های سبک نگارش را نیز در نظر می‌گیرد، در شناسایی نویسنده، موفق‌تر عمل می‌کند.

۳-۳. سبک‌سنجی^{۳۲}

یکی از زیرمجموعه‌های اساسی انتساب نویسنده، استخراج ویژگی‌های مناسب برای نشان دادن سبک نوشتاری هر شخص است که تحت عنوان «سبک‌سنجی» شناخته می‌شود (Howedi et al., 2020, p. 1955). سبک‌سنجی عبارت است از مجموعه تکنیک‌هایی که با بررسی ویژگی‌های سبکی، امکان تشخیص سبک نویسنده را فراهم می‌کنند. این ویژگی‌ها معمولاً با روش‌های آماری، محاسبه می‌شوند (Escobedo et al., 2013, p. 605; Stamatatos, 2009, p. 541). در روش‌های سبک‌سنجی، از بسامد طیف وسیعی از ویژگی‌های زبانی برای کمی کردن^{۳۳} سبک استفاده می‌شود؛ مانند مؤلفه‌های نحوی^{۳۴}، عناصر ساختاری^{۳۵}، طول پاراگراف^{۳۶} و ... (Wright, 2014, p. 14).

از سال ۱۹۶۴ تا ۱۹۹۰، مطالعات سبک‌سنجی، در صدر قرار داشت و بهره‌گیری از خصوصیات واژگانی، از جمله تکرار کلمات، از سنتی‌ترین روش‌های سبک‌سنجی به‌شمار می‌رفت که تاریخ آن به قرن ۱۹ میلادی برمی‌گردد (فرهمن‌دپور و همکاران، ۲۰۱۲، ص ۲۷). با گسترش روزافزون تکنولوژی، امکان استفاده از روش‌های آماری دقیق‌تر و با خطای محاسباتی پایین‌تر، برای پژوهشگران فراهم شده است. در این تحقیق، از نظریه‌هایی برای تشخیص نویسنده استفاده می‌شود که ثابت شده است از دقت محاسباتی خوبی برخوردارند.

۳-۴. نظریه‌های غنای واژگانی

غنای واژگانی یکی از رایج‌ترین روش‌های تشخیص نویسنده است که در سال‌های اخیر، مورد اقبال زیادی قرار گرفته است. توابع غنای واژگان، تلاش‌هایی برای تعیین تنوع واژگان یک متن به

شمار می‌روند (Stamatatos, 2009, p.540). در طول صد سال گذشته، مطالعات مختلفی در این زمینه انجام شده و فرمول‌های آماری متعددی برای محاسبه غنای واژگانی متون ارائه شده و هدف تمامی آن‌ها ارائه روشی بوده است که خروجی آن، وابسته به اندازه متن نباشد (Torruella & Capsada, 2013, p. 448). طبق منابع موجود، قدیم‌ترین تعریف از غنای واژگانی، توسط یول ۳۷ بیان شده است (Tanaka-Ishii & Aihara, 2015, p. 483) که آن را K نامیده و به نظر او مستقل از حجم متن مورد بررسی است (Yule, 1944, p. 52-53):

$$K = 10000 \frac{S_2 - S_1}{S_1^2} = \frac{\sigma_d^2 + M_d^2 - M_d}{NM_d^2} = \frac{1}{N} \left(\frac{\sigma_d^2}{M_d^2} + 1 \right) \quad (۱)$$

به طوری که:

$$S_0 = N = S(f_x) \quad , \quad S_1 = S(f_x X) \quad , \quad S_2 = S(f_x X^2)$$

$$M_c = \frac{S_1}{W} \quad , \quad \sigma_c^2 = \frac{S_2}{W} - \frac{S_1^2}{W^2}$$

به جای استفاده از یک معیار واحد، برخی از محققان - برای دستیابی به نتایج بهتر - از ترکیب برخی توابع غنای واژگانی استفاده کرده‌اند (Stamatatos et al, 2000, p. 478) که عبارت‌اند از: K یول (1944)، R هونور ۳۸ (1979)، W بروننت ۳۹ (1978)، S سیشل ۴۰ (1975)، و D سیمپسون ۴۱ (1949) که به صورت زیر تعریف می‌شوند:

$$K = \frac{10^4 (\sum_{i=1}^{\infty} i^2 V_i - N)}{N^2} \quad (۲)$$

$$R = \frac{100 \log N}{\left(1 - \left(\frac{V_1}{V}\right)\right)} \quad (۳)$$

$$W = N^{V^{-\alpha}} \quad (۴)$$

$$S = \frac{V_2}{V} \quad (۵)$$

$$D = \sum_{i=1}^V V_i \frac{i(i-1)}{N(N-1)} \quad (۶)$$

جایی که:

V_i : تعداد کلماتی که i مرتبه استفاده شده‌اند

N : تعداد کل کلمات متن

V : تعداد کلمات غیر تکراری

α : معمولاً ۰.۱۷ فرض می‌شود.

خروجی نهایی برای محاسبه غنای واژگانی، از ترکیب این پنج معادله به دست می‌آید.

۳-۵. سند صحیفه سجادیه

در استناد صحیفه سجادیه به امام سجاد (ع)، میان علما هیچ اختلافی وجود ندارد و تاکنون هیچ شبهه و یا دلیل موثقی مبنی بر نادرستی انتساب آن به امام سجاد (ع)، وارد نشده است. علاوه بر بررسی اسناد این کتاب شریف در پژوهش‌های مختلف، با چهار دلیل، درستی انتساب آن به امام سجاد (ع) قابل اثبات است:

۱- وجود شرح‌های گوناگون و بی‌شماری که توسط علما و بزرگان بر صحیفه سجادیه نوشته شده و در تمامی آن‌ها این کتاب شریف، اثر امام سجاد (ع) دانسته شده است.
 ۲- فقها در مسائل فقهی به برخی از دعاهای این کتاب شریف که مضامین فقهی دارند، استناد کرده‌اند.

۳- برخی از بزرگان فقه و علم رجال، صدور آن را از امام سجاد (ع) معلوم و قطعی و از جمله بدیهیات و در نتیجه، بی‌نیاز از اثبات دانسته‌اند؛ از جمله: صاحب جواهر (النجفی الجواهری، ۱۹۸۳، ج. ۱۱/ص. ۱۵۸) و آیت الله بروجردی (المنتظری، ۱۹۹۵، ص. ۴۰).

