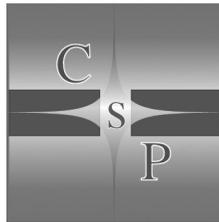


Aspects of Iranian Linguistics

Aspects of Iranian Linguistics

Edited by

Simin Karimi, Vida Samiian and Donald Stilo



Cambridge Scholars Publishing

Aspects of Iranian Linguistics, Edited by Simin Karimi, Vida Samiian and Donald Stilo

This book first published 2008 by

Cambridge Scholars Publishing

12 Back Chapman Street, Newcastle upon Tyne, NE6 2XX, UK

British Library Cataloguing in Publication Data

A catalogue record for this book is available from the British Library

Copyright © 2008 by Simin Karimi, Vida Samiian and Donald Stilo and contributors

All rights for this book reserved. No part of this book may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior permission of the copyright owner.

ISBN (10): 1-84718-639-4, ISBN (13): 9781847186393

Table of Contents

| | |
|--|-----|
| Introduction Simin Karimi, Vida Samiian, Donald Stilo | 1 |
| Recent advances in Persian lexicography Mohammad Reza Bateni | 3 |
| A link grammar parser for Persian Jon Dehdari and Deryle Lonsdale | 19 |
| Classifiers, plural and definiteness in Persian Lewis Gebhardt | 35 |
| Optionality and variation: a stochastic OT analysis of M/p-echo reduplication in Colloquial Persian Saeed Ghaniabadi | 57 |
| Markedness and bare nouns in Persian Jila Ghomeshi | 85 |
| The emergence of ergativity in Iranian: reanalysis or extension? Geoffrey Haig | 113 |
| The noun phrase in Hawrami Anders Holmberg and David Odden | 129 |
| Expressions of future in Classical and Modern New Persian Carina Jahani | 153 |
| Raising and control in Persian Simin Karimi | 177 |
| Event structure of verbal nouns and light verbs Gholamhossein Karimi-Doostan | 209 |
| Differential object marking in a Medieval Persian text Gregory Key | 227 |

| | |
|--|-----|
| Marking of arguments in Balochi ergative and mixed constructions Agnes Korn | 249 |
| Inversion and topicalization in Farsi discourse: A comparative study Shahrzad Mahootian | 277 |
| Aspects of agrammatical language in Persian Reza Nilipour | 289 |
| The individuating function of the Persian ‘indefinite suffix’ Daniel Paul | 309 |
| Some remarks on Persian suffix –râ as a general and historical issue Ludwig Paul | 329 |
| The Ezafe as a head-marking inflectional suffix: evidence from Persian and Kurmanji Kurdish Pollet Samvelian | 339 |
| Two sets of mobile verbal person agreement markers in the Northern Talyshi language Donald Stilo | 363 |
| Mood and modality in Persian Azita Taleghani | 391 |
| On the ergativity in the Pamir languages Antje Wendtland | 419 |
| Index | 435 |

Introduction

Simin Karimi, Vida Samiian, Donald L. Stilo

This volume consists of twenty papers presented at the First International Conference on Aspects of Iranian Linguistics, which was held at the Max Planck Institute for Evolutionary Anthropology in Leipzig, Germany, June 17–19, 2005. The conference included thirty-one papers organized in panels, seven presentations displayed at the poster sessions, and two invited keynote addresses by Dr. Gilbert Lazard and Dr. Mohammad Reza Bateni. This volume is dedicated to Dr. Mohammad Reza Bateni and hopes to honor and celebrate his impressive contribution to modern linguistics in Iran.

Born and raised in Iran, Dr. Bateni received his B.A. in English Language and Literature in 1960 in Iran, and a postgraduate diploma in English and Linguistics from Leeds University in England (with distinction) in 1961, followed by an M.A. in Linguistics from the same university in 1963. He was a graduate student at the University of London from 1963 till 1965, and obtained his Ph.D. in General Linguistics and Ancient Iranian Languages from Tehran University in 1967. He was the first professor to teach modern linguistics at the newly established Linguistics Department at Tehran University in 1967. He served as a Professor in that department until 1987. He has been an Academic Advisor and Head of the Research Unit of Farhang Moaser Publishing House (a major publisher of reference books in Iran) from 1987 until the present.

Dr. Bateni has published many books, seven of them in the areas of General Linguistics and Persian Linguistics, which have been used at various universities nationwide in Iran. Some of these books have been reprinted as many as twenty times. He is also the author of numerous papers on various aspects of linguistics, specifically syntax and semantics. Dr. Bateni is the Editor in Chief of five English-Persian dictionaries published by Farhang Moaser, the first of which appeared in 1993, and won the Best Book of the Year Award in Iran. A revised and enlarged version of this volume was published in 1997. His two-volume English-Persian dictionary appeared in 2006, followed by a Dictionary of Phrasal Verbs (English-Persian) in 2007, and a Dictionary of Idioms (English-Persian) in 2008. He is currently working on his first Persian-English Dictionary. Dr. Bateni has been a research associate and visiting scholar in the linguistics departments at such prestigious universities as MIT and the University of California, Berkeley.

Many Iranian linguists have directly or indirectly benefited from Dr. Bateni's teaching and mentorship over the years. In fact, one of the editors of this volume, Simin Karimi, had the honor of being his advisee at Tehran University in 1976–77. He is considered the major founder of modern linguistics in Iran, and enjoys enormous respect among students and scholars who work on different aspects of Iranian linguistics in Iran and abroad.

This volume consists of twenty papers, including the paper Dr. Bateni presented at the conference in Leipzig. The editors wish to thank the authors of these articles for their valuable contributions. Each single paper was evaluated by two reviewers. Our special thanks go to all those who took the time to do a thorough review of the papers and provide detailed comments: Brian Aghbayani, Miriam Butt, Andrew Carnie, Mohammad Dabir-Moghaddam, Ali Darzi, Lewis Gebhardt, Jila Ghomeshi, Chris Golston, Geoffrey Haig, Heidi Harley, Martin Haspelmath, Anders Holmberg, Carina Jahani, Arsalan Kahnemuyipour, Gholamhossein Karimi-Doostan, Agnes Korn, Will Lewis, Shahrzad Mahootian, Karine Megerdooonian, Reza Nilipour, David Odden, Ludwig Paul, Mohammad Rasekh, Anna Roussou, Pollet Samvellian, Martha Shulte-Nafeh, Azita Taleghani, Rudy Troike, and Wendy Wilkins. Also, special thanks go to the Max Planck Institute for Evolutionary Anthropology and the local organizers of the conference in Leipzig, Claudia Schmidt and Claudia Büchel, who did a fantastic job in organizing the meeting. Many thanks also to the Departments of Linguistics at the University of Arizona and California State University, Fresno for co-sponsoring the conference. Sumayya Racy and Jeff Punske, two graduate students at the University of Arizona helped with the correspondence with the authors and the collection of the articles. Many thanks to them. Very special thanks to Adam Baker, another graduate student at the University of Arizona who did the final editing and formatting of the volume.

