

The Pārsīg Database

Hossein Mostafavi Kashani¹

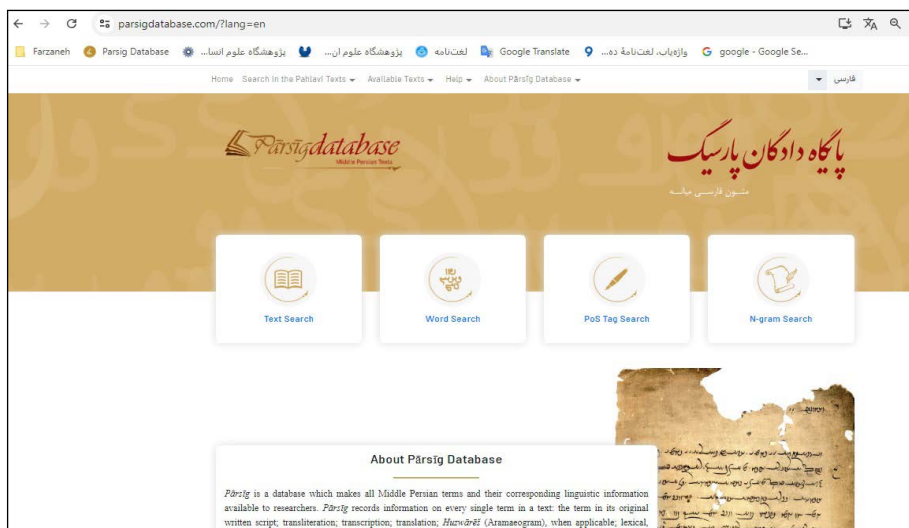
Received: 21 March 2023 / Accepted: 20 June 2023

Middle Persian, referred to as “Pārsīg” in its writings, is a descendant of Old Persian and serves as the ancestor of Dari Persian. This language was prevalent from around the fourth century BCE (the fall of the Achaemenid Empire) to the ninth century CE (the third century AH). Zoroastrians, Manichaeans, and Christians have produced works in this language, and the Sasanian kings have inscribed tablets. Understanding this language is essential for comprehending the intellectual heritage of Iran during the Sasanian era and the early Islamic period, as well as for strengthening the Dari Persian language. Several dictionaries have been compiled for this language, including notable works such as *A Manual of Pahlavi* by Samuel Nyberg (1964-1974), *A Concise Pahlavi Dictionary* by MacKenzie (1971), and *Pahlavi Dictionary* by Bahram Farahvashi (1967). None of these dictionaries encompass all the words of the Pahlavi language. In five volumes, the *Pahlavi Dictionary* by Yadollah Mansouri (2015-2021) includes all the words of this language found in Middle Persian texts.

The printed dictionaries of the Middle Persian language do not encompass the translation of all texts, the complete frequency of words, and the mention of each example along with its translation. The possibility of providing individual word examples accompanied by their translation and linguistic analysis is rare in printed dictionaries. Providing the ability to search for all the words of a language with examples, translations, and linguistic analysis is more feasible using computers.

To compile a comprehensive corpus of Middle Persian, it is necessary first to read all of its texts, transcribe, transliterate, and translate them. The frequency of each word should be accompanied by all its available evidence and the translation of these pieces of evidence, and each word should be linguistically analyzed.

1. Alumnus of the University of Tehran, Department of the Ancient Iranian Culture and Languages, Tehran, Iran.
Email: hmk1368@gmail.com.



The ultimate goal of Pārsīg database is to have all Middle Persian texts translated and their words examined at six levels: transcription, transliteration, translation, determination of their syntactic category, morphological analysis, and determination of their lexical status. This includes textual references, references to existing manuscripts, and the provision of a critical edition of each text. For the first time, Middle Persian texts are extensively studied and analyzed at this level. Online dictionaries of the Middle Persian language also do not have such extensive coverage. “Phalavazh” is the first online dictionary of the Middle Persian language that has been developed solely based on nine Middle Persian texts.² It only includes transliteration, Persian meanings, and references to Pahlavi texts. Furthermore, there is no possibility for its expansion. The online dictionary “Pahlavica” is also only comprised of transcription, transliteration, English and Persian meanings, word etymology, and an example of each word’s usage provided under its entry.³ It does not include a linguistic analysis of the vocabulary and is not based on a database of all Middle Persian texts.