۴- دلیل چهارم که از همه مهم‌تر است ادعای متواتر بودن آن توسط عده‌ای از بزرگان است. «خبر متواتر عبارت است از خبری که برای مخاطب و شنونده، آرامش نفس و اطمینان خاطر بیاورد؛ آن‌گونه سکون و آرامش که شک را از دل بزدايد و درخت یقین را در سرزمین دل بکارد و این آرامش خاطر، از راه گزارش جماعت بسیاری باشد که به حسب عادت، توافق این همه انسان بر کذب، محال و ممتنع باشد؛ نظیر علم ما به وجود مکه، به کلام الله بودن قرآن و ...» (محمدی، ۲۰۰۸، ج. ۳/ص. ۱۱۳). در علم اصول فقه، خبر متواتر، دارای حجیت قطعی است و نیازی به اثبات و بررسی ندارد (همان، ص. ۱۱۶).

از جمله افرادی که این ادعا را مطرح کرده‌اند، سید صدر - فقیه، اصولی، محدث، مفسر و از مراجع تقلید شیعی - است که صحیفه سجادیه را همانند قرآن و از متواترات دانسته است (الصدر، 1996، ص. ۲۸۵). مرحوم آقا بزرگ تهرانی نیز در کتاب *الذریعة صحیفه سجادیه* را به دلیل آنکه در هر عصری و هر طبقه‌ای از مردم، روایت می‌شده است، از جمله متواترات دانسته است (آقابزرگ الطهرانی، ۱۴۰۳ق، ج. ۱۵/ص. ۱۸). علاوه بر این، مرحوم مجلسی (ره) در *بحار الأنوار* مدعی است که ۵۶۱۰۰ سند برای آن وجود دارد (مجلسی، ۱۴۴۰ق، ج. ۱۰۷/ص. ۶۱). بنابراین ادله فراوانی برای اثبات درستی انتساب صحیفه سجادیه به امام سجاد (ع) وجود دارد که ذکر تمامی آن‌ها از حوصله بحث حاضر، خارج است و به همین مقدار، اکتفا می‌شود.

۳-۶. سند مناجات خمس عشر

مناجات خمس عشر، شامل ۱۵ مناجات از امام سجاد (ع) بوده که مرحوم قمی (ره) در *مفاتیح الجنان* (قمی، ۲۰۰۰، ص. ۹۹) آورده است. شیخ حر عاملی (ره) نیز این مناجات‌ها را از امام سجاد (ع) دانسته و در کتاب *الصحیفه السجادیة الثانیة* ذکر کرده است (حر عاملی، ۲۰۰۰، صص. ۲۵-۵۷). علامه مجلسی (ره) نیز در *بحار الانوار* اشاره کرده است که برخی از اصحاب امام (ع) در کتاب‌های خود، این مناجات را از امام سجاد (ع) روایت کرده‌اند (مجلسی، ۱۴۴۰، ج. ۹۱/ص. ۱۴۲). با وجود این، در هیچ‌یک از این منابع، سند کامل مناجات خمس عشر ذکر نشده است؛ لذا این مناجات از جمله روایات مرسل به‌شمار می‌رود.

۳-۶-۱. حدیث مرسل

حدیث (خبر) مرسل - که مقابل خبر مسند است - اصطلاحی است در علم حدیث و بدین معنی است که راوی، مطلبی را بدون واسطه از معصوم (ع) نقل کرده باشد؛ یا با واسطه‌ای که آن را عمداً یا سهواً ذکر نکرده است (العاملی، ۱۹۸۱، ص. ۱۰۶)؛ چه یک نفر از سلسله راویان حذف شده باشد یا بیشتر (صدر، ۱۹۵۷، ص. ۱۸۹). همچنین ممکن است این حذف، از اول یا از وسط و یا از آخر سلسله راویان انجام شده باشد و یا اینکه به‌صورت مبهم و مجمل ذکر شود؛ مانند: عن رجل، عن بعض اصحابنا، عن عدة، عن اناس و... (محمدی، ۲۰۰۸، ج. ۳/ص. ۳۰۶). احادیث مرسل از جمله «اخبار ضعیف» به‌شمار رفته و لذا از نظر فقهی، قابل استناد قطعی نیستند. فقها راه‌هایی برای حجت دانستن این‌گونه احادیث ذکر کرده‌اند؛ از جمله: ۱- اگر اصحاب و

فقهای امامیه به روایت عمل کرده باشند، ضعف سند، جبران می‌شود. ۲- اگر برخی از روایان حدیث، مورد اطمینان باشند که از اشخاص ضعیف، نقل حدیث نمی‌کنند، می‌توان به روایت استناد کرد. ۳- مضمون حدیث، آن قدر متقن و محکم باشد که انسان مطمئن شود جز از امام معصوم (ع) نمی‌توانسته صادر شده باشد (محقق داماد، ۱۴۰۶ق، ج ۴/صص ۴۶-۴۷).

از آنجایی که علامه مجلسی و شیخ حر عاملی، بدون ایراد هرگونه شبهه‌ای، مناجات خمس عشر را در آثار خود ذکر کرده‌اند، می‌توان این مناجات‌ها را از معصوم (ع) دانست اما انتساب آن به امام سجاد (ع) نیازمند اثبات است. در این تحقیق، با استفاده از تکنیک‌های سبک‌سنجی، درستی این انتساب، مورد بررسی قرار می‌گیرد.

۷-۳. روش تحقیق

در جستار حاضر، درستی انتساب مناجات خمس عشر به امام سجاد (ع)، با نمونه‌گیری از دعاهای صحیفه سجادیه و مقایسه میزان غنای واژگانی آن‌ها با مناجات خمس عشر، مورد بررسی قرار می‌گیرد. بنابراین لازم و ضروری است که جامعه آماری نمونه، به‌گونه‌ای تعریف شود که بتواند شیوه نگارش امام سجاد (ع) در صحیفه سجادیه را انعکاس دهد؛ لذا یکی از مهم‌ترین مسائل، چگونگی انتخاب دعاها بر حسب تعداد کلمات آن‌هاست.