We are especially thankful to the Roshan Cultural Heritage Institute for its valuable contribution. Without the financial support of this institute neither the conference nor this volume would have been realized.

Simin Karimi, Vida Samiian, Donald L. Stilo

Recent Advances in Persian Lexicography

Mohammad Reza Bateni

1 Introduction

The art of lexicography in Persian has a long history. I am not going to deal with this subject from a historical point of view. Rather, as the title of this chapter suggests, I will confine myself to touching upon some major reference works, especially dictionaries, which have appeared in the past twenty years, which have been based a specific methodology and which have had some impact on the field of lexicography in general and on their users in particular. A word of caution is in order here. No account of such a topic could be either complete or absolutely objective. Bear in mind that this is only one account of the recent advances in Persian lexicography.

The subject matter can be divided into three sections. Section two deals with monolingual Persian dictionaries, or more technically, with dictionaries whose source language and target language are Persian. Section three is concerned with the recent bilingual dictionaries. This domain can be divided into two subsections. Subsection (a) takes a foreign language as the source language and Persian as the target language, such as German-Persian or English-Persian dictionaries. Subsection (b) takes Persian as the source language and a foreign language as the target language, such as Persian-German or Persian-English dictionaries. Section four deals with specialized or technical dictionaries. These works usually take English as the source language and Persian as the target language in an area of specialization; for instance, an English-Persian dictionary of law. There are very few works which are compiled the other way around, i.e. with Persian as their source language.

2 Monolingual dictionaries

If you asked an educated Iranian two decades ago about the best Persian dictionary or dictionaries, he would mention *Mo'in's Persian Dictionary* and/or *Dehkhoda's Loghatname*, which is a kind of encyclopedic dictionary. In fact, the modern era of dictionary-making for the Persian language started after Dehkhoda's and Mo'in's works. I am excluding these two famous works from my discussion, however, on the grounds that they lack a

coherent methodology, being Arabic-oriented in their definitions of words and drawing heavily on classical poetry as their sources of information. I am not going to elaborate on these reasons for their exclusion.

The first breakthrough in Persian dictionary-writing was made by a group of three authors: G.H. Sadri Afshar, Nasrin Hakami and Nastaran Hakami. The volume is entitled *Farhang-e Moaser-e Farsi*, “A Dictionary of Contemporary Persian.” Work on this dictionary started in 1981 and took 10 years to finish. It was released in 1990 by Kalame Publishers. During the next ten years, it was revised and expanded three times. In 2002, the dictionary changed hands and its fourth edition was published by Farhang-e Moaser Publishers. The fifth edition of this work is in preparation and will be published very soon. Relatively speaking, *Farhang-e Moaser-e Farsi* is a very good dictionary. It is the only Persian dictionary which has undergone continual revision, and really justifies its title “A Dictionary of Contemporary Persian.” It is also an award-winning dictionary.

This same group of three co-authors have produced another book of reference. It is entitled *Farhang-e Farsi-ye A’lâm* “A Persian Dictionary of Proper Names.” This work was published by Farhang-e Moaser Publishers in May 2005. It contains 14,000 entries, provides valuable information about famous persons, places, books, works of art, historical events—such as wars and treaties—and many other topics of interest. It is in fact an educational supplement to their Persian dictionary.

Next in chronological line comes a two-volume dictionary compiled by Abolhassan Najafi and published by Morvarid Publishers in 1999. It is entitled *Farhang-e Farsi-ye Amiyâne*, which I translate, “A Dictionary of Colloquial or Spoken or Informal Persian.” I think the word *âmiyâne* in the title is misleading. In relation to language, *âmiyâne* means “vulgar,” “slang,” “substandard” or “non-standard,” whereas the entries included in the book are far from being slang or substandard. We find entries such as *rooze gereftan* “to fast or to go on fasting” or *rooze shekastan* “to break one’s fast” in this volume. Apart from this difficulty in naming, the dictionary fills a large gap that previously existed in Persian lexicography. It covers an area of the language which had been totally ignored by the older dictionaries, and which had been lightly touched upon by the recent ones. The work is based on years of data collection through reading a large number of novels and short stories, and through listening to people’s everyday conversation. It confines itself to the colloquial expressions common in Tehran and draws on sources written in the past 80 years for citations and examples. This dictionary is not a replacement for general Persian dictionaries, but it is a great and badly-needed supplement to them.

The most recent Persian dictionary is *Farhang-e Bozorg-e Sokhan* “*Sokhan Comprehensive Dictionary of Persian*,” an eight-volume dictionary compiled under the supervision of Hassan Anvari as the Editor-in-Chief, and published by Sokhan Publishing Co. in 2002. The dictionary was put together over a period of eight years. The research and compilation was carried out, according to the publisher, by a group of more than one hundred contributors, including consultants, specialist editors, general editors, lexicographers, translators, designers and so forth. Apart from a large number of current periodicals, some 450 literary works, old and new, were used in the process of extracting the entries. The dictionary includes 80,000 main entries, 40,000 subentries and 170,000 citations and examples. These citations are quoted from the works of 800 writers and poets and, according to the publisher, cover a period of 12 centuries, from the ninth to the twenty-first century A.D. In some cases, in order to show the semantic changes of a word, three citations are given: one from the 9th century, another from the 14th or 15th century and a third one from the 20th or 21st century. In order to avoid confusion, the meanings of a word which are not used anymore are labeled “obsolete.” For instance, *tavile* has three meanings. In contemporary Persian, it means “a place where domesticated animals are kept.” This meaning is unmarked or unlabeled, whereas the other two, “a string of pearls or of other precious stones” and “a rope by which animals were tied up,” are marked “obsolete.” For the sake of brevity, I will not go into more details about this dictionary. In brief, I believe *Farhang-e Bozorg-e Sokhan* is the best Persian dictionary available at the present time.

