
LANGUAGE IN INDIA

Strength for Today and Bright Hope for Tomorrow

Volume 14:3 March 2014
ISSN 1930-2940

Managing Editor: M. S. Thirumalai, Ph.D.
Editors: B. Mallikarjun, Ph.D.
Sam Mohanlal, Ph.D.
B. A. Sharada, Ph.D.
A. R. Fatihi, Ph.D.
Lakhan Gusain, Ph.D.
Jennifer Marie Bayer, Ph.D.
S. M. Ravichandran, Ph.D.
G. Baskaran, Ph.D.
L. Ramamoorthy, Ph.D.
C. Subburaman, Ph.D. (Economics)
Assistant Managing Editor: Swarna Thirumalai, M.A.

Corpus in Persian Lexicography

Zahra Zare and M. Balakumar

Abstract

In linguistics studies, the corpus is a collection of written or oral texts that can be used as the base for analysis and description of language. With the help of corpus, we can study the phonetic, syntactic, social and other aspects of languages. Linguists emphasize the importance of language corpus in their researches. In this article, at first we discuss the resources and component elements of corpus including the written and oral resources of the corpus in Persian lexicons in the Indian subcontinent and cotemporary lexicons of Iran. A sample of the corpus in the following Dictionaries of Persian, viz., Dekhoda, Moein and Sokhan lexicons and some examples of computer corpus of Persian language are also given. The non-electronic corpora created prior to computer era are reviewed. Electronic corpora are created with the help of computer which are easy to access for analysis.

Language in India www.languageinindia.com ISSN 1930-2940 14:3 March 2014

Zahra Zare and M. Balakumar
Corpus in Persian Lexicography

Key words: corpus, language lexicons, Persian Lexicography, computer

Introduction

Meyer (.2002), said: “A corpus is a collection of texts or part of texts upon which some general linguistics analysis can be collected “.

It is applied to a collection of written or spoken language data that could be employed to describe and analyze a language. Undoubtedly, any research on languages and their description and analysis, including lexicography, grammar compilation, dialectology, etc. will not yield useful results without relying on corpora. (Khatibi, 2007, p.14).

Many linguists have emphasized the importance of corpus in most of their researches in the past. However, relying on real linguistic data has become common in the modern era and has turned into one of the essential requirements of many theoretical and applicable studies such as studies on dialectology, grammar compilation, and quick examination of linguistic data. Thus, a specialized branch in the field of computational linguistics has emerged. This new branch was created under the name of corpus linguistics in the last decades of the twentieth century and became one of the most active and applicable branches of linguistics within a short period of time. The term "corpus" is especially used by structural linguists, who always emphasize that the description of a language or dialect should be based on collecting and analyzing data (Assi, 2000).

The language corpus can be of any size depending upon the purpose. It can be big or very big, inclusive and represent the whole of a language or kind of that language, in the form of notes or computer files which include complete texts or the chosen parts of them or continuous parts of texts or a select part of quotations and notations and list of vocabularies. The language corpus can be specially provided for a study or can include a big and formless collection of various texts which are used for different purposes. (Assi, 2000).

Resources (Components)

Language in India www.languageinindia.com ISSN 1930-2940 14:3 March 2014

Zahra Zare and M. Balakumar

Corpus in Persian Lexicography

Corpus is the most important factor contributing to compiling a comprehensive monolingual general dictionary. The components of a corpus are as follows:

- 1- Written documents: (from the oldest texts to the era of compiling dictionaries, including books and pamphlets on literature, religion, and history, legal and judicial documents, agreements, letters, stories, itineraries, magazines, newspapers, classified advertisements, etc.
- 2- Spoken documents: including telephone conversations, daily conversations, radio and TV debates, face-to-face interviews, classroom discussions, personal quotations, etc.)
- 3- Documents obtained through field research: (These documents are especially useful for compiling terms used in different professions and jobs.)
- 4- Linguistic knowledge of lexicographer.

Lexicographers are also a source of information which is to be used in dictionaries. The linguistic knowledge of lexicographers and their lexicographical skill guarantees the success of dictionary projects. (Hashemi-Minabad, 2007, p.111).

Lexicographers with critical and analytical insight are capable of removing shortages in extracted data and correcting flaws observed in such data. They can compile good dictionaries thanks to their knowledge about written language and spoken language as well as stylistic, contextual, semantic and grammatical nuances.