در زمینه انتساب نویسندگی، پژوهش‌های اندکی به صراحت بر اندازه متن مورد مطالعه، متمرکز شده‌اند و درمورد حداقل تعداد کلمات مورد نیاز، اتفاق نظر وجود ندارد. برخی از پژوهشگران، انتساب نویسندگی را برای متون کوتاه بررسی کرده و دریافته‌اند که استفاده از ۲۵۰۰ کلمه در هر نمونه، نتیجه قابل اعتمادی را ارائه می‌دهد (Howedi et al., 2020, p.1335).

از آنجایی که مطابق ادعا، خروجی‌های معادلات پنج‌گانه مورد استفاده در پژوهش حاضر، وابسته به طول متن نیستند، دو جامعه آماری انتخاب می‌شود: جامعه آماری اول، شامل دعاهایی است که تعداد کلمات مساوی از آن‌ها - ۸۰ کلمه اول - انتخاب شده‌اند (چون کوتاه‌ترین دعای صحیفه سجادیه، دارای ۸۰ کلمه است) و جامعه آماری دوم، شامل دعاهایی با تعداد کلمات مختلف است؛ بنابراین، علاوه بر مقایسه غنای واژگانی نمونه‌ها، وابستگی یا عدم وابستگی این معادلات به طول متن نیز بررسی خواهد شد. همچنین برای مقایسه غنای واژگانی دعاهای صحیفه

سجاده و مناجات خمس عشر، چهار مناجات اول، پنجم، دهم و پانزدهم را به عنوان نمونه، انتخاب کردیم.

۴. تطبیق نظریه‌ها

همان‌طور که در بخش ادبیات نظری، توضیح داده شد، غنای واژگانی دعا‌های صحیفه سجاده و مناجات خمس عشر، از ترکیب مقدار متغیرهای پنج‌گانه و به دست آوردن مدل رگرسیون خطی^{۴۲} آن‌ها محاسبه می‌شود. رگرسیون خطی را می‌توان به شکل زیر فرموله کرد:

$$y_i = b_0 + z_1 b_{1i} + z_2 b_{2i} + \dots + z_r b_{ri} + e_i \quad (7)$$

بنابراین با قرار دادن پنج متغیر در آن به صورت زیر در می‌آید:

$$y = b_0 + b_1 K + b_2 R + b_3 W + b_4 S + b_5 D + e_i \quad (8)$$

که در آن y نشان‌دهنده غنای واژگانی در نمونه‌های انتخاب‌شده و e_i خطای آماری است. در ابتدا، مقدار تمامی متغیرها را برای نمونه‌های انتخاب‌شده به عنوان جامعه آماری اول، محاسبه کردیم و با استفاده از برنامه اس.پی.اس.اس. مقادیر مورد نیاز برای به دست آوردن معادله، به شرح زیر محاسبه شد:

جدول ۱: خلاصه مدل برای جامعه آماری اول

Table 1: Model Summary for the First Statistical Population

Model Summary				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.999 ^a	.998	.995	.00526
a. Predictors: (Constant), D, S, R, K, W				

R^2 تناسب مدل رگرسیون خطی را با داده‌های آماری نشان می‌دهد و مقدار آن بین ۰ و ۱ متغیر است. هر چه بالاتر باشد، تناسب بهتری دارد. همان‌طور که جدول ۱ نشان می‌دهد، مقدار R^2 برابر با ۰.۹۹۵ بوده و خطای آماری بسیار کوچک است (۰.۰۰۵)؛ بنابراین مدل رگرسیون بسیار مناسب بوده و دقت محاسبات نیز بسیار بالاست.

جدول ۲: جدول ANOVA برای جامعه آماری اول

Table 2: ANOVA Table for the First Statistical Population

ANOVA ^a						
Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	.050	5	.010	364.066	.000 ^b
	Residual	.000	4	.000		
	Total	.051	9			

a. Dependent Variable: Y
b. Predictors: (Constant), D, S, R, K, W

همچنین مقدار F (364.066) و مقدار $\text{Sig} = 0.000 < 0.05$ نشان‌دهنده مناسب بودن مدل رگرسیون خطی برای صحیفه سجادیه است، زیرا بیشتر تغییرات متغیر وابسته در مدل دیده می‌شود. برای تخمین پارامترها و در نتیجه، به دست آوردن مدل رگرسیون خطی از جدول زیر استفاده می‌کنیم:

جدول ۳: جدول محاسبه b_0, b_1, b_2, b_3, b_4 و b_5 برای جامعه آماری اول

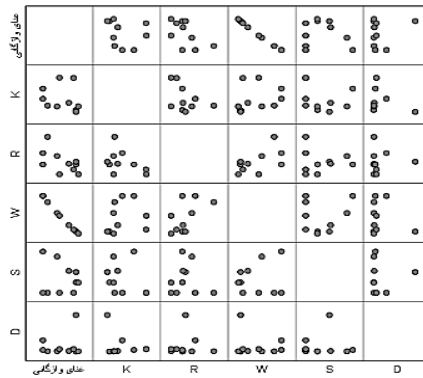
Table 3: Calculation Table b_0, b_1, b_2, b_3, b_4 and b_5 for the First Statistical Population

Coefficients ^a						
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	2.242	.040		55.674	.000
	K	-1.606E-5	.000	-.040	-1.027	.363
	R	.000	.000	-.042	-.955	.394
	W	-.165	.007	-.962	-23.471	.000
	S	-.050	.081	-.017	-.621	.569
	D	.004	.008	.013	.524	.628

a. Dependent Variable: Y

پس معادله رگرسیون خطی برای جامعه آماری اول به صورت زیر است:

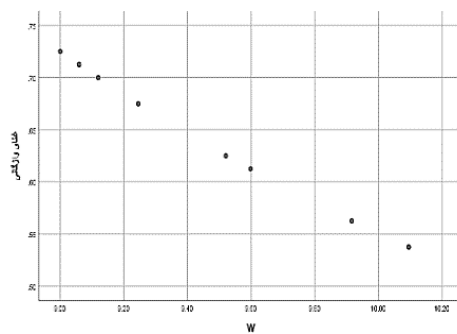
$$Y = 1.69744736 - 0.00344369 K - 0.00122531 R - 0.11750253 W - 0.50246854 S + 34.48886474 D$$



شکل ۱: نمودار Matrix Scatter برای جامعه آماری اول

Figure 1: Matrix Scatter for the First Statistical Population

هر ضریب (K، R، W، S و D) که بتای^{۴۳} بزرگتر داشته باشد، بیشترین تأثیر را بر تعیین غنای واژگانی، بر عهده دارد. همانطور که جدول ۳ نشان می‌دهد، بالاترین مقدار آن، مربوط به متغیر W است؛ بنابراین، متغیر W مهم‌ترین نقش را در تعیین غنای واژگانی دعاهای انتخاب‌شده در جامعه آماری اول ایفا می‌کند.