It is important to mention that there have been two projects for compiling Persian dictionaries which have failed for various reasons. One of them was the first computerized data-based scheme. The project belonged to Agah, a private publisher, and I was its Editor-in-Chief. The software was written by a group of graduate students of Sharif University of Technology, and was working fairly well though the OCR (the optical character reader) available to us was not satisfactory, and the data collection needed a good deal of keyboarding. In spite of all this, we collected more than 5 million words. The first three letters of the Persian alphabet, “â,” “a,” and “b,” and part of the fourth “p” were completed when the publisher ran out of money, and the whole project collapsed. The work on this project took about two and a half years, from 1997 to mid 1999.

Although it is regrettable that this project did not materialize, I learned a lot about the advantages of a data-based dictionary, and believe that one cannot be so sure about the usefulness of the entries included in a dictionary if it does not make use of a database. Here is one example. I remember that we

were searching our limited database for the word *arike* “throne.” We came up with more than one hundred instances of the use of this word. But the word *arike* alone or *arike-ye saltanati* “the imperial throne” occurred only five times whereas *arike-ye ghodrat* “the seat of power” occurred nearly a hundred times. Supposing that our database was balanced, this means that the inclusion of *arike-ye ghodrat* in a Persian dictionary has priority over the single word *arike*. Since the dictionaries I have discussed in this section had no computerized database to depend on, it has resulted in the inclusion of the entry *arike* and the exclusion of the figurative and more common expression of *arike-ye ghodrat*.

Another project for writing a Persian dictionary that failed miserably belonged to Markaz-e Nashr-e Daneshgahi (Iran University Press), a semi-governmental organization. This project started in 1986 and came to a halt in 2004. That is, it was in the making for 18 years before it stopped completely. According to its latest Editor-in-Chief, approximately 60% of the work was done, and the cost was about half a billion tooman (roughly 500,000 U.S. dollars). I have been told that approximately another half a billion toomans is required to finish the job. The reason given for dropping the project is lack of funds; but in my view, lack of funds is the effect not the cause. The real cause seems to be mismanagement. The main point is that the project has collapsed and there is little chance for it getting off the ground again. If the project is dropped forever, which is most likely, it means that half a billion toomans has gone down the drain.¹ The government has proven time and again that it is inefficient at managing its affairs, and the collapse of this project proves this point once again.

The next section is devoted to a discussion of bilingual dictionaries.

3 Bilingual dictionaries

In this section, I will first take up those bilingual dictionaries which take a foreign language as their source language and Persian as their target language. The discussion of Persian-other language dictionaries will follow in the second part of this section.

¹This did happen. The project was totally abandoned. On the face of it, it was transferred to the Academy of Persian Language and Literature, to gather dust there.

3.1 Foreign language-Persian

This subsection is divided into English-Persian, French-Persian, German-Persian and other languages-Persian. It goes without saying that English-Persian dictionaries have the high priority. English is almost exclusively the second language taught at schools, and is the first option in the Iranian colleges and universities, though this is not generally officially mentioned. Therefore it is natural that more research and effort are being devoted to English, and that more English-Persian dictionaries are produced.

Soleyman Haiim was a great pioneer in this field. He was a genius. Almost every Persian lexicographer has been influenced by him to varying degrees. In fact, English-Persian dictionary writing in Iran is associated with the name of Haiim. Here is a short story on this very issue. A friend of my son, as he was a school boy, asked him the following question: "What kind of work is your father doing?" My son replied "He is writing an English-Persian dictionary." "Oh, I see," the boy replied "he is writing Haiim!" In spite of all this, times have changed and the time of Haiim dictionaries is over, although their publisher, Farhang-e Moaser, is trying to keep them alive by revising and updating them. Some years after Haiim's work, a five-volume dictionary called *The New Unabridged English-Persian Dictionary* was compiled by Abbas Aryanpur Kashani and published by Amirkabir Publishers in 1963. I will not discuss these two works here, and will concentrate on more recent dictionaries which really started after Haiim's and Aryanpur's works.

In 1985, Farhang-e Moaser Publishers asked me to compile a small English-Persian dictionary for tourists. Some time later, they changed their mind and asked me to expand it to a fully-fledged one-volume dictionary. My assistant Fatemeh Azarmehr and I spent about 8 years on this dictionary, which finally appeared in 1992, under the title of *Farhang-e Moaser English-Persian Dictionary*.

Two criticisms were leveled at this dictionary. The first was that the dictionary did not include phrasal verbs, idioms and so on. This criticism was quite justified. The reason for this shortcoming was that the financial resources of the publisher at that time did not allow him to wait longer than 8 years and spend more money on this project. So it had to be marketed as it was.

The second criticism was about its phonetic transcription. Before taking up this point, an introductory remark is in order. As a student at Teachers' Training College in Tehran, I had an English professor by the name of Dr. Nye. One day a fellow student of mine asked her "what kind of English are we supposed to learn, British English or American English?" She promptly

replied “You learn neither British English nor American English; if you learn any English at all, you’ll learn Persian English.” If we translate this answer into technical terms, it means that the pronunciation of adult learners of a foreign language, in this case the one of adult Persian learners of English, can hardly become identical with the pronunciation of native speakers, but can approximate to this ideal to various degrees. Of course there are exceptions. Here I do not intend to try and explain the scientific basis of this phenomenon.

With all of this in mind, I devised a simplified system of phonetic transcription which suited Persian learners of English and also represented the nearest approximation of the pronunciation of a native speaker of English. This simplified version of phonetic transcription came under attack from a particular section of the learned society, that is, from linguists and students of linguistics, from professors of English, English teachers and some others. In brief, it came from those who were familiar with the IPA (the International Phonetic Alphabet) in the written shape, although the majority of them were not able to translate the IPA phonetic symbols into the correct pronunciation. All this criticism did not shake my belief that I was on the right track, and I have continued to use this system of transcription in subsequent editions.

In spite of the shortcomings and criticisms, the work was recognized by the government as the award-winning dictionary in 1994. The dictionary won the award for the relative accuracy of what it contained and for its breaking with the past and ushering in a new era of dictionary-making. Before I close the discussion on this work, I should add that in spite of the fact that the dictionary won the award, I refused to receive the prize because I did not like to shake hands with the president at the time, Mr. Rafsanjani, and Mr. Larijani, then the Minister of Culture and Islamic Guidance, and I am glad I did so.

The second edition of this work appeared in 1997, with the help of four assistants. The first edition included about 30,000 entries; the second edition brought the number of entries to about 50,000. A CD-ROM of the second edition was also released. The academic work on the third edition of this dictionary is completed and is now in press.² It is a two-volume dictionary with phrasal verbs, idioms, common collocations, and thousands of examples accompanied by their Persian translations. It includes entries like “bird influenza,” “mad-cow disease,” “SARS,” “nuclear terrorism,” many of which are not included in the 2005 edition of some of the English-English dictionaries. We have checked and rechecked the contents of the second edi-

²This work has been published in the meantime, and is now in bookstores (the editors).

tion and, to my disappointment and regret, we have found inaccuracies and errors which have been corrected and for which we ought to apologize to the public. It must be mentioned that all three editions of this dictionary have been compiled in the Research Unit of Farhang-e Moaser. Without the facilities provided by the publisher in this unit nothing of this kind could have been produced.