According to Hashemi Minabad (2007), Corpus can be drawn from to following sources:

1-Written Corpus

Written corpus is regarded as the most important linguistic corpus. Even dictionaries that are claimed to emphasize spoken language greatly benefit from written corpora. For example, Ahmad Shamlu who has allocated his "*Street Dictionary*" to vulgar language has quoted written documents in many cases. One of the main reasons behind the great emphasis on written documents is the availability of such documents compared to spoken documents. The holiness and importance of written language is another reason.

Written corpora are of different types which are used depending on the type and audience of the dictionary. Written documents could include textbooks, fiction works, non-fiction works, newspapers, publications, previous dictionaries, and reference books.

2-Textbooks

High-school textbooks are among the main sources of the corpus of a general dictionary. These textbooks enjoy great vocabulary diversity, cover many basic topics, and almost include all important fields. Some terms found in such books are too specialized to be used in general language. For example, some very specialized terms found in the "*carpet-weaving*" textbook of the first grade of high school include "Shakheh-Pich Toranjdar" (twisted branch with citrons), "Talfighi Toranji Dasteh Gol" (combination of citrons and flowers), "Talfighi Toranji Kaf Sadeh" (combination of citrons and simple backgrounds), "Talfighi Lachak Toranj Kaf Sadeh" (combination of citrons, scarves and simple backgrounds).

It is obvious that textbooks contain many words used in different professions and fields that should be extracted. So, the language used to describe and define words in dictionaries should be closely related to the language of textbooks. In addition, textbooks could also contribute to the language and expression of dictionaries. The corpus of general dictionaries does not cover university textbooks since they are too specialized. However, general textbooks such as "*General Persian*", "*The History of Islam*", "*Islamic Insight*", and "*Physical Education*" could be covered by the corpus of general dictionaries.

3-Fiction Works

Novels, stories, short stories and plays make up a major part of written corpus since fiction reflects a wide range of human experience and are widely read.

Literary texts and classic works often make up the main source of Persian dictionaries. Therefore, the end product does not properly describe the contemporary spoken language.

Fiction works have different style and types, covering various topics. Historical, science-fiction, adventure, police, entertainment and general stories are proportionally included in written corpora.

4-Non-Fiction Works

More attention is paid to non-fiction works today. Non-fiction works include books on general, scientific and technical topics, itineraries, biographies, guide books, and collections of articles. These works cover different fields such as arts, civilization, economics, history, politics, psychology, sports, nutrition, health, housekeeping, cooking, animal life, environmental health, industry, education, religion, cinema, etc.

5-Newspapers and Publications

Newspapers and periodical publications contain many modern and contemporary linguistic elements, reflecting the latest linguistic and non-linguistic developments. Newly-coined words, definitions of new words and existing terms, and colloquial elements are widely found in such texts. An advantage of newspapers is that they cover various topics and their language is a current daily language.

6-Letters and Catalogues

Formal, administrative and personal letters have their own specific words that should be recorded by general dictionaries. Catalogues, brochures and classified advertisements often contain numerous and various elements of the contemporary and daily language which are rarely found in other texts.

7-Previous Dictionaries

Previous general dictionaries are one of the important sources of data. Sometimes, a dictionary is compiled based on another one. Sometimes, a concise dictionary is compiled based on a comprehensive one or vice versa. Quoting previous dictionaries to write a new one has a long history. Data provided by other dictionaries could be employed to review data collected for the new dictionary and make a comparison. However, this should not be done without doing required research.

8-Reference Books

Reference books, encyclopedias, and dictionaries on language difficulties, such as *let's write correctly* by Abolhasan Najafi, are among the resources that are used to describe and define words and extract data. (p, 105-111),

In the past, preparing linguistic corpora to compile a dictionary was a time-consuming, laborious and costly task. For example, Allameh Dehkhoda studied Persian texts in verse and prose for many years, preparing a trove of notes containing words, evidence, and references. The collection of linguistic data which Dehkhoda prepared through the traditional note-taking method has been unique among Old Persian texts and even contemporary texts in terms of size and precision. (Khatibi, 2007,p.14).

Corpus in Persian Dictionaries in the Indian Subcontinent (13th – 19th century)

Lexicography became significantly popular in the Indian subcontinent as of the twelfth century, after Islamic dynasties such as Ghurid and Mogul came to power in the subcontinent and Persian language was spoken in the Indian court. The *Ghavas Dictionary* by Fakhroddin Mobarakshah Ghaznavi was the first dictionary compiled in that era. The *Loghat-e Fors* dictionary, rather than texts, made up the main corpus of this dictionary. (Siaghi Dabir,1989,P.33).