شکل ۲: نمودار Y-W برای جامعه آماری اول

Figure 2: Y-W diagram for the First Statistical Population

جهت مقایسه غنای واژگانی دعا‌های صحیفه سجادیه و مناجات خمس عشر، متغیر W را برای چهار مناجات منتخب نیز محاسبه کردیم.

جدول ۴: مقدار متغیر W در نمونه‌ها

Table 4: Variable Value of W in the Samples

مقدار متغیر W	مناجات	مقدار متغیر W	دعا
11.85	مناجات اول	9.598	دعا ۳۴
11.14	مناجات پنجم	9.0596	دعا ۱۹
10.33	مناجات دهم	9.06	دعا ۱۰
9.72	مناجات پانزدهم	9.0008	دعا ۲۹
		9.246	دعا ۳۳
		9.916	دعا ۳۵
		10.0945	دعا ۳۸
		10.0945	دعا ۵۴
		9.11993	دعا ۲۶
		9.5206	دعا ۲۷

طبق جدول ۴، میانگین مقدار متغیر W برای دعا‌های صحیفه سجادیه، ۹.۴۷ و برای مناجات‌های خمس عشر، ۱۰.۷۶ بوده و تنها ۱۲.۶٪ $(9.47/10.76-9.47) = 12.6\%$ اختلاف دارند؛ لذا در جامعه آماری اول، اختلاف قابل ملاحظه‌ای میان غنای واژگانی دعاها و مناجات‌های منتخب، وجود ندارد.

در مورد جامعه آماری دوم، دعا‌های صحیفه سجادیه را بر حسب تعداد واژگان، به سه گروه: ۱- دعا‌های کوتاه، ۲- دعا‌های متوسط و ۳- دعا‌های طولانی، دسته‌بندی کردیم. بدین صورت که دعاهایی را که تعداد واژگان آن‌ها کمتر از ۲۵۰ کلمه (یک صفحه) است به‌عنوان دعا‌های کوتاه، دعاهایی را که تعداد واژگان آن‌ها بین ۲۵۰ تا ۱۰۰۰ کلمه (۱ تا ۴ صفحه) است به‌عنوان دعا‌های متوسط و دعاهایی را که تعداد واژگان آن‌ها بیش از ۱۰۰۰ کلمه (بیش از ۴ صفحه) است به‌عنوان دعا‌های طولانی در نظر گرفتیم.

جدول ۵: دسته‌بندی دعاهای صحیفه سجادیه بر حسب تعداد واژگان

Table 5: Classification of the Prays of Al-Sahifa al-Sajjadiyya According to the Number of Words

تعداد کلمات	دعا
دعاهای کوتاه	۵۳-۴۳-۴۱-۴۰-۳۸-۳۶-۳۵-۳۴-۳۳-۳۰-۲۹-۲۸-۲۶-۱۹-۱۸-۱۵-۱۱-۱۰-۹-۸-۷
دعاهای متوسط	۴۲-۳۹-۳۷-۳۲-۳۱-۲۷-۲۵-۲۴-۲۳-۲۲-۲۱-۱۷-۱۶-۱۴-۱۳-۱۲-۶-۵-۴-۳-۲-۱ ۵۴-۵۲-۵۱-۵۰-۴۹-۴۸-۴۶-۴۴
دعاهای طولانی	۴۷-۴۵-۲۰

با توجه به تعداد دعاهایی که در هر دسته قرار می‌گیرند، ۳ دعا از دعاهای کوتاه، ۶ دعا از دعاهای متوسط و ۱ دعا از دعاهای طولانی را به‌طور تصادفی، به‌عنوان جامعه آماری انتخاب کردیم؛ بنابراین جامعه آماری دوم، شامل دعاهای زیر است:

جدول ۶: دعاهای منتخب برای جامعه آماری دوم

Table 6: Selected Prays for the Second Statistical Population

تعداد کلمات	دعا	گروه
۱۶۵	۹ دعا	کوتاه
۸۰	۱۸ دعا	
۱۵۸	۵۳ دعا	
۲۷۰	۲ دعا	متوسط
۳۷۷	۴ دعا	
۴۵۷	۲۵ دعا	
۴۳۹	۳۷ دعا	
۷۴۳	۴۴ دعا	
۴۰۶	۵۱ دعا	
۲۸۸۴	۴۷ دعا	طولانی

مقدار متغیرهای پنج‌گانه را برای جامعه آماری دوم نیز محاسبه کردیم و با استفاده از برنامه اس.پی.اس.اس. مقادیر مورد نیاز برای به‌دست آوردن معادله رگرسیون خطی به شرح زیر محاسبه شد:

جدول ۷: خلاصه مدل برای جامعه آماری دوم

Table 7: Model Summary for the Second Statistical Population

Model Summary				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.998 ^a	.995	.990	.00760

a. Predictors: (Constant), D, W, S, R, K

همان‌طور که جدول ۷ نشان می‌دهد مقدار R^2 برابر با ۰.۹۹ و خطای آماری بسیار کوچک است (۰.۰۰۸) و این نشان می‌دهد که مدل رگرسیون بسیار مناسب بوده و دقت محاسبات بسیار بالاست.