Moving on, acknowledgement of a great contribution to the field of English-Persian lexicography is in order at this point. This contribution was made by three co-authors: Ali-mohammad Haghshenas, Hossein Samei and Nargues Entekhabi. This is a two-volume dictionary entitled *Farhang-e Moaser Millennium English-Persian Dictionary*. The compilation of this dictionary took 13 years, which was started in 1985 in the Research Unit of Farhang-e Moaser, and was finished in 1998. It took the publisher an additional period of three years to prepare this work for publication, which was finally distributed in 2001. According to the preface to the dictionary, it contains more than 55,000 entries and subentries; 30,000 idioms, collocations and proverbs; more than 300,000 Persian equivalents, and 254 practical notes on usage. This dictionary won the award for the best reference book in Iran in 2002.

Similar to any other dictionary, *The Millennium Dictionary*, as it is called for short, is not free of inaccuracies, which mainly result from heavily depending on, and uncritically following, Haiim's work. The consequence is that many of the errors found in Haiim's Dictionary are reflected in this work as well. Some examples are provided in the next two paragraphs.

Persian equivalents given for the English word *skin* are *charm*, *pust*, and *timâj*, meaning, respectively, "leather," "skin" and "goatskin." As far as I know, "leather" is not synonymous with the untreated unprocessed skin of animals. For example, you speak of "a leather coat," "a leather belt" or "a leather handbag," but you do not say "a skin coat," "a skin belt" or "a skin handbag."³ As for *timâj*, it is a special kind of leather made by tanning the skin of a goat. The source of this inaccuracy can be found in Haiim's Dictionary, where *pust* and *charm* are used synonymously as equivalents for the word "skin." Apparently the authors have concluded "if *charm* why not *timâj*, if "leather" why not "goatskin.""

Here is another example. The order of Persian equivalents for the English word *bed* in terms of usage should be *takht*, *takhtekhâb*, (loosely speaking) *rakhtekhâb*, and *bestar*. In the dictionary in question, the order is reversed

³It should be noted that compound nouns, such as *sheepskin*, *snakeskin*, *eelskin*, and *leopardskin* are not under discussion here. It is the word *skin* that is being dealt with in this chapter.

from the proper order, which is similar to the order you find in Haiim's dictionary. The overconfidence of the authors in Haiim's work is quite striking. In spite of all this, *The Millennium Dictionary* is a great and valuable work and is one of the bestselling dictionaries.

Another recent dictionary worth mentioning is *Nashr-e Now Concise English-English-Persian Dictionary*, compiled by Mohammad-Reza Ja'fari and published by Fakhte Publishers in 1997. As apparent in the title, the book is arranged in a peculiar way. The dictionary is based on *Longman Handy Learner's Dictionary*, which appeared in 1991. The English-English dictionary is wholly transferred into this dictionary and then the Persian equivalent of each English definition is given in the same order as in the English text. In order to help the reader understand the nature of this volume, here I provide an example of its content.

Consider the word *horse* as an entry. The first definition of this entry that appears in this work is "a large four-legged solid-hoofed animal with flowing mane and tail used from early times to carry loads, for riding, etc." The Persian equivalent of this definition is given as *asb*, which is correct. And it goes on like this for other definitions and expressions.

The author of this dictionary believes that learners of English learn more if they see English definitions and their Persian equivalents side by side. This could be true. There is, however, another advantage in employing this method. That is, if the author or translator makes a mistake, it will be found out by this side-by-side methodology. Here is an example. Let us take the entry "skin" again. The first definition says "natural outer covering of the body," which is correctly translated as *pust*. The second definition says "skin of an animal used to make leather." That is, the author does not use "leather" but "used to make leather." The Persian equivalent of this definition is given as *charm* "leather" which is wrong. Here again, Haiim has been copied uncritically. In short, some of the errors and inaccuracies of Haiim's work have been perpetuated through copying and recopying. In spite of all this, I think Mohammad-Reza Ja'fari's work is a good and useful dictionary.

In 1998 a six-volume dictionary appeared, entitled *The Aryanpur Progressive English-Persian Dictionary*, compiled by Manoochehr Aryanpur Kashani, and published by Jahan-Rayane. This dictionary can be considered as the second edition of the work of Abbas Aryanpur Kashani, the father of Manoochehr Aryanpur Kashani, which had appeared 40 years earlier. No doubt there are a number of improvements in this edition which should be taken into account, but the basic issue with this work is that you do not know whether it is meant to be an English-Persian dictionary or an encyclopedia. For the sake of brevity, I will cite only one example here. One of the en-

tries in the dictionary is *Henry*. The English translations for this entry are (1) “a male proper name” (abbreviated to Hank, Harry and Hal), (2) “The name of several Kings of England and other European countries, including Henry the Seventh (the King of England born 1457 and died 1509) and Henry the Eighth (1491-1547) who converted the English into Protestantism.” You find the names of various people and places in this dictionary from André Gide (French writer) to Katharine Hepburn (American actress) to the Libyan Desert.

Furthermore, this dictionary contains many inaccuracies and errors. We have, for example, the Persian phrase *bâ hamkâri-ye*, which the author has translated word for word as “with the collaboration of ...,” while the more natural English version is “in collaboration with....” The less preferred English translation of “with the collaboration of” appears on the cover of the volume. Another example is about the first entry in the dictionary which is the letter A. The first meaning provided for this entry is “the first letter in the English alphabet,” which is correct. However, the author has provided a second meaning which states that “any of the sounds that this letter in English has,” which, as far as I understand, is nonsensical. To sum up, this six-volume dictionary is, in my opinion, of very poor quality.

Last in line of the English-Persian dictionaries comes *Farhang-e Moaser School Dictionary*, compiled by Nargues Entekhabi and published by Farhang-e Moaser Publishers in 2003. It is designed for English learners aged between 9 and 14. The dictionary contains 3500 most frequently used words in English. It won the award offered by *The Children’s Book Council* for the best children’s book in 2003. The book is attractively typeset, bound and illustrated.