It became necessary to compile more complete Persian dictionaries in the Indian subcontinent after great Persian-speaking scholars such as Amir-Khosro Dehlavi, Sa'eb Tabrizi, and Bidel Dehlavi emerged there.

The *Jahangiri Dictionary* was written by Mir Jamaloddin Hossein Ebn-e Fakhroddin Hasan Anju Shirazi during Mogul King Jahangir. It was regarded as an important evolution in Persian lexicography. The corpus of this dictionary included:

- 1- Previous dictionaries, including 53 dictionaries compiled to that date.
- 2- Texts in verse. Only one text in prose (*The Naser Khosrow's Itinerary*) has reportedly been among the corpus of this dictionary.
- 3- Pahlavi words. The author has recorded and defined a number of Pahlavi words, known as "Zand" and "Pazand" words among lexicographers, in his dictionary with the help of the minority Zoroastrian population living in India, known as "Persians".

Language in India www.languageinindia.com ISSN 1930-2940 14:3 March 2014

Zahra Zare and M. Balakumar

Corpus in Persian Lexicography

- 4- Other references such as *Zakhireye Khwarazmshahi*, *Nozhat ol Gholub*, and *Ajayeb ol Boldan* (Khatibi, 2007, p.21).

A new chapter began in Persian lexicography, and the corpus of Persian dictionaries and words increased following the compilation of the *Borhan-e Ghatte Dictionary* by Mohammad Ebn-e Hossein Khalaf Tabrizi. The corpus of this dictionary included:

- 1- Previous dictionaries and poetical works
- 2- Arabic dictionaries
- 3- Pahlavi words known as "Zand" and "Pazand"
- 4- Dasatiri words, namely words coined by Azar Kayan in a book named *Dasatir-e Asemani* during the reign of the Mogul emperor Akbar and attributed to a false prophet named Sasan the Fifth. (Khatibi, 2007,p.20&21).
- 5-Compilation of dictionaries continued after the Mogul empire collapsed and Britain dominated over India. Among the most important dictionaries was *The Anenderaj*, which used a relatively large number of Arabic dictionaries as its corpus. . (Siaghi Dabir,1989, p.123)

The last important Persian dictionary authored in India was *The Nezam Dictionary* by Mohammad Ali Daeoleslam which has been compiled based on a new method. The corpus of this dictionary was essentially different from those of previous dictionaries:

- 1- The corpus of the dictionary included not only complex poetical words, but also spoken words.
- 2- It paid attention to texts in both verse and prose.
- 3- "Zand" and "Pazand" words have been omitted.
- 4- Words and their spellings have been collected through referring to people.
- 5- Tehrani dialect has been employed to record words.(Daeoleslam, 1983,P.2).

Corpus in Contemporary Lexicography in Iran

As of the sixteenth century, dictionaries were also compiled in Iran from time to time concurrent with the compilation of Persian dictionaries in India. All of the dictionaries authored in Iran were influenced by those written in India. *The Majma el-Fors-e Soruri* was one of the

oldest and most important dictionaries authored in Iran during that period. The dictionary, compiled during the reign of Shah Abbas I, was totally based on *the Jahangiri Dictionary* .

In the contemporary era and concurrent with the evolution of Persian language, basic changes were also observed in the corpus of Persian dictionaries. For example, the corpus of *The Nafisi Dictionary* by Nazem ol-Atebba Nafisi included European words - e.g. words from Persian-English dictionaries by Richardson, Wollaston, Johnson, and Steingass - in addition to previous dictionaries both in Persian and Arabic (Khatibi, 2007,p.24&25).

Corpus in the Dekhoda Dictionary

The Dekhoda Dictionary is regarded as the last Persian dictionary written based on traditional methods. It is also a collection of all the dictionaries compiled previously and covers almost all the words found in important handwritten and printed Persian and Arabic dictionaries. In addition, it contains thousands of sarcastic words and phrases as well as examples from prose and verse texts and other scientific and literary books none of which could be found in other Persian and Arabic dictionaries. *The Dekhoda Dictionary* also covers many Turkish, Mogul, Indian, French, English, German and Russian words which were used in Persian but had not been included in dictionaries compiled before. (Moein, 1996, p.394).