جدول ۸: جدول ANOVA برای جامعه آماری دوم

Table 8: ANOVA Table for the Second Statistical Population

ANOVA ^a						
Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	.050	5	.010	172.809	.000 ^b
	Residual	.000	4	.000		
	Total	.050	9			

a. Dependent Variable: غنای واژگانی
b. Predictors: (Constant), D, W, S, R, K

همچنین مقدار F (172.809) و مقدار $0.05 < \text{Sig} = 0.000$ نشان‌دهنده مناسب بودن مدل رگرسیون خطی است، زیرا بیشتر تغییرات متغیر وابسته در مدل دیده می‌شود. برای تخمین پارامترها و در نتیجه به‌دست آوردن مدل رگرسیون خطی از جدول ۹ استفاده می‌کنیم:

جدول ۹: جدول محاسبه b0، b1، b2، b3، b4 و b5 برای جامعه آماری دوم

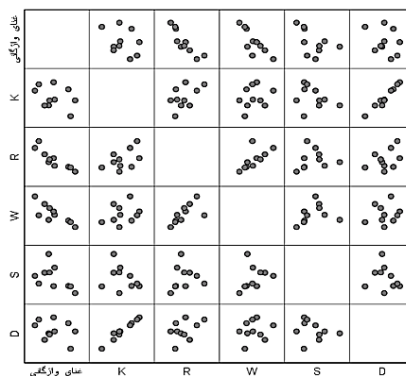
Table 9: Calculation Table b0, b1, b2, b3, b4 and b5 for the Second Statistical Population

Coefficients ^a						
Model	Unstandardized Coefficients			Standardized Coefficients	t	Sig.
	B	Std. Error	Beta			
1	(Constant)	1.334	.073		18.160	.000
	K	.000	.000	.473	1.191	.300
	R	-.002	.000	-.662	-12.098	.000
	W	-.027	.006	-.323	-4.612	.010
	S	-1.001	.233	-.301	-4.294	.013
	D	-4.073	2.925	-.506	-1.393	.236

a. Dependent Variable: غنای واژگانی

پس معادله رگرسیون خطی برای جامعه آماری دوم به صورت زیر است:

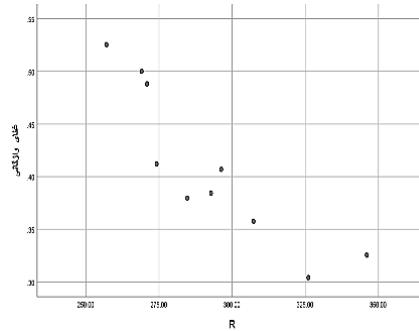
$$Y = 1.33353411 + 0.0003413009 K - 0.00178466 R - 0.02682203 W - 1.00136447 S - 4.07342408 D$$



شکل ۳: نمودار Matrix Scatter برای جامعه آماری دوم

Figure 3: Matrix Scatter for the Second Statistical Population

همان‌طور که جدول ۹ نشان می‌دهد، بالاترین مقدار ضریب بتا مربوط به متغیر R است؛ بنابراین، متغیر R مهم‌ترین نقش را در تعیین غنای واژگانی دعوای انتخاب‌شده در جامعه آماری دوم ایفا می‌کند.



شکل ۴: نمودار Y-R برای جامعه آماری دوم

Figure 4: Y-R Chart for the Second Statistical Population

جهت مقایسه غنای واژگانی دعا‌های صحیفه سجادیه و مناجات خمس عشر، متغیر R را برای چهار مناجات انتخاب شده، محاسبه کردیم.

جدول ۱۰: مقدار متغیر R برای نمونه‌ها در جامعه آماری دوم

Table 10: The Value of the Variable R for the Samples in the Second Statistical Population

مقدار متغیر R	مناجات	مقدار متغیر R	دعا
270.6165968701	مناجات اول	269.0443509	دعا ۲
262.7675769128	مناجات پنجم	270.8838905	دعا ۴
240.4553250325	مناجات دهم	274.1616513	دعا ۹
217.0397211682	مناجات پانزدهم	284.6440637	دعا ۱۸
		296.2541396	دعا ۲۵
		307.3088368	دعا ۳۷
		326.0519959	دعا ۴۴
		346.0231257	دعا ۴۷
		292.7554397	دعا ۵۱
		257.0261102	دعا ۵۳

طبق جدول ۱۰، میانگین مقدار متغیر R برای دعا‌های صحیفه سجادیه، ۲۹۲.۴ و برای مناجات‌های خمس عشر، ۲۴۷.۷ بوده و تنها ۱۸٪ $(247.7 - 292.4) / 247.7 = 18\%$ اختلاف دارند؛ لذا در

جامعه آماری دوم نیز اختلاف قابل ملاحظه‌ای میان غنای واژگانی دعاها و مناجات‌های منتخب، وجود ندارد.

در نهایت، جهت بررسی وابستگی خروجی نظریه‌ها به طول متن، مقدار این متغیرهای محاسبه‌شده در دو جامعه آماری را با یکدیگر مقایسه می‌کنیم:

جدول ۱۱: مقادیر متغیرهای W و R در جامعه آماری اول و دوم

Table 11: The Values of W and R Variables in the First and Second Statistical Populations

مقدار متغیر R در جامعه آماری دوم	مقدار متغیر R در جامعه آماری اول	مقدار متغیر W در جامعه آماری دوم	مقدار متغیر W در جامعه آماری اول	تعداد کلمات جامعه آماری دوم
270.8838905	216.86	11.52543586	9.598	۳۷۷
274.1616513	204.672	12.08553512	9.0596	۱۶۵
284.6440637	208.61	11.59814965	9.06	۸۰
296.2541396	190.31	12.4189311	9.0008	۴۵۷
269.0443509	205.534	11.37813809	9.246	۲۷۰
307.3088368	244.68	13.14258825	9.916	۴۳۹
326.0519959	221.1699	13.881587	10.0945	۷۴۳
346.0231257	204.58	12.04322511	10.0945	۲۸۸۴
292.7554397	197.357	12.7500931	9.11993	۴۰۶
257.0261102	190.309	10.8974157	9.5206	۱۵۸
89 (34.6%)	30.87 (16.2%)	2.98 (27.5%)	1.09 (12.1%)	اختلاف حداکثر و حداقل مقدار

بر اساس داده‌های جدول ۱۱، علی‌رغم وجود تفاوت میان مقدار متغیر W و R در جامعه آماری اول و دوم، اولاً، مقدار این متغیرها متناسب با افزایش تعداد کلمات، افزایش نیافته است؛ بدین معنی که به‌عنوان مثال، تعداد کلمات دعا ۴۷ صحیفه سجاده، ۲۸۸۴ کلمه بوده، اما اختلاف مقدار متغیر R در آن با دعای ۱۸ که دارای ۸۰ کلمه است $\approx 21.6\%$ (۲۸۴.۶/۳۴۶-۲۸۴.۶) است که متناسب با میزان اختلاف کلمات نیست، زیرا تعداد کلمات دعای ۴۷، تقریباً ۳۶ برابر تعداد کلمات دعای ۱۸ است.