Now we turn to French-Persian dictionaries. Almost all modern French-Persian dictionaries have been developed in the Research Unit of Farhang-e Moaser, and are compiled by Moahmmad-Reza Parsayar. The largest one is *Farhang-e Moaser French-Persian Dictionary*, compiled by the same author mentioned above, edited by Hormoz Milanian, and published by Farhang-e Moaser Publishers in 2001. The third reprint of this work appeared in 2003.

Previous work by Mohammad-Reza Parsayar consists of two volumes: *Farhang-e Moaser Pocket French-Persian Dictionary (Dictionnaire de Poche, Français-Persan)*, first published in 1995 and reprinted several times since, and *Farhang-e Moaser Shorter French-Persian Dictionary (Petit Dictionnaire Français-Persan)*, first published in 1993, revised for second edition, and published in 2002. The most recent work by this author is *Farhang-e Moaser School French-Persian Dictionary (Dictionnaire de l’école Français-Persan)*, published in 2005. This work is one of a series of dictionaries

designed for learners of foreign languages aged between 9 and 14. All of these dictionaries are patterned on the same basis.

This summarizes the review of French-Persian dictionaries. I am not aware of any other recently published French-Persian dictionary worth mentioning in this chapter.

Now, let us turn to German-Persian dictionaries. A recent and serious work in this area is *A German-Persian Dictionary (Deutsch-Persisch Wörterbuch)*, compiled by Faramarz Behzad, and published in 2002 by Kharazmie Publication and Distribution Co. in Tehran, in collaboration with Logos Verlag, Bamberg, Germany. It is the result of 12 years of research and contains more than 30,000 contemporary German words and almost the same number of expressions, idioms and collocations.

Although the author states that the dictionary is compiled in such a way that it is useful both to German and Persian users, it is believed that the balance is tipped in favor of the German users. All the helpful hints about the German words and the arrangement of their Persian equivalents are given in German. Mr. Behzad's work is the best German-Persian dictionary available in terms of contents, methodology and form.

In 2002, a two-sided tourist-oriented dictionary was published by Langenscheidt Publishers in Germany. It is compiled by Khosro Naghed and Mohsen Naghed. This *Persian-German, German-Persian Dictionary (Persisch-Deutsch, Deutsch-Persisch Wörterbuch)* is not a serious work, but it serves its purpose as a handy tourist dictionary. You come across some confusion in the Persian section, rising apparently as a consequence of having been laid out by people who did not know Persian.

The last item to be mentioned here is *Farhang-e Moaser School German-Persian Dictionary (Schulwörterbuch, Deutsch-Persisch)*, compiled by Nargues Entekhabi, and published by Farhang-e Moaser Publishers in 2003. This is another book in the same series of dictionaries designed for learners of foreign languages aged between 9 and 14, all of the same format.

A distinguished work among the bilingual dictionaries is *Farhang-e Moaser Arabic-Persian Dictionary*, compiled by Abdolnabi Ghayyem, and published by Farhang-e Moaser Publishers in 2002. This is the first Arabic-Persian dictionary which does not presuppose any knowledge of Arabic in order to be able to look up a word. How does this work? Here is an example.

Suppose you are looking for the meaning of the word *masjed* "mosque." If you know the order of the Arabic alphabet, you right away find the word in this dictionary, whereas in traditional Arabic-Persian dictionaries you should already know that *masjed* is a derivative of the verb *sajada*, meaning "to kneel before God," and that *masjed* is the name of a place where this action

takes place. In this type of organization, if the word *masjed* is listed at all, it is listed under *sajada*. The irony is that if you know enough Arabic to be sure that *masjed* is a derivative of *sajada* and is probably listed under *sajada*, in all probability you do not need to look it up because the presupposed knowledge of Arabic tells you that *masjed* is the name of the place where the action of *sajada* takes place. In short, since *Farhang-e Moaser Arabic-Persian Dictionary* does not presuppose this knowledge, it is the best of its kind available at the present time.

Another item under this heading is *Azerbaijani Turkish-Persian Dictionary*, compiled by Behzad Behzadi, and published by Farhang-e Moaser Publishers, in 2003. As indicated in the title, the source language in this dictionary is the variety of Turkish which is spoken in Azarbaijan, Iran, and not the one spoken in Turkey. It should also be mentioned that the book was previously published by another publisher. From a lexicographical point of view, the dictionary is far from being ideal, but unfortunately it is the only work of its kind available at the moment.

We close this subsection by introducing a very important work entitled *Sogdian-Persian-English Dictionary*, compiled by Badrozzaman Gharib and published by Farhangan Publications in 1995. Sogdi is one of the Middle-Iranian languages which was spoken from approximately the second to the thirteenth century A.D. over a vast area of Middle and Eastern Asia, from Samarkand to the Great Wall in China. This dictionary is a monumental contribution to the field of Persian lexicography.

3.2 Persian-Foreign language

Now we come to the second subsection of bilingual dictionaries. Unfortunately, we have very few items to present here. What has been done in this domain is next to nothing. Let us start with Persian-English works.

In 2003 we witnessed the appearance of a very good dictionary entitled *Alborz Persian-English Dictionary* compiled by Seyf Ghaffari, and published by Alborz Publishers. This is a practical, up-to-date dictionary, well-organized by the author but unattractively presented by the publisher. Like any other dictionary, it is not fault-free. For example, “old age” and “senility” are given as synonymous equivalents for the Persian word *piri*, although *senility* does not mean “old age” or *piri*. It is a decline or deterioration of mental functioning, especially of short-term memory and alertness, caused by old age or disease. In other words, senility is a condition often, but not always, accompanied by old age, but not “old age” per se. In spite of inaccuracies of this sort, *Alborz Persian-English Dictionary* is a good and useful

dictionary.

The most recent work under this heading is *Aryanpur Progressive Persian-English Dictionary*, a four-volume dictionary compiled by Manoocher Aryanpur Kashani, and published by Jahan-Rayane in 2004. This dictionary could have been excluded from this review by the criteria I had set myself for choosing the items, but I decided to mention it here without making any comments on it.

A two-sided dictionary, Persian-English, English-Persian, was published by Farhang-e Moaser Publishers in 2005. It is called *Farhang-e Moaser Bilingual Pocket Dictionary*. It is a very useful, handy dictionary, and the first of its kind produced in Iran.

A totally new Persian-English dictionary is being compiled by Karim Emami, and is to be published by Farhang-e Moaser in near future.⁴

Under the heading of Persian-French, all we have is *A Persian-French Dictionary*, compiled by Gilbert Lazard with the assistance of Mehdi Ghavam-Nezhad. The dictionary was designed mostly for French users. Therefore, it was arranged from left to right, with a detailed introduction in French. It was first published in the Netherlands in 1990. The original version was available in Iran for some time. Then piracy took over. An Iranian publisher changed the order of the book and made it right to left, in accordance with the Persian writing tradition, clumsily translated the author's introduction, and boldly published it in 1991.