The corpus of the *Dekhoda Dictionary* was larger and more diverse than all the corpora used by previous and later dictionaries. The components of that corpus could be categorized as follows:

- 1- Most of Persian and Arabic dictionaries compiled to that date.
- 2- Persian texts in verse; for the first time, Dekhoda read Ferdowsi's *Shahnameh* and other collections of poems, extracting words and evidence from them.
- 3- Persian texts in prose.
- 4- Widely-used Persian words; in many cases, Dekhoda would approach people to collect words and their meanings.
- 5- Historical and geographical announcements related to Iran, Muslim countries, and some non-Muslim countries.

Besides, Dehkhoda explained scientific, literary, grammatical and philosophic issues as well as viewpoints and beliefs of sects and ideologies, adding aspects of an encyclopedia to his dictionary. The publication date of this dictionary (1945-1980) is an important period in the history of Persian lexicography since it has been used as the main corpus for all the dictionaries that were authored later. (Khatibi, 2007, p.26)

Corpus in the Moein Dictionary

The publication of *The Moein Dictionary* was another important turning-point in the history of Persian lexicography, indicating great progress in different fields and opening a new path for later lexicographers. Some features of *The Moein Dictionary* include: reasonable order of entries, sub-entries, definitions, and examples; Latin transcription; scientific etymology; and separation of proper nouns from words.

Domination of Moein over the legacy of Persian language, especially Persian dictionaries, and employment of modern lexicography methods has made his dictionary one of the main reference books used by speakers of Persian language.

The elements of the corpus used in *The Moein Dictionary* could be categorized as follows:

- 1) Persian texts from the beginning until the author's lifetime, including over 343 books and pamphlets.
- 2) Persian, Arabic and European dictionaries.
- 3) Arabic, Turkish and European encyclopedias such as *Mu'jam al-Buldan* by Yaqut, *Qamus al-A'alam* (Turkish), *Kashf al-zunūn* by Haji Khalifa, *Atlas of Iranian Geography*, *Encyclopedia of Iranian Geography*, *Encyclopedia Britannica*, *Brockhaus Enzyklopädie*, *Grand Larousse encyclopédique*, and *the Dehkhoda Dictionary*, which has been classified as an encyclopedia by Moein.
- 4) Words and names approved by the Persian Language & Literature House of Culture.
- 5) Latest works on the etymology of Persian words by Iranologists, including *Grundriss der iranischen Philologie* by Wilhelm Geiger and Ernst Kuhn, *Grundriss der neupersischen Etymologie* by Paul Horn, and *Persian Studies* by Hübschmann.

- 6) Notes taken from daily conversations and interviews with authors, poets, critics, and artists in different fields.
- 7) Pictures, tables and maps. (Khatibi, 2007,p.27&28).

Corpus in *The Sokhan Comprehensive Dictionary*

The Sokhan Comprehensive Dictionary is another important dictionary that has contributed to the evolution of Persian lexicography. It was published in 8 volumes in 2002 under the supervision of Dr. Hasan Anvari. If this dictionary is divided into two parts of old words and contemporary words, the latter part will be of more significance. The dictionary's introduction titled "*Word Selection*" has briefly pointed to the corpus used during its compilation process. Part of the introduction reads: "The linguistic corpus on which our dictionary has been based includes about two million examples of word usage mostly extracted from about four hundred old and modern texts. Our main focus has been on the current Persian. Our use of about one hundred and fifty contemporary short stories and novels as well as other sources in current Persian is indicative of such a focus. As for Old Persian, we should admit that our resources have been limited compared to the wide range of old texts" (Anvari, 2002, p.24).

Therefore, the failure to use sufficient older Persian texts is the main linguistic setback of this dictionary. Although the bibliography and content of the dictionary suggest that the most important Persian texts have been among its references, it has seemingly referred to other Persian dictionaries, especially *the Dekhoda Dictionary*, or other dictionaries of old texts rather than Persian texts to select older words, phrases or usages as well as current phrases. The only difference is that it has used a scientific method to list entries and information related to each word and phrase. A review of the corpus of old texts used in *The Sokhan Dictionary* reveals that it shares a large number of references with *the Dekhoda Dictionary*. Among the old translations and interpretations of the Quran, *Tafsir al-Tabari* is among the references of *The Sokhan Dictionary*. (Khatibi, 2007,p.29)

Computational Corpus Linguistics

Today's lexicography is closely interwoven with the use of computer. However, the traditional note-taking method has been used to compile general Persian dictionaries, including

Language in India www.languageinindia.com ISSN 1930-2940 14:3 March 2014

Zahra Zare and M. Balakumar

Corpus in Persian Lexicography

The Contemporary Persian Dictionary and *The Sokhan Comprehensive Dictionary*, in Iran over the past two decades.