ثانیاً، طبق ردیف آخر جدول، اختلاف بیش‌ترین و کم‌ترین مقدار متغیر W ، $15.4\% = (12.1 - 27.5)$ و متغیر R نیز $18.4\% = (34.6 - 16.2)$ بیش‌تر از جامعه آماری اول بوده که بسیار ناچیز است و با در نظر گرفتن تفاوت موقعیت و شرایط موجود در زمان ایراد هر دعا و همچنین تفاوت نوع مخاطب و لزوم تطبیق سطح واژگان به‌کار گرفته شده با سطح درک او، قابل اغماض و چشم‌پوشی است؛ لذا مقادیر این متغیرها در هر دو جامعه آماری از یک ثبات نسبی برخوردار بوده و خروجی نظریه‌های استفاده شده در این تحقیق، وابسته به طول متن نیست.

۵. نتیجه

در سال‌های اخیر، موضوع شناسایی نویسنده، اهمیت زیادی یافته است، زیرا با پیشرفت علم، دیگر وجود متن‌های بدون سند و مدرک صحیح، پذیرفتنی نیست. از قرن ۱۹ میلادی، نخستین تلاش‌های پژوهشگران جهت دستیابی به روشی مطمئن برای تشخیص نویسنده واقعی یک اثر، آغاز شد. مسئله اصلی در تحقیقات پژوهشگران دهه‌های اخیر، بهره‌مندی از دقیق‌ترین نظریه‌ها و روش‌های محاسباتی است که بدون وابستگی به طول متن، قادر به تشخیص نویسنده باشند. در پژوهش حاضر، از ترکیب پنج نظریه مختلف برای محاسبه غنای واژگانی - که ادعا شده است مستقل از طول متن هستند و دقت محاسباتی خوبی دارند - استفاده کرده و با استفاده از نرم افزار اس.پی.اس.اس. و ترسیم نمودارهای رگرسیونی مربوطه، مقایسه‌ای میان غنای واژگانی مناجات خمس عشر و دعاهای صحیفه سجادیه، انجام دادیم.

نتایج نشان می‌دهند که اولاً، طبق جدول‌های ۱ و ۷، دقت محاسبات انجام‌شده، بسیار بالا بوده و لذا خروجی نظریه‌ها، قابل اعتماد و استناد است. ثانیاً، طبق جدول ۱۱، خروجی نظریه‌های به‌کار گرفته شده، وابسته به طول متن نبوده و متناسب با افزایش تعداد واژگان، افزایش چشم‌گیری نداشته است. ثالثاً، طبق جدول‌های ۴ و ۱۰، میان غنای واژگانی مناجات خمس عشر و دعاهای صحیفه سجادیه - در هر دو جامعه آماری - اختلاف چندانی وجود ندارد؛ بنابراین صحت انتساب مناجات خمس عشر به امام سجاد (ع) - از منظر تکنیک‌های سبک‌سنجی - اثبات می‌شود.

در پایان، تأکید می‌شود که جهت بررسی درستی یا نادرستی انتساب یک اثر به نویسنده‌ای مشخص، لازم و ضروری است که فاکتورهای متفاوتی در نظر گرفته شوند. از آنجا که یک مقاله، ظرفیت بررسی همه جوانب را ندارد، در این پژوهش به استفاده از جدیدترین و دقیق‌ترین

روش‌های سبک‌سنجی، بسنده شد و صدور حکم کلی و قطعی در مورد درستی این انتساب، به پژوهش‌های بیشتر واگذار می‌شود.

۶. پی‌نوشت‌ها

1. writing style
2. new rhetoric
3. reasoning
4. logic
5. rationality
6. stylistics
7. new linguistics
8. quantitative research
9. Mendelhall
10. word-length Distribution
11. lexical Method
12. syntactic or Grammatic Method
13. language-model Method
14. vocabulary richness
15. text length
16. Bozkurt
17. Oliveria
18. Howedi
19. Rabab'ah
20. Torruella & Capsada
21. statistical stylistics
22. linguistic markers
23. stylistic markers
24. linguistic stylistics
25. quantity
26. authorship attribution (AA)
27. plagiarism
28. text classification (TC)
29. topic identification
30. language identification
31. genre detection
32. stylometry
33. quantify
34. Syntactic parts
35. structural elements

36. paragraph length
37. Yule
38. honore
39. Brunet
40. Sichel
41. Simpson
42. linear regression
43. standardize coefficients beta

۷. منابع

- آقابزرگ الطهرانی، م. (۱۴۰۳ق). *الذريعة إلى تصانيف الشيعة*. بيروت: دار الأضواء.
- أميدوار، أ.، و أميد علی، أ. (۲۰۱۵ م). *دراسة أسلوبية في صحة نسبة الديوان المنسوب إلى الإمام علي(ع) على أساس معادلة يول*. *اللغة العربية وآدابها*، ۱۱(۱)، ۵۹-۸۱.
- بلیث، ه. (۱۹۸۹م). *البلاغة والأسلوبية*. ترجمة و تقديم و تعليق: محمد العمری. ط ۱. منشورات دراسات أساس فاس.
- چراغی و ش، ح.، خسروی، ک.، و لطفی، ا. (۱۳۹۲ ش). *سبک‌شناسی مناجات التائبین امام سجاد (ع) بر مبنای رویکرد ساختارگرایی*. *لسان مبین*، ۴(۱۱)، ۸۲-۱۰۱.
- الحر العاملی، م. (۱۳۷۹ ش). *الصحيفة السجادية الثانية*. تحقیق: فارس حسون کریم. قم: مؤسسة المعارف الإسلامية.
- زنگویی، س.، و نعمتی شمس‌آباد، ح. (۱۳۹۲ ش). *شناسایی نویسندگان پیام‌های الکترونیکی از طریق واکاوی نوع و سبک نگارش آن‌ها مبتنی بر روش‌های یادگیری ماشین*. *پربازش و مدیریت اطلاعات*، ۲۹(۲)، ۴۵۳-۴۷۶.
- الصدر، ح. (۱۳۵۴ ش). *نهاية الدراية*. تحقیق: ماجد الغرابوی. تهران: نشر مشعر.
- الصدر، ح. (۱۳۷۵ ش). *تأسيس الشيعة لعلوم الإسلام*. بی جا: اعلمی.
- العاملی، ح. (۱۳۶۰ ش). *وصول الأخبار إلى اصول الأخبار*. تحقیق: عبداللطیف الكوهکمری. قم: مجمع الذخائر الإسلامية.