As it usually goes, a new thief can steal from an old one. Recently I saw that another publisher in Ghom had published the same pirate copy in its own name, and had gone one step further, giving the date of the first publication of this work as 2002, that is, the date that its own pirate copy had appeared. Well, so far we have seen the history of the publication, but what about the content?

No doubt, Lazard's work is a valuable contribution to Persian lexicography. Nevertheless, I am not in full agreement with him in some respects. First, I think that the inclusion of many dated, old-fashioned entries in a concise dictionary of this kind is not justified. For example, *khâkestardân* meaning *cendrier* or "ashtray"; *arkhâlogh* an old-fashioned piece of clothing, *takhvij* (from Arabic *khowf*) meaning "intimidation."

Secondly, many informal familiar Persian expressions are translated word for word, losing sight of their real connotations. For example, *qorbân-e kesi raftan* or *qorbân-e kesi shodan*" is translated *être prêt à donner sa vie*

⁴This dictionary, entitled *Farhang-e Moaser Kimia Persian-English Dictionary*, was published posthumously by Farhang-e Moaser Publishers in 2006. This is the best Persian-English Dictionary available at the present time.

pour qqn, meaning “be ready to give one’s life for someone.” That is not the case at all. When a mother says to her child *qorboonet beram* she does not mean “I am ready to give my life for you.” If she means anything at all, she means “I love you” or something of this kind. The same applies to the expression *pishmarg-e kesi shodan* which is word for word translated into French as *mourir avant qqn*, meaning “to die before someone.” When a person uses this expression, he/she does not mean this at all. We Iranians are very generous with the use of these expressions, but they are either void of meaning or at most they imply some sort of affection.⁵ What I said here does not reduce the value of Lazard’s work. It still remains a great contribution to Persian lexicography in recent years.

There are no other items to mention in this subsection. We now turn to the final section which is concerned with specialized dictionaries.

4 Specialized dictionaries

The situation here is not neatly organized. It is absolutely chaotic. Compilation of specialized dictionaries has mushroomed in recent years, running to a hundred or more. A great number of those so-called technical dictionaries are just worthless. They do no good, and rather add to the already existing confusion in technical terminology. What I am trying to do here is to impose some arbitrary order on them and then cite some reliable examples.

4.1 Compilation

By compilations I mean those English-Persian technical dictionaries which are put together by an academic organization or by individual authors. Below is a sample list. Bear in mind, however, that (a) the list is by no means exhaustive and (b) the ordering is random.

- *A Dictionary of Mathematics and Statistics*. 1991. Published by the Iranian Mathematical Society in collaboration with Iran University Press.
- *A Dictionary of Physics*. 1987. Iran University Press.
- *A Dictionary of Chemistry*. 1991. Iran University Press.

⁵These Persian expressions could be compared to frozen, cliché-like greetings in English; for example, the formal greeting of “how do you do?” Also consider the Persian expression *khodâ hâfez* “God protect you”, which is not used as a blessing at all, but rather as a simple greeting.

- Soltanpur, Zahra, compiler. 1995. *A Dictionary of Metallurgical Engineering*. Iran University Press.
- Dahi, M. R. and F. Moslehi Moslehabadi, compilers. 1996. *A Dictionary of Agriculture*. Iran University Press.
- Daneshgahi, Jahad. 1991. *A Dictionary of Mathematics*. Sharif University of Technology.
- Mozafarian, V., compiler. 2004. *Trees and Shrubs of Iran*. Farhang-e Moaser Publishers.
- Baraheni, Mohammad-Naghi et al., compilers. 1996. *A Dictionary of Psychology and Related Fields*. Farhang-e Moaser Publishers.
- Golriz, Hassan, compiler. 2001. *A Dictionary of Money, Banking and International Finance*. Farhang-e Moaser Publishers.
- Mohammadifar, Mohammad-Reza, compiler. 2004. *A Dictionary of Computers*. Farhang-e Moaser Publishers.
- Ashouri, Daryoush, Nashr-e Markaz compiler. 1996. *A Dictionary of Human Sciences*.
- Majd Law Dictionary. 1999. Majd Publications.
- Nowruzi, Mehdi, compiler. 1995. *A Dictionary of Political Terms*. Ney Publishing House.
- Mohajer, Mehran and Mohammad Nabavi, compilers. 2003. *A Lexicon of Literature and Literacy Discourse*. Agah Publishing House.
- Pazargadi, A. and G. Tofangdar, compilers. 1991. *A Dictionary of Sports*. Soroush Press.

4.2 Translation

English-Persian technical dictionaries

In the translation category, I include those English-Persian technical dictionaries which are direct translations of an English-English technical dictionary. Below is a sample list. Once again I emphasize that the list is not exhaustive and that the ordering is random.

- Mofidi, Mostafa, translator. 1996. *Farhang-e Feshorde-ye Pezeshki* (a translation of *A Concise Medical Dictionary*). Oxford University Press. Farhang Publishers.
- Afzali, M. R., translator. 2000. *Farhang-e Oloum-e Mohandesi* (a translation of *McGraw-Hill Dictionary of Engineering* by S.P. Parker). Daneshyar Publishers.
- Hosnavi, Reza and Daryoush Farsayi, translators. 2001. *Farhang-e Tashrihi-ye Computer* (a translation of *Microsoft Press Computer Dictionary*). Daneshyar Publishers.
- Hooshmand-vizheh, M., translator. 2001. *Farhang-e Pezeshki-ye Dorland* (a translation of *Dorland's Medical Dictionary*). Tootiya Publishers.
- Harris, C. M., editor. Translated by M. R. Afzali and M. Hashemzade-Homayouni. 2002. *Farhang-e Me'mari va Sakhteman* (a translation of *A Dictionary of Architecture and Construction*). Daneshyar Publishers.

Persian-English technical dictionaries

In this area, what has been done is next to nothing. All I know is the following: *A Dictionary of Legal Parlance*, compiled by Kianoosh Seyed Hosseini and Ahmad-e Heydari, published by Mowj Publications 1998.

It should be added that many of English-Persian technical dictionaries have an index which enables the reader to use them as Persian-English dictionaries as well. It is unfortunate that so little has been done in this direction; but that is the way it is, “earning is always harder than borrowing.”