The Persian Language Comprehensive Dictionary, presently under compilation by the lexicography department of the Persian Language & Literature House of Culture, is the first Persian dictionary to be based on a large computational corpus of words taken from Persian texts. (Khatibi, 2007.p30.).

Computational Corpora of Persian Language and Literature

Over the past two decades, the Persian Language with some limited Computational corpora has been developed. That has a great impact on the study of Persian Language and literature. Some of the important corpora include:

1. Dorj 3

The Compact Disc (CD) of the Mehr Argham Rayaneh Company

A new compact disc (CD) released by this company contains 178 Persian works in verse and prose from 101 authors and poets, including texts from the oldest times to the current era. This CD is the first and best reference for different types of research on Persian language and literature in terms of data size and application. (Khatibi, 2007,p31).

2. Persian Language Database

This database has been created at the Institute for Humanities and Cultural Studies under the supervision of Dr. Mostafa A'asi. It is available at the institute's website. The database, designed to create a large collection of various corpora of Persian language, includes selected literary, scientific, artistic and political Persian texts in written and spoken forms. (A'asi,1997). (www.pldb.ihcs.ac.ir)

3. The Persian Language Corpus of the Agah Publications

This corpus was prepared under the administration of Dr. Mohammadreza Bateni within a lexicography project at the Agah Institute in 1997-1999. It contains over 20 million words. Various usages of words from the current Persian language have been included in this corpus based on their level of importance since the project had adopted a stylistic and sociological approach toward Persian. The corpus includes various texts such as stories, novels, political, philosophic, social and scientific texts, news items, articles, newspaper reports, contemporary magazines, classified advertisements, texts related to different aspects of daily life, and school textbooks. (Khatibi, 2007,p31).

Some features of this corpus are as follows:

- 1- This corpus enables lexicographers to access each word and the whole examples and usages related to it within the framework of lexicography and corpus linguistics standards with the help of a keyword tool.
- 2- It enables lexicographers to put words in alphabetical order based on their previous and next words.
- 3- It contains different formats and images of verbal contexts for word usage.
- 4- It is a very effective tool for recording lexical homophony.
- 5- It accurately determines the frequency of each word since texts have been added to it constantly, enabling lexicographers to select entries depending on the purpose and extent of their dictionaries (Khatibi, 2007, p.32).

4. Database of Persian Language and Literature Texts

This project was approved at the Persian Promotion Council under the supervision of Dr. Mostafa Musavi in 2001. It started off one year later. Based on the project, all Persian language books in different fields authored after 1921 were supposed to be added to the database. The required software was designed and tested, and Persian texts authored during the ninth century and the first half of the tenth century, were added to it in the first phase of the project. However the project was stopped in 2003. (khatibi,2007,p.33)

5. The Computational Corpus of the Iranian Academy of Persian Language & Literature

Iranian Academy of Persian Language & Literature has prepared two corpora:
Language in India www.languageinindia.com ISSN 1930-2940 14:3 March 2014
Zahra Zare and M. Balakumar
Corpus in Persian Lexicography

1-Ganj Vajeh (treasury of words), is a corpus of Persian-Persian, English-Persian-English, German-Persian-German, French-Persian-French, Italian-Persian-Italian, and Spanish-Persian-Spanish dictionaries.

This project, carried out by the word selection department, included words and their meanings found in 375 dictionaries in different fields such as the humanities, medical science, agriculture, engineering, etc. The size of this corpus has reached to 6 million words so far. It also covers over 3000 words found in the glossaries of scientific books. The project is still underway. (Khatibi, 2007, p.33)

1- The corpus of the lexical units of Persian texts

This computational corpus of lexicography currently contains the lexical units of about 800 titles of Persian books (a total of over 100 volumes) and 7 million records. If each example contains at least 10 lexical units, the whole corpus will have about 70 million lexical units and examples. The corpus's software program has been designed with FoxPro in the user interface.