In the Pahlavi section, the Titus database only provides transliteration and occasionally transcription of texts without their translation.⁴ The transliteration and word meanings in Titus are based on the MacKenzie dictionary. Other projects have also been proposed for the development of an online Middle Persian dictionary, such as “the Middle Persian Dictionary Project” led by Shaul Shaked and Carlo G. Cereti (2005), and *The Dictionary of Manichaean Texts* at the School of Oriental and African Studies (SOAS), which have not yet been presented as a database.⁵

2. <http://pahlavazh.ihcs.ac.ir>

3. <http://pahlavica.org>

4. <http://titus.uni-frankfurt.de>

5. <http://www.soas.ac.uk/nme/research/manichaeandict>

The Pārsīg database has so far presented three categories of Middle Persian texts: Zoroastrian Middle Persian texts, Zand texts, and Sasanian and post-Sasanian private inscriptions. The database offers four search options:

1. Text Search: In this search, uploaded texts, including the Pahlavi script, transliteration, and Persian translation, are provided along with references to Pahlavi texts. Some texts also include English or French translations. Clicking on each word reveals its specifications, including Pahlavi spelling, transcription, transliteration, meaning, syntactic category, morphological analysis, references to Pahlavi texts, and critical editions. A critical edition image is also available for each text.

Text Search
Transcription and Persian translation of Middle Persian texts

Text Type: Sasanian and Post-Sasanian private inscriptions
Type of Inscription: Darband-Caucasus Inscr
Inscription Name: Chapter 1

Transcription and Translation | Edited Text

کتابه اسفند

Search

Home Search in the Pahlavi Texts Available Texts Help About Pārsīg Database فارسی

Text Search
Transcription and Persian translation of Middle Persian texts

Text Type: Zoroastrian Middle Persian texts
Text Name: abdñ ud sahīgññ f sīstān
Chapter: All

Transcription and Translation | Edited Text

All

Search

Double-click on Pahlavi or Transcription to view tags for each word

0- pad nām f yazdān. [abdñ ud sahīgññ f sīstān]. (Jamasp-Asana, 1913, p.25)

به نام بردن، شکفتن و برهمنگی سپستان. (کهنسب و جانیپور، 1398)

2. Word Search: In this section, it is possible to search for a word based on its Pahlavi spelling, transliteration, transcription, Persian meaning, or its root. The search provides all the occurrences of the word in the available texts, and in addition to the word's specifications (as mentioned in the previous section), the translations of the examples are also provided. This allows the searcher to access different usages of the word in various contexts and texts.

Word Search

Vocabulary List - Variety and Frequency

Text Type
Zoroastrian Middle Persian texts

Text Type
All

Chapter

Pahlavi
Transcription
Transliteration
Meaning (Persian)
Lemma

↵

Click on the numbers to view tags for each word

۱۴۳۱

handarz T ādurbād T māraspandān : 35 46 57 102 103 139
 handarz T dānār T dānāg : 53 53
 mēnōg T xrad : 1/35 1/61
 zand T wahman yasn : 9/5

3. Syntactic Tag Search: This section includes twelve predetermined syntactic categories: noun, adjective, determiner, verb, adverb, pronoun, preposition, postposition, conjunction, *Ezafē*, number, and verbal particle. Each category has its own subcategories. As an example, we can mention the verb category, which includes features such as polarity (positive and negative), type (simple, causative, and reflexive), transitivity (intransitive, transitive, and passive), number (singular and plural), person (first person, second person, third person), tense (present, past, future), aspect (perfect, imperfect, optative, conditional), mood (subjunctive, imperative), impersonal (modal, infinitive, participle of necessity), and clitic.

PoS tag search

Grammatical category of Middle Persian words

Text Type
Zoroastrian Middle Persian texts

Text Name
All

Chapter

Grammatical Category

Proper. Singular. Positive. God

Click on the numbers to view tags for each word

*ahlišwag

wizdagihā T zādisparam : 35/40 35/40

Thus, examples of each syntactic category are provided in the available texts within the database, and the corresponding translations for each example are given in Persian. With the final completion of the corpus, it will become feasible to compile a comprehensive Middle Persian grammar, as there will be examples, translations, and references to the existing texts for each syntactic category and its subcategories. Determining the syntactic category of words is based on the collection of labels for contemporary Persian syntactic categories, authored by Bijankhan et al (2011) and structured by Ghayoomi (2014). However, some modifications have been made to the framework due to the differences between Middle Persian and contemporary Persian. For example, in contemporary Persian, adjectives are

always singular in number, but in Middle Persian, an adjective can be either singular or plural. Therefore, the capability of determining the number has been added to the search features to search within the adjective category.

4. Collocation Search: In this section, in addition to the main search word, there is the possibility of adding up to three preceding words and three following words. Thus, the search can include phrases with a maximum of seven words. This searchability feature can be helpful in retrieving repeated expressions in Pahlavi texts, including idiomatic and formulaic expressions.