5 Conclusion

In this chapter, I have tried to present a survey of what has happened in the field of Persian lexicography in the past two decades. To summarize, we have witnessed the publication of some major monolingual (Persian-Persian) dictionaries. We have also seen that in the area of bilingual dictionaries, where foreign languages are used as the source and Persian as the target language, several serious and important works have been produced. In sharp contrast, where Persian is used as the source and foreign languages as the target language, not much work has been done. A large number of bilingual specialized dictionaries have been published, but the credibility of many of

them is in doubt. In spite of the fact that many significant contributions have been made to the field of Persian lexicography, still a great deal of work remains to be done. A final point worth mentioning is that Farhang-e Moaser, judging by the number and quality of its publications, is emerging as the main publisher of reference books, particularly of dictionaries, in Iran.

A Link Grammar Parser for Persian

Jon Dehdari and Deryle Lonsdale

1 Introduction

Persian is an Indo-European language with interesting morphological and syntactic properties. Verbs can express tense and aspect, and they agree with the subject in person and number. Nouns can host pronominal clitics marked for dative and genitive pronouns, as in the word بدستت *bedāstæt* “to your hand.” Verb forms like دیدیمتان *didimetān* “we saw you” host accusative pronominal clitics. Standard morphophonological changes such as epenthesis, assimilation, and deletion occur at morpheme boundaries. Some vowel harmony also occurs, however, as in the word نمی‌آیم *nemiāyæm* “I’m not coming”, where the negation prefix *næ* changes to *ne* (see Mahootian 1997: 306–8). Glossed morphological analyses for these words are given in (1).

- (1) a. be+ dāst +æt
DAT+ hand +2.S.GEN
“to your hand”
- b. di +d +im +etān
see +PAST +1.P.NOM +2.P.ACC
“We saw you.”
- c. næ +mi +ā +æm
NEG +DUR +come +1.S.NOM
“I am not coming.”

Issues naturally arise when using an orthography originally designed for a Semitic language. Some morphophonological phenomena, such as vowel harmony, do not show up in the orthography. Also, the distinction between affixes, clitics, and words is further complicated with the use of zero-width joiners, zero-width non-joiners, word spaces, and narrow no-break spaces¹ (Megerdooian 2000c). Since the genitive *ezāfe* marker is not normally visible in the written form, ambiguities in syntactic part-of-speech assignment and semantic roles can arise in text-based parsing.

¹See also <http://www.laits.utexas.edu/persian/persianword/persianwp.htm>

Another consideration in parsing Persian is the morphosyntactic relation of light verb constructions. Nouns, adjectives, or prepositional phrases (among other categories) can combine with light verbs like “do” (*kærdæn*) and “have” (*dāštæn*). The resulting word pair usually derives new, non-compositional meaning. Thus دوست—داشتن *dust—dāštæn* (lit. “friend—have”) means “to like.” Megerdoomian (2002) offers an in-depth treatment of these constructions.

Computational processing of Persian appears to be somewhat under-explored. One notable exception is the Shiraz project,² which employs a unification-based morphology engine (Megerdoomian 2000b) and a chart parser (Amtrup et al. 1999) for Persian-English machine translation (Amtrup et al. 2000). Another approach somewhat closer to the one we have taken underlies the Perslex engine (Riazati 1997), a Persian two-level morphology processor, the public availability of which is unclear. Finally, a Persian stemming implementation has been developed for information retrieval purposes (Tashakori et al. 2002).

In this paper we discuss efforts to develop a new system for parsing Persian, called Persian LG. This enterprise was viewed as worthwhile and appropriate since it involves novel ways of integrating morphological and syntactic processing. Its modules are also built on open-source technologies, like two-level morphology and link grammar parsing, which have been used in similar applications.

2 Morphological preprocessing

The first component in our processing approach performs morphological decomposition. This can be carried out by a lexicon-dependent morphological analyzer or a lexicon-independent stemmer, as seen in Fig. 1. After the input text has been morphologically decomposed, it is then syntactically parsed. The PC-Kimmo morphological analyzer and the Persian LG syntactic parser make use of a lexicon, which will be discussed below.

2.1 Morphological analyzer

The lexicon-dependent morphology engine is based on the two-level finite-state approach (Koskenniemi 1983) and uses the PC-Kimmo engine, which is capable of recognizing (i.e. analyzing or parsing) and generating

²<http://crl.nmsu.edu/shiraz>

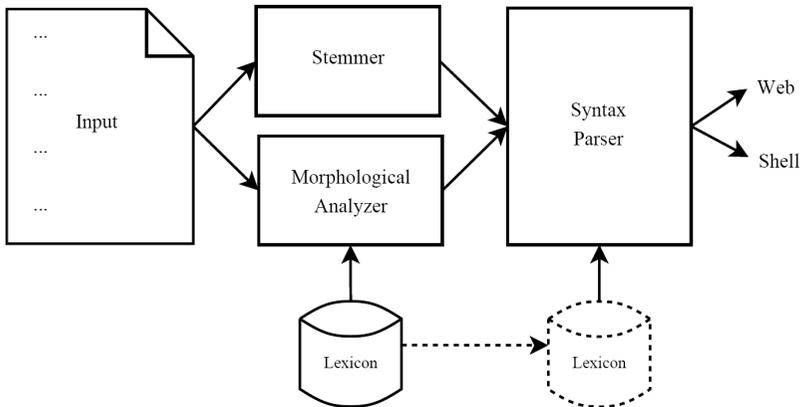


Figure 1: An overview of the parsing system.

(i.e. synthesizing or composing) word-based morpheme sequences. Similar morphology engines have been developed for such languages as Arabic (Beesley 1997; Beesley and Karttunen 2003), Turkish (Ofrazier 1994), Armenian (Lonsdale and Danielyan 2005), Oriya (Shabadi 2003), and many others. As mentioned above, the Shiraz (Megerdooian 2000b) and Perslex (Riazati 1997) projects have also developed morphology engines for Persian.

Practical considerations prompted us to implement a transliteration scheme for Persian. For example, our morphology and syntax engines do not accommodate non-Roman characters. Perl software assures a straightforward conversion between commonly used character sets (e.g. ISIRI 3342,³ CP-12564,⁴ and UTF-8) and our romanized input/output. A strict 1:1 correspondence to the orthography of Persian allows lossless conversion.

The PC-Kimmo morphology engine has three principal components: a set of rules, a collection of morpheme lexicons, and a phrase-structure grammar. The engine mediates two levels of a word (the lexical and surface representations) via a dozen rules that specify systematic morphophonological changes. Traditional grammars were used to conceptualize and develop appropriate rules (Mace 2003; Mahootian 1997). These rules were compiled via the KGen rule compiler⁵ into state transition tables which collectively specify a finite-state transducer architecture.