Type a lexical unit in the search field and the relevant sub-entries, examples, references and authorship date are shown immediately. If all the sub-entries and examples are required, click on the sub-entry count to observe the frequency of the lexical unit. Then, click on "export to Excel" option to access all the examples in chronological order. Another capability of this program is that it enables searching into the huge body of examples. Click on a button on the left of the example and a large number of examples will be provided in chronological order in the Excel. Prefixes, infixes, and suffices could be easily searched within the text to examine their behavior in words and contexts. (Khatibi, 2007, p.32&33).

Conclusion

Compiling the contents of the lexicons is the duty of a lexicographer. Undoubtedly, any kind of research about the language, description and analysis of language including lexicography, writing syntax, studying dialects, etc., will have no useful results without depending on the language corpus. Because the body of linguistic contents is too much (in most

of the languages present in the world) to become part of the corpus, we can at times create the corpus as the sample. Signifying the limits of corpus is the duty of the linguist.

In the tradition of lexicography in Iran, from the beginning till later years, the biggest corpus of lexicon was made of poetic texts but in the contemporary age, the prose has been used too in order to extract words. Today, we cannot compile a lexicon which is compiled with the new method of lexicography without established and determined corpus.

Suggestions

The type of linguistic corpus depends on the limits of lexicon. The lexicographer should decide about the extent to which he wants to bring in the old, obsolete and dialectical entries. This fact will help him in selecting the texts based on the historical arrangement and importance, people, places and times. It is required that a lexicographer has to review a big mass of published material every year in order to coordinate with the new words and definitions, linguistic changes and to update lexicons on the whole. The vocabulary contents and the texts that can be read on the computer are shared between lexicographers. They help encourage the compilation of better lexicons and prevent the copying of one lexicon from another. Because the computer processes the information faster and more accurately, the lexicographer considers the computer record more useful than copying from other lexicons.

References

- 1- Anvari, Hasan (2002), "*Dictionary References, the Sokhan Comprehensive Dictionary*", Sokhan Publications, Tehran, p 23.
- 2- Assi, Mostafa (1997), "*Persian Language Databases*", Collection of Articles Submitted to the Third Conference on Linguistics, Alameh Tabatabayi University and Institute of Humanities and Cultural Studies, Tehran
- 3- Assi, Mostafa (2000). *Language Corpora to Corpora Linguistic*, Collection of Articles Submitted to the Fifth Conference on Linguistics, Alameh Tabatabayi University and Institute of Humanities and Cultural Studies, Tehran, pp.484-495
- 4- Daei Al Eslam, Mohammad Ali, (1983). *Farhange Nezam*. Second Pub, Knowledge Institution. Tehran.

- 5- Dehkhoda, Aliakbar (1998), "*Loghat Nameh , Introduction to the Dehkhoda Dictionary*", pp 399-405.
- 6- Hashemi-Minabad, Hasan (2007), "*Corpus in Lexicography*", *Collection of Articles on Lexicography*, Tehran, Iranian Academy of Persian Language & Literature, Asar
- 7- Hashemi-Minabad, Hasan (2007), "*Corpus or, Collection of Articles on Lexicography*", Tehran, Iranian Academy of Persian Language & Literature, Asar
- 8- Khatibi, Abolfazl (2007), "*Persian Comprehensive Dictionary, Corpus in Persian Lexicography and Computational Corpus*", *Lexicography Magazine*, Num. 1, pp 4-49.
- 9- Meyer, Charles F., ed. (2002) *English corpus linguistics: An introduction*, Cambridge University Press.
- 10- Moein, Mohammad (1996), "*Persian Dictionary*", Amirkabir Publications, Tehran, ninth edition.
- 11- Siaghi, Dabir, Mohhamad. (1989) *Persian Dictionaries (Farhang ghoneh ha)*. Espargh Publication. Tehran
- 12- www.pldb.ihcs.ac.ir

=====

Zahra Zare
 Ph.D. Research Scholar
 Department of Studies in Linguistics, KIKS
 University of Mysore
 Manasagangotri
 Mysore-570006
 Karnataka, India
zahrazare92@yahoo.com

Dr. M. Balakumar
 Reader-cum-Research Officer
 Central Institute of Indian Languages (CIIL)
 Masagangotri
 Mysore-570006
 Karnataka, India

Language in India www.languageinindia.com ISSN 1930-2940 **14:3 March 2014**

Zahra Zare and M. Balakumar
 Corpus in Persian Lexicography