- قمی، ع. (۱۳۷۹ش). *مفاتیح الجنان*. تهران: برهان.
- لطفی، ا. (۱۳۹۱ش). *سبک‌شناسی مناجات خمس عشر امام سجاد (ع)*. پایان‌نامه کارشناسی ارشد. استاد راهنما: حسین چراغی وش. دانشگاه لرستان.
- متقی‌زاده، ع.، حاجی‌خانی، ع.، و مدیری، س. (۱۴۰۱ش). تفاوت سبک نامه‌های نهج‌البلاغه از منظر «غنائی واژگانی» و «رویکرد انفعالی» (مطالعه موردی نامه‌های ۲۳، ۳۰ و ۷۳). *جستارهای زبانی*، ۱۳(۱)، ۲۳۵-۲۶۲.
- مجلسی، م. (۱۴۴۰ق). *بحار الأنوار الجامعة لدرر أخبار الأئمة الأطهار*. بیروت: دار إحياء التراث العربی.
- محقق داماد، م. (۱۴۰۶ق). *قواعد فقه*. تهران: مرکز نشر علوم اسلامی.
- محمدی، ع. (۱۳۸۷ش). *شرح اصول فقه*. قم: دار الفکر.
- مدیری، س. (۱۳۹۹ش). *المقارنة بين نهج البلاغة والصحيفة السجادية على أساس الأسلوبية الإحصائية*. رسالة أعدت لنيل شهادة الماجستير. الأستاذ المشرف: عیسی متقی‌زاده. جامعة تربیت مدرس.
- مرادی، م.، و بحرانی، م. (۱۳۹۴ش). تشخیص خودکار جنسیت نویسنده در متون فارسی. *پردازش علائم و داده‌ها*، ۴، ۸۳-۹۴.
- مسجدی، ه.، عادل، م.، امیریان، م.، و زارعیان، غ. (۱۴۰۰ش). نگرشی به «متن‌کاوی» در پژوهش‌های زبانی: رویکرد رایانشی در تحلیل متون. *جستارهای زبانی*، ۱۲(۶)، ۴۹۹-۵۳۱.
- مشهدی نوش‌آبادی، م. (۱۳۹۴ش). نخستین تبیین احوال و مقامات عرفانی در مناجات خمس عشره. *مطالعات عرفانی*، ۲۲، ۲۱۵-۲۴۰.
- مصلوح، س. (۱۹۹۲م). *الأسلوب، دراسة لغوية إحصائية*. ط ۳. القاهرة: عالم الكتب.
- المنتظری، ح. (۱۴۱۶ق). *البدر الزاهر فی صلاة الجمعة والمسافر*. قم: مکتب آیه الله العظمی المنتظری.

- النجفی الجواهری، م. (۱۳۶۲ش). *جواهر الکلام فی شرح شرائع الإسلام*. تحقیق: عباس قوچانی. بیروت: دار إحياء التراث العربی.
- وطن‌خواه، ر، خاتمی، ا، و غلامرضایی، م. (۱۴۰۰ش). تبیین رویکرد نظری سبک‌شناختی ملک الشعرای بهار. *جستارهای زبانی*، ۱۲(۲)، ۴۷۷-۴۲۱.
- یاحقی، م، و ایزانلو، ع. (۱۳۸۵ش). سبک‌سنجی، نقد و بررسی شیوه آماري کیوسام در انتساب یک اثر. *مجله دانشکده ادبیات و علوم انسانی دانشگاه خوارزمی*، ۵۳، ۱۹۰-۱۵۱.

References

- Agha Bozorg Tehrani, M. (1982). *Az-Zaree'a*. Eeirut: Dar Al-Azvae. [In Arabic]
- Ameli, H. (1982.). *Good people access to the assets of the news*. Investigated by: Abdul Latif Al-Kokhamari. Qum: Islamic Relics Complex. [In Arabic]
- Al-Montazeri, H. (1995). *The full moon in the Friday prayer and the traveler*. Qum: The Office of the Great Ayatollah al-Montazeri. [In Arabic]
- Al-Sadr, H. (1996). *Shi'a founding the sciences of Islam*. Tehran: A'alami. [In Arabic]
- Bagirzade, Z.M. (2019). Rhetoric, linguistics and stylistics. *Russian Linguistic Bulletin*, 3(19). 41-43.
- Biber, D. (1995). *Dimensions of register variation: A cross-linguistic compariso*, Cambridge University Press.
- Bleeth, H. (1989). *Rhetoric and style*, translated by Mohammad Al-Omari. Publications on the principles of Fas. [In Arabic]
- Bozkurt, I.N., Baghioglu, O., & Uyar, E. (2007). Authorship Attribution, performance of various features and classification methods. *Computer and information sciences*, 1-5.
- Brunet, E. (1978). *Vocabulaire de Jean Giraudoux: Structure et Evolution*. Slatkine.