In the Zoroastrian Middle Persian text section, forty texts have been uploaded so far. Some of them are fully presented, while others still need to be completed, and the completion of the website's data will continue until all the Middle Persian texts are uploaded. For some of the texts, previously available translations have been used, and some of the Middle Persian texts have been newly translated by Farzaneh Goshtasb, the founder and manager of the platform, and Nadia Hajipour. In most cases, it seems that translations have been set aside to preserve Pahlavi words as much as possible in the Persian translation without considering whether their meanings are still retained in Persian. In these translations, efforts have been made to keep the Middle Persian text untouched as much as possible and transfer Pahlavi words and sentences into Persian translation. However, alternative translations on the website have made efforts to translate Pahlavi words into words that convey their meanings and are also understandable in present-day Persian rather than simply transcribing Pahlavi words into Persian. Additionally, in sections where previous translations of the texts have been used, the translations have been critically reviewed, and in many cases, alternative translations have been suggested, which serve as a guide for better understanding the texts. For example, notes have been made for the text *Wizīdagīhā ī Zādisparam*, including notes related to Introduction: 1; 1: 3, 10, 15, 16, 28, 29. For some texts, existing European translations are also included, allowing the reader to compare the Persian and European translations.

In the section of the Zand texts, translations of sections of the Zand of Yasna have been uploaded so far. These texts have been translated into Persian for the first time by Shole Hesarpouladi and the present researcher. The third section, which relates to the Sasanian and post-Sasanian inscriptions, is translated by Cyrus Nasrollahzadeh. The Middle Persian texts of the Manichaean Manuscripts are also translated based on the selected texts of the Manichaean Manuscripts by Boyce (1975), to be uploaded soon. Regarding these texts, in addition to transliteration and Persian translation, by clicking on each word, information including Manichaean spelling, calligraphy, transliteration, meaning, syntactic category, word derivation, reference to Boyce's *Reader*, and reference to the available version will be available.

The achievements of realizing the ultimate goal of the Pārsīg database, namely the upload of all Middle Persian texts as described, are as follows:

1. Access to the translation of all Middle Persian texts.
2. Access to all Middle Persian language words, along with Pahlavi spelling, calligraphy,

transliteration, meaning, syntactic category, and all their attestations with Persian translation and, in some cases, their translation into European languages.

3. Access to all instances of each syntactic category of the Pahlavi language, along with its subcategories, thereby enabling the compilation of a comprehensive Pahlavi grammar with all the examples and frequencies of each (sub-)category.
4. Access to the usage and frequency of phrases (up to seven words) in Middle Persian texts with the possibility of identifying formulaic expressions.

The function of this database should not be limited to extracting statistical information. It is to be noted that to get to know all the meanings of a Middle Persian word (technical and idiomatic meanings included), one needs to analyze the word in all the different contexts in which it occurs. Access to different and accurate translations of Middle Persian texts will be handy to gain such an understanding. Accurate grammatical and historical syntax analyses of Persian (in all three periods: Old, Middle, and Modern) are achievable only when instances of all grammatical (sub-)categories are available. The Pārsīg database moves towards achieving these objectives through text translations on one hand and the manual tagging of word-by-word grammatical categories on the other. The project members manually perform calligraphy, transliteration, and text translation, while the responsibility for grammatical tagging of all texts lies with the database's founder.

The information in this database can be utilized in producing historical dictionaries, various Middle and Modern New Persian language dictionaries, scientific terminology selection, semantics, phonetics, syntactic analysis, and studying the historical evolution of Persian vocabulary (Goshtasb and Hajipour Artarani 2021, 264).

References

- Bijankhan, Mahmood, Javad Sheykhzadegan, Mohammad Bahrani, and Masood Ghayoomi. 2011. "Lessons from building a Persian written corpus: Peykare." *Language Resources and Evaluation* 45, no. 2: 143-64.
- Boyce, Mary. 1975. *A Reader in Manichaean Middle Persian and Parthian*. Tehran: Liège.
- Faravashi, Bahram. 1967. *Pahlavi Dictionary*. Tehran.
- Ghayoomi, Masood. 2014. "From HPSG-based Persian Treebanking to Parsing: Machine Learning for Data Annotation." PhD diss., Freie Universität Berlin, Germany.
- Goshtasb, Farzaneh, Masood Ghayoomi, and Nadia Hajipour Artarani. 2021. "Goshtasb, Farzaneh, Masood Ghayoomi, and Nadia Hajipour Artarani." *Zabānshenāxt* 12, no. 1: 255-80.
- MacKenzie, David Neil. 1971. *A Concise Pahlavi Dictionary*. London: Routledge.
- Mansouri, Yadollah. 2015-2021. *Pahlavi Dictionary*. 5 vols. Tehran: Shahid Beheshti University Press.
- Nyberg, Henrik Samuel. 1964-1974. *A Manual of Pahlavi*. 2 vols. (1964, vol.1; 1974, vol.2). Wiesbaden: Harrassowitz.
- Shaked, Shaul and Carlo Giovanni Cereti, 2005. "A Middle Persian Dictionary: Project Proposal." *Orientalia Romana 8: Middle Iranian Lexicography* (Proceedings of the Conference held in Rome, 9-11 April 2001), edited by Carlo Giovanni Cereti and Mauro Maggi, 181-90.