³<http://www.isiri.org/std/3342.htm>

⁴<http://www.microsoft.com/globaldev/reference/sbcs/1256.htm>

⁵http://www.sil.org/pckimmo/about_pc-kimmo.html

```

RULE 0:i <= V1 +:0 ___ [ Pn | m A n | t A n | C A n | e:0 ] #
RULE e:0 <=> V1:V1 +:0 0:i ___ #

```

Figure 2: Two sample PC-Kimmo rules.

Fig. 2 shows two sample rules. These two rules often work in tandem and with other rules to handle epenthesis, even in situations where the surrounding vowels are not visible on the surface. The first rule states that a surface letter ‘i’ must (<=) delete (i.e. correspond to null) if preceded by a morpheme-final long vowel and if followed by any of the various suffixes listed, including an *ezāfe* particle. The second rule, an iff (<=>) rule, complements the first by always deleting the *ezāfe* particle (lexical ‘e’) when word-final and when preceded by a morpheme-final long vowel and an epenthesized ‘i’. So بوی *bui buye* “smell of” might appear in a text and would resolve to بو + bu+e at the lexical level. This knowledge substantially contributes to the subsequent syntactic parsing stage.

A lexicon licenses valid morpheme sequences and specifies various properties of lexical and grammatical morphemes (Antworth 1990). The lexicon system contains seven different repositories for three types of morpheme-related information: fully vowelled lexical forms, English glosses, and featural/constraint information. Each lexical category (V, N, P, A) has its own lexicon; other lexicons have been developed for affixes, proper nouns, and function words (prepositions, conjunctions, numbers, etc.). Unfortunately, no freely usable and easily adaptable machine-readable Persian lexicon was readily available during development. In our case lexical information was derived from standard reference dictionaries, such as Steingass (1892).

The third component, the word grammar, permits display of a word’s morphological structure in hierarchical format. A set of 16 context-free word-formation rules specify and constrain Persian lexical substructure.

Recognition of a word returns all parses in three possible formats: a sequence of lexical morphemes, a corresponding sequence of English morphemic glosses, and an optional word-structure parse tree. The top part of Fig. 3 displays recognition results for the surface form *bui*, which is lexically *bu+e*. Gloss information is also displayed (“smell+EZ”). Fig. 3 (middle) recognizes *نمی گویم* *nemiguyim* “We are not saying.” The first rule of Fig. 2 epenthesizes a *y* between the present tense verb root *گو* *gu* “say” and the first person plural suffix *یم* *im*. A rule governing vowel harmony with the negative prefix *nae* would have also been employed, had the text been fully vowelled. The final part of Fig. 3 shows recognition of a complex verbal

```

PC-KIMMO>recognize bui
bu+e smell+EZ
PC-KIMMO>recognize nmi-guim
n+mi-+gu+im NEG+DUR+say.PRES+1P
PC-KIMMO>recognize nmi-binmC
n+mi-+bin+m+C NEG+DUR+see.PRES+1S+3s.object

```

Figure 3: Morphological decomposition for three words of varying complexity.

| | Baseline | PC-Kimmo | Perstem |
|-----------------|----------|----------|---------|
| <i>Accuracy</i> | — | 86% | 91% |
| <i>Coverage</i> | 82% | 92% | 97% |

Table 1: Evaluation of the two morphological analyzers

form, نمی بینمش *nemibinæmeš* “I don’t see it.”

The morphology engine has undergone considerable development, but some work remains to be done. Less commonly needed morphophonological rules still have to be written, and of course more lexicon development is necessary to extend coverage. In spite of this limited lexicon, recent evaluations have shown promising results. We used a random sampling of 500 unseen words from corpora that we built from Kayhan news⁶ and BBC Persian news.⁷ When words had multiple parses, the first parse was used. When words were not recognized, the word was taken in its entirety. The baseline, where no words were morphologically decomposed, was correct 82% of the time. More than 92% of the test words were morphologically analyzed correctly, as is seen in Table 1. Of the words which were morphologically decomposed, 96% were analyzed correctly.

2.2 Stemmer

While the Persian PC-Kimmo engine provides excellent accuracy, it is currently unable to perform morphological operations on words not found in its lexicon. We developed a lexicon-independent stemmer/shallow morphological parser that can be used as an alternative to the PC-Kimmo engine,

⁶<http://www.kayhannews.ir>

⁷<http://www.bbc.co.uk/persian>

or in conjunction with it. The stemmer, *Perstem*, is written in Perl and uses regular expression substitutions to separate inflectional morphemes, and optionally remove affixes.⁸ The stemmer currently has 76 substitution rules, which replace one pattern of text with another.

نمی گوییمتان → nmi-guiimtAn → n+_mi-+_gu_+0_+im_+tAn → n mi gu
im tAn
نمی گویمتان → nmi-guiimtAn → n+_mi-+_gu_+0_+m_+tAn → n mi gu m
tAn
کتاب های → ktAb-hAi → ktAb_+-hA_+e → ktAb hA e

Figure 4: The stages of *Perstem*'s morphological decomposition for the words *nemi-guyimetān*, *nemi-guyæmetān*, and *ketāb-hāye*.

Fig. 4 shows the decomposition stages for the verbs *nemi-guyimetān* “we do not tell you”, *nemi-guyæmetān* “I do not tell you”, and the nominal fragment *ketāb-hāye* “books of.” The morphemes in the final stage serve as the input for the syntactic parser, where they are linked with other words and morphemes. This ensures agreement between features, such as number and person.

Currently *Perstem* is the primary means of morphological decomposition for Persian LG, due to its flexibility and robustness. *Perstem* can process about 12200 words per second on an UltraSPARC-III machine, ten times faster than the PC-Kimmo analyzer. Using the previously mentioned testing words, *Perstem* correctly analyzed 97% of the words.⁹ Of the words that were morphologically decomposed, 91% were analyzed correctly. The use of a lexicon clearly helps eliminate incorrect analyses, but requires more processing time and extensive development time. *Perstem*'s coverage is comparable with Megerdoomian (2004), as is the accuracy of Persian PC-Kimmo. Preliminary testing has shown that integrating the PC-Kimmo engine with *Perstem* markedly increases coverage with only a small loss in accuracy compared with the PC-Kimmo results. Tashakori et al. (2002) uses stemming evaluation metrics that are not comparable with the aims of this paper.

⁸*Perstem* may be downloaded at <http://sourceforge.net/projects/perstem>

⁹Evaluation data are found at <http://ling.ohio-state.edu/~jonsafari>