- Cheraghivash, H., Khosravy, K., & Lotfy, E. (2014). Stylistics of Imam Sajjad's Monajat Altaebeen in the light of structuralism. *Lisan-I Mubin*, 4(11), 82–101. [In Persian]
- Escobedo, F., Cruz, C., Sierra, G., & Soto, J. (2013). Analysis of stylometric variables in long and short texts. *Procedia-Social and Behavioral Sciences* 95, 604–611.
- Farahmandpoor, Z., Nikmehr, H., Mansoorizade, M., & abibzadeh Ghamsary, O. (2013). A novel intelligent Persian authorship system based on writing style. *Soft Computing Journal*, 1(2), 26–35. [In Persian]
- Fowler, R. (1987). *A Dictionary of Modern Critical Terms*. Routledge and Kegan Paul Ltd.
- Gharechapogh, F., Motamanfar, M., & Vafadar, M. (2017). A new way to identify the author of texts by combining particle mass optimization algorithms and backup vector machines. *Computing Science Journal (CSJ)*. Num, (12), 2–12. [In Persian]
- Honore, A. (1979). Some simple measures of richness of vocabulary. *Association for Literary and Linguistic Computing Bulletin*, 7(2), 172–177.
- Howedi, F., Mohd, M., Aborawi, Z., & A.Jowan, S. (2020). Authorship attribution of short historical arabic texts using stylometric features and a KNN classifier with limited training data. *Journal of Computer Science*, 16(10), 1334–1345.
- Howedi, F., & Mohd, M. (2014). Text classification for authorship attribution using Naïve Bayes classifier with limited training data. *Computer Engineering and Intelligent Systems*, 5(4), 48–56.
- Al-Hur Al-Amili, M. (2000.). *The second Sahifah Sajjadiah*. Investigation: Faris Hassoun Karim. Qum: Islamic Knowledge Foundation. [In Arabic]

- Lotfi, A. (2010). *Stylistics of the prays of Khams Ashar Imam Sajjad*. [Master Thesis, University of Lorestan]. Lorestan [In Persian]
- Majlesi, M. (2019). *Behar Al-Anwar*. Arab Heritage Revival House. [In Arabic]
- Mashhadi nushabadi, M. (2015). The first explanation of mystical Ahwāl and Maqāmāt in Khamsa ‘Ashara Monājāt. *Mysticism Studies*, 1(22), 215–240. [In Persian]
- Masjedy, H., Adel, M., Amirian, M., & Zareian, G. (022). An overview of text mining in language studies: The computational approach to text analytics. *Language Related Research*, 12(6), 499–531. [In Persian]
- Maslouh, S. (1992). *Style, statistical linguistic study*. The world of the books. [In Arabic].
- Mendelhal, T.C. (1887). *The characteristic curves of composition*. *Science*, 9(214), 237–246.
- Modiri, S. (2018). *Comparison between Nahj al-Balagha and Al-Sahifa al-Sajjadiyya on the basis of statistical stylistics*. [Master thesis, Tarbiat Modares University]. Tehran [In Arabic]
- Mohaghegh damad, M. (1985). *Rules of jurisprudence*. Islamic Sciences Publishing Center. [In Persian]
- Mohammadi, A. (2008). *Description of the principles of jurisprudence*. Dar Al-fekr. [In Persian]
- Moradi, M., & Bahrani, M. (2013). Automatic gender identification in persian text. *Signal and Data Processing*, (4), 83–94. [In Persian]
- Motaghizadeh, I., Hajikhani, A., & Modiri, S. (2022). Differences between the style of the Nahj al Balagha's letters in terms of “Vocabulary Richness” and “literary style” (Letters 23, 30 and 73 as case study). *Language Related Research*, 13(1), 235–262. [In Persian]

- Najafi Al-Jawaheri, M. (1983). *The jewels of speech in explaining the laws of Islam*. Investigate: Abbas Qoshani. Beirut: Arab Heritage Revival House. [In Arabic]
- Oliver, Jr. W., Justino, E., & Olivera, L.S. (2013). Comparing compression models for authorship attribution. *Forensic Science International*, 5(4), 100-104.
- Omidvar, A., & Omidali, A. (2015). Stylistics research on correctness of relation of the poem related to Imam Ali (PBUH) based on Yule's equation. *Language and Literature Arabic*. 11(1), 81-59. [In Arabic]
- Perelman, Ch. (1971). *The new rhetoric*. Reidel publishing company.
- Qomi, A. (2000). *Heaven keys*. Borhan. [In Persian]
- Rabab'ah, A., Al-Ayyoub, M., Jararweh, Y., & Aldwairi, M. (2016). Authorship attribution of Arabic tweets. *13th Computer Systems and Applications*. Agadir, Morocco. 1-6.
- Sadr, H. (1975). *End of know-how*. Investigate: Majed Al-Gharabawi. Hairy Publishing. [In Arabic]
- Sichel, H. (1975). On a distribution law for word frequencies. *Journal of the American Statistical Association*, (70), 542-547.
- Simpson, E. (1949). Measurement of Diversity. *Nature*, 16(3), 650-688.
- Stamatatos, E. (2009). A survey of modern authorship attribution methods. *Journal of the American Society for Information Science and Technology*, 60(3), 538-556.
- Tanaka-Ishii, K., & Aihara, Sh. (2015). Computational constancy measures of textx- Yule's K and Renyi's entropy. *Computational Linguistics*, 41(3), 481-502.
- Torruella, J., & Capsada, R. (2013). Lexical statistics and tipological structures: A measure of lexical richness. *5th Conference on Corpus Linguistics*. 447-454.
- Vatankhah, R., Khatami, A., & Gholamrezaei, M. (2021). Explaining the theoretical approach of stylistic features of Malek osh-Sho'arā Bahar. *Language Related Research*, 12(2), 421-447. [In Persian]

- Wright, D. (2014). *Stylistics versus statistics: A corpus linguistic approach to combining techniques in forensic authorship analysis using Enron emails*. The University of Leeds. School of English.
- Yahaghi, M., & Izanlu, A. (2006). Stylometric, critique of Qiosam's statistical method in assigning a work. *Journal of the Faculty of Literature and Humanities*, 14(53), 151–190. [In Persian]
- Yule, G. (1944). *The statistical study of literary vocabulary*. Cambridge at the University Press, University Printing House.
- Zangoei, S., & Nemati Shamsabad, H. (2014). Identify the authors of electronic messages through the analysis of the type and style based on machine learning technique. *Information Processing and Management*, 29(2), 453–476. [In Persian]
- Zhao, Y., & Zobel, J. (2005). Effective and scalable authorship attribution using function words. *Information Retrieval Technology*, 3(2), 174-